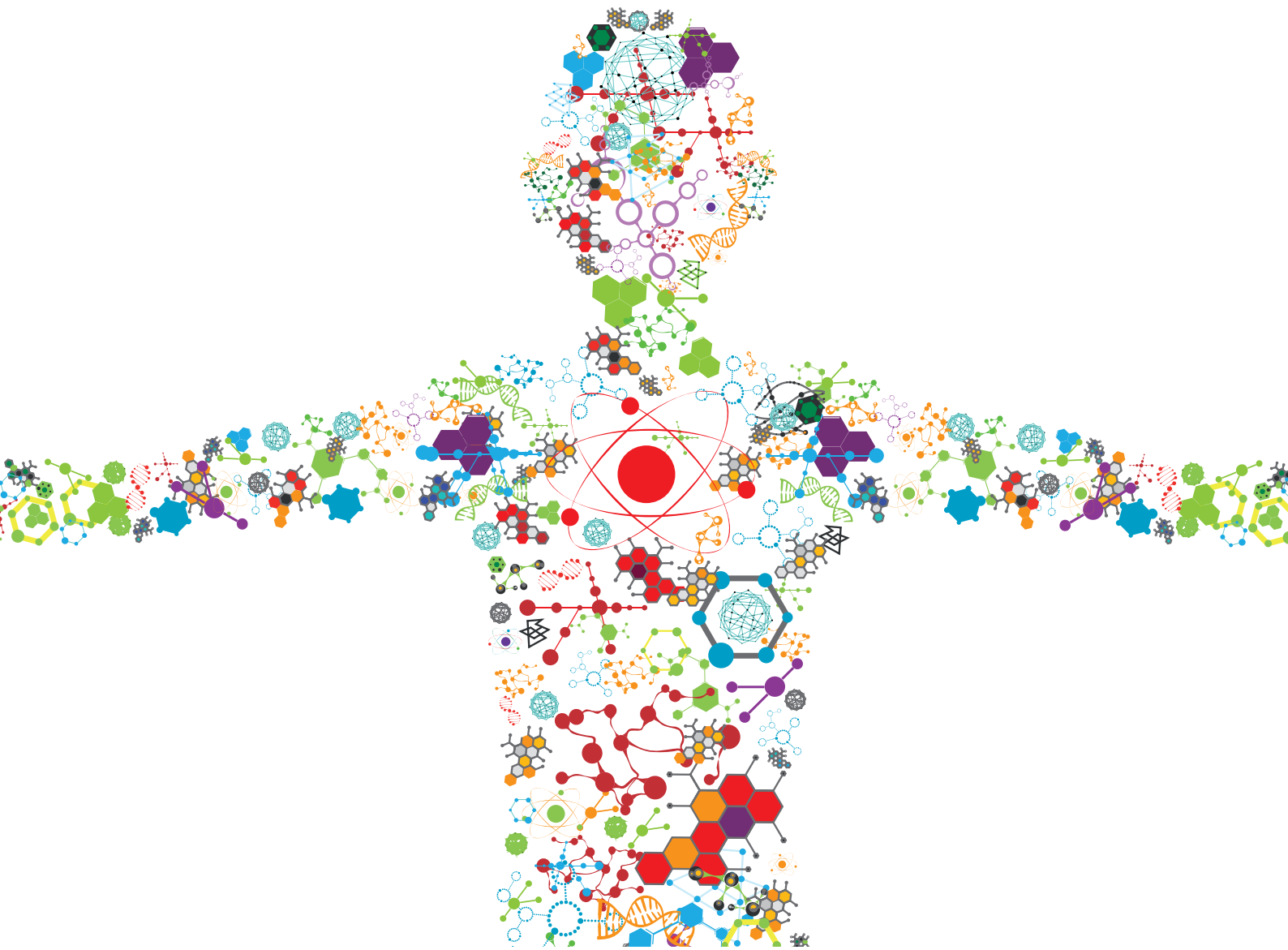
MULTI-OMICS TECHNOLOGIES FOR OPTIMIZING SYNTHETIC BIOMANUFACTURING

EDITED BY: Young-Mo Kim, Chris Petzold, Eduard Kerkhoven and Scott E. Baker

PUBLISHED IN: Frontiers in Bioengineering and Biotechnology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-237-0
DOI 10.3389/978-2-88974-237-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MULTI-OMICS TECHNOLOGIES FOR OPTIMIZING SYNTHETIC BIOMANUFACTURING

Topic Editors:
Young-Mo Kim, Pacific Northwest National Laboratory (DOE), United States
Chris Petzold, Lawrence Berkeley National Laboratory, United States
Eduard Kerkhoven, Chalmers University of Technology, Sweden
Scott E. Baker, Pacific Northwest National Laboratory (DOE), United States

Citation: Kim, Y.-M., Petzold, C., Kerkhoven, E., Baker, S. E., eds. (2022). Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-237-0

Table of Contents

frontiers in Bioengineering and Biotechnology

EDITORIAL
published: 15 December 2021
doi: 10.3389/fbioe.2021.818010

Editorial: Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing

Young-Mo Kim[1,2]*, *Christopher J. Petzold*[2,3], *Eduard J. Kerkhoven*[4] *and Scott E. Baker*[2,5]

[1]Integrative Omics Group, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States, [2]Department of Energy, Agile BioFoundry, Emeryville, CA, United States, [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [4]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, [5]Functional and Systems Biology Group, Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States

Keywords: biomanufacturing, multi-omics analysis, synthetic biology, DBTL cycle, metabolic engineering

Editorial on the Research Topic

Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing

Industrial manufacturing endures as an essential human activity yielding a variety of useful products; it plays a significant role in the global economy with huge impacts in everyday life. However, the manufacturing process requires consumption of various raw materials (especially petroleum derivatives), generates a variety of harmful waste products, causes pollution, and is energetically inefficient. Biological manufacturing from sustainable, affordable, and scalable feedstocks potentially enables the displacement of the entire portfolio of currently available products produced by industrial processes, enabling the manufacturing of renewable and eco-friendly products (Clomburg et al., 2017). Thus, successful development of a robust biomanufacturing strategy and technology platform, based on the latest advances in synthetic biology and chemical catalysis, will decrease both the cost and production time compared with previous manufacturing processes. Development of biomanufacturing processes using a synthetic biology platform requires the multidisciplinary efforts of science and engineering fields including molecular biology, microbiology, genetic engineering, informatics, metabolic modeling and chemical or process engineering (El Karoui et al., 2019).

In this research topic, Amer and Baidoo discussed the importance of using multi-omic analytical approaches to monitor and improve the biomanufacturing process. These approaches include genomics, transcriptomics, proteomics, metabolomics and fluxomics (**Figure 1**). The multi-omics data acquired from the biomanufacturing process not only provides potential solutions to low production efficiency by identifying underlying metabolic bottlenecks or pathway sinks, but also guides the understanding of how these modified biological systems function. Furthermore, such multi-omics technologies are constantly innovated and improved to expand molecular detection coverage, obtain data with increased accuracy by using new or novel analytical instruments, achieve better computational algorithms, and create wider and deeper databases to support a growing variety of biological host systems. Roy et al. described a combined computational tool to optimize the DBTL (Design-Build-Test-Learn) cycle in biomanufacturing process by collecting, visualizing, and utilizing large multi-omics datasets from various biological systems and emphasized their importance in the following metabolic engineering processes with machine learning approaches.

Gao et al. compared microflow and nanoflow liquid chromatography-selected reaction monitoring (LC-SRM) methods for analysis of hundreds of targeted peptides associated with 132 proteins in major pathways of *Pseudomonas putida,* a versatile bacterial host for production

OPEN ACCESS

Edited and reviewed by:
Jean Marie François,
Institut Biotechnologique de Toulouse (INSA), France

***Correspondence:**
Young-Mo Kim
young-mo.kim@pnnl.gov

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *18 November 2021*
Accepted: *29 November 2021*
Published: *15 December 2021*

Citation:
Kim Y-M Petzold CJ, Kerkhoven EJ and Baker SE (2021) Editorial: Multi-Omics Technologies for Optimizing Synthetic Biomanufacturing. Front. Bioeng. Biotechnol. 9:818010. doi: 10.3389/fbioe.2021.818010

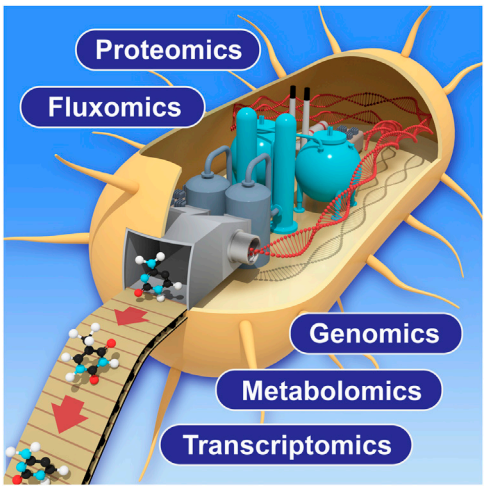


FIGURE 1 | Schematic view of multi-omics application to biomanufacturing process. Improvement of each technology will enhance the measurement coverage and accuracy in future applications.

of bioproducts and biofuels *via* metabolic engineering. The increased throughput and accuracy of protein measurement will not only reduce the DBTL cycle time in future applications, but is, in addition, easily applied to other biomanufacturing host organisms.

Fletcher and Baetz reviewed the toxicity of phenolic compounds which are produced from pretreatment or hydrolysis of natural lignocellulosic biomass based on functional genomics and transcriptomics approaches, especially to the important model organism and industrial bioproduction host strain, *Saccharomyces cerevisiae*. Information regarding physiological tolerances of toxic phenolic compounds may be applied and evaluated in other host strains for future improvement. In that regard, Garcia et al. developed the genome-scale metabolic model of *Clostridium thermocellum* for efficient conversion of lignocellulosic biomass which has unique preference for its anaerobic and thermophilic growth attributes. This model will provide a useful tool to understand physiological and metabolic parameters associated with potential future biomanufacturing process.

Pinheiro et al. studied a xylose metabolism by *Rhodosporidium toruloides*, an oleaginous yeast with significant emerging potential in industrial applications, using a detailed physiological characterization interpreted with absolute proteomics and genome scale metabolic models. Kim et al. performed a multi-omics analysis on *R. toruloides* and the transcriptomics, proteomics, metabolomics and RB-TDNA sequencing data improved the current genome-scale model to make it a more exhaustive and accurate metabolic network model.

Pomraning et al. integrated high-throughput proteomics and metabolomics data as part of a DBTL cycle focused on improving production efficiency of 3-hydroxypropionic acid (3HP) in engineered *Aspergillus pseudoterreus* strains. This was the first report of 3HP production in a filamentous fungus amenable to industry-level biomanufacturing of organic acids at high titer and low pH. Chroumpi et al. studied another filamentous fungus *Aspergillus niger* for better understanding of pentose catabolic pathways by deletion of the key genes. The high-throughput multi-omics data (i.e., transcriptome, metabolome and proteome) generated on the mutant strains revealed that these genes are critical for metabolic pathways but not as critical for growth of *A. niger* on more complex biomass substrates, which raises fundamental questions on nutrient acquisition during growth on various carbon sources.

Wu et al. investigated the metabolic potential of *Zymomonas mobilis* for conversion of glucose and xylose to 2,3-butanediol. This study used calculated thermodynamic and kinetic parameters to generate insights of *Z. mobilis* metabolism. They also performed pathway and dynamic flux balance analysis to understand metabolic potential and production efficiency for future industrial applications. Nitta et al. acquired metabolomics and transcriptomics data on antibiotic producing strain, *Streptomyces coelicolor* to understand the functional connections between the production of antibiotic, actinorhodin and the level of cAMP. They found that higher levels of cAMP improved cell growth and production of actinorhodin, which was confirmed by the metabolomic and transcriptomic data.

We conclude by emphasizing that high-throughput multi-omics data play a critical role to unravel the complexities of metabolic engineering to improve production efficiency and product titer produced by a variety of industrial microbes. In addition, generation of multi-omics datasets accelerates the adoption and subsequent application of artificial intelligence approaches such as machine learning to design of improved microbial bioproduction host systems (Lawson et al., 2021). In terms of technological perspectives, enhanced high-throughput measurement and improved coverage of multi-omics analyses with higher accuracy will not only benefit in shortened DBTL cycle times for the metabolic engineering process, but also will lead to improved fundamental understanding of engineered biosystems. Refining tools and analytical platforms will benefit manipulating, modifying, and reshaping potential host systems. The long-term outcomes of these efforts will impact the world and our future by decarbonizing the current manufacturing processes *via* an environmental-friendly manner.

AUTHOR CONTRIBUTIONS

Y-MK, CP, EK, and SB served as co-editors for the Research Topic: Multi-omics technologies for optimizing synthetic biomanufacturing. Y-MK conceived of the idea for the research topic, and all the authors contributed to writing the editorial.

FUNDING

The work was supported by Agile BioFoundry (http://agilebiofoundry.org), funded by the United States Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, under Award No. DE-NL0030038. Pacific Northwest National Laboratory (PNNL) is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830.

REFERENCES

Clomburg, J. M., Crumbley, A. M., and Gonzalez, R. (2017). Industrial Biomanufacturing: The Future of Chemical Production. *Science* 355, aag0804. doi:10.1126/science.aag0804

El Karoui, M., Hoyos-Flight, M., and Fletcher, L. (2019). Future Trends in Synthetic Biology-A Report. *Front. Bioeng. Biotechnol.* 7, 175. doi:10.3389/fbioe.2019.00175

Lawson, C. E., Martí, J. M., Radivojevic, T., Jonnalagadda, S. V. R., Gentz, R., Hillson, N. J., et al. (2021). Machine Learning for Metabolic Engineering: A Review. *Metab. Eng.* 63, 34–60. doi:10.1016/j.ymben.2020.10.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 19 August 2020
doi: 10.3389/fbioe.2020.01008

Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in *Rhodotorula toruloides*: Insights for Future Biorefinery

Marina Julio Pinheiro[1,2†], Nemailla Bonturi[1†], Isma Belouah[1], Everson Alves Miranda[2] and Petri-Jaan Lahtvee[1]*

[1] Institute of Technology, University of Tartu, Tartu, Estonia, [2] Department of Materials and Bioprocess Engineering, School of Chemical Engineering, University of Campinas, Campinas, Brazil

OPEN ACCESS

Edited by:
Chris Petzold, Lawrence Berkeley National Laboratory, United States

Reviewed by:
Zongbao K. Zhao, Dalian Institute of Chemical Physics (CAS), China
John Gladden, Sandia National Laboratories (SNL), United States

****Correspondence:***
Petri-Jaan Lahtvee
petri.lahtvee@ut.ee

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *02 June 2020*
Accepted: *31 July 2020*
Published: *19 August 2020*

Citation:
Pinheiro MJ, Bonturi N, Belouah I, Miranda EA and Lahtvee P-J (2020) Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in Rhodotorula toruloides: Insights for Future Biorefinery. Front. Bioeng. Biotechnol. 8:1008. doi: 10.3389/fbioe.2020.01008

The use of cell factories to convert sugars from lignocellulosic biomass into chemicals in which oleochemicals and food additives, such as carotenoids, is essential for the shift toward sustainable processes. *Rhodotorula toruloides* is a yeast that naturally metabolises a wide range of substrates, including lignocellulosic hydrolysates, and converts them into lipids and carotenoids. In this study, xylose, the main component of hemicellulose, was used as the sole substrate for *R. toruloides*, and a detailed physiology characterisation combined with absolute proteomics and genome-scale metabolic models was carried out to understand the regulation of lipid and carotenoid production. To improve these productions, oxidative stress was induced by hydrogen peroxide and light irradiation and further enhanced by adaptive laboratory evolution. Based on the online measurements of growth and CO_2 excretion, three distinct growth phases were identified during batch cultivations. Majority of the intracellular flux estimations showed similar trends with the measured protein levels and demonstrated improved NADPH regeneration, phosphoketolase activity and reduced β-oxidation, correlating with increasing lipid yields. Light irradiation resulted in 70% higher carotenoid and 40% higher lipid content compared to the optimal growth conditions. The presence of hydrogen peroxide did not affect the carotenoid production but culminated in the highest lipid content of 0.65 g/g_{DCW}. The adapted strain showed improved fitness and 2.3-fold higher carotenoid content than the parental strain. This work presents a holistic view of xylose conversion into microbial oil and carotenoids by *R. toruloides*, in a process toward renewable and cost-effective production of these molecules.

Keywords: microbial oil, carotenoids, *Rhodotorula toruloides*, Genome-scale modelling, xylose, biorefinery, absolute proteomics

INTRODUCTION

Rhodotorula toruloides is considered one of the most promising oleaginous yeasts for industrial applications. This microorganism is a natural producer of lipids (microbial oil) and high-value compounds, such as carotenoids and enzymes for pharma and chemical industries (L-phenylalanine ammonia-lyase and D-amino acid oxidase) (Park et al., 2017). The microbial

oil, which is primarily composed of triacylglycerides (TAGs), is a potential raw material for oleochemicals that can be used as biodiesel, cosmetics, and coatings as well as a replacement of vegetable oil in fish feed (Unrean et al., 2017; Blomqvist et al., 2018; Yang et al., 2018). Carotenoids are important molecules for different industries, such as the food, chemical, pharmaceutical and cosmetics industries. In addition to its colorants properties, carotenoids can be metabolised into vitamin A and have antioxidant activity that has been explored, for example, in the prevention of cancer, immune diseases and as skin protection against radiation (Stahl and Sies, 2007; Du et al., 2016; Kot et al., 2019). The global market for carotenoids should reach US$2.0 billion by 2022 (BBC Research, 2018), while the global demand for fatty acids (FAs) and alcohols is expected to reach over 10 Mt in 2020 (Adrio, 2017).

In addition to the ability to produce a variety of relevant compounds, *R. toruloides* can consume a range of carbon and nitrogen sources (Park et al., 2017; Lopes et al., 2020b), including lignocellulosic hydrolysates (Bonturi et al., 2017; Lopes et al., 2020a). Following cellulose, hemicellulose is the second most abundant fraction of lignocellulose, and this fraction consists of polymerised five-carbon sugars, mainly xylose. The efficient utilisation of xylose by a microorganism is essential to improve the conversion of lignocellulosic materials into target compounds, thus increasing the economic viability of the biotechnological processes in biorefineries. Therefore, understanding of the metabolic mechanisms involved in the production of lipids and carotenoids from xylose by *R. toruloides* is crucial to further improve the titres, yields and rates of this bioprocess.

The cellular content of lipids and carotenoids is affected by several factors, including medium composition and cultivation conditions (Mata-Gómez et al., 2014). Previous studies have described the increase in carotenoid production in the presence of oxidative stress, such as hydrogen peroxide (H_2O_2) and light irradiation. In the presence of H_2O_2 (5 mmol/L), a five-fold increase in carotenoid production by *Rhodotorula mucilaginosa* was observed (Irazusta et al., 2013), illustrating how optimisation of cultivation conditions can improve production yields of the desired metabolites. Under light irradiation (4,000 lux), the production of carotenoids and lipids by *Rhodotorula glutinis* increased 60% (Gong et al., 2020). The cellular response mechanism against oxidative stress is not clear for yeast, including *R. toruloides*. This condition is associated with the presence of reactive oxygen species (ROS), including H_2O_2, superoxide (O_2^-) and hydroxyl radicals (HO·). ROS are potent oxidants that can damage all cellular components, including DNA, lipids and proteins. Microbial cells possess two defensive systems against oxidative damage: enzymatic and non-enzymatic. The former is mainly constituted by enzymes superoxide dismutase and catalase, and the latter is involved in direct scavenging of ROS or recycling of oxidised compounds, such as ascorbate, glutathione, α-tocopherol, and carotenoids (Irazusta et al., 2013).

Integration of large-scale data sets is crucial for a better understanding of the genomic organisation and metabolic pathways in living cells. Complete genome sequences are available for several *R. toruloides* strains (Kumar et al., 2012; Zhu et al., 2012; Morin et al., 2014; Hu and Ji, 2016; Sambles et al., 2017; Coradetti et al., 2018). The lipid formation process during different growth phases of cultivation on glucose has been investigated through proteomic analysis (Liu et al., 2009) and compared with cells grown on xylose (Tiukova et al., 2019b). Multi-omics analyses have identified metabolism modification under nitrogen (Zhu et al., 2012) and phosphate limitation (Wang Y.et al., 2018). The latter studies have identified higher lipid accumulation under nitrogen or phosphate limitation, which has been correlated to higher activation in nitrogen recycling but also lipid degradation and autophagy. Carotenoid production from glycerol was investigated using global metabolomics, revealing reduced abundance of metabolites involved in TCA and amino acid biosynthesis (Lee et al., 2014).

Genome-scale metabolic models (GEMs) are another powerful tool to understand and provide a holistic view of metabolic fluxes, energy and redox metabolism or even suggest targets for metabolic engineering. GEMs are constructed based on the available genome sequence of a specific organism, thus providing a summary of the metabolic network (Kerkhoven et al., 2015). Regarding *R. toruloides*, a number of metabolic models are available to assess lipid production (Bommareddy et al., 2015; Castañeda et al., 2018), and two reports of GEMs are available (Dinh et al., 2019; Tiukova et al., 2019a). Lopes et al. (2020a) reported the first study that combined data from cultivations of *R. toruloides* under different carbon sources with the GEM. The approach proved to be useful for understanding metabolic fluxes and identifying targets to improve lipid production either by metabolic engineering or process optimisation.

Therefore, the current study aimed at providing a holistic view of lipid and carotenoid production by *R. toruloides* using xylose as a sole carbon source by combining detailed physiological characterisation with the quantitative proteomics and GEM analysis. Oxidative stress (H_2O_2 and light irradiation) and adaptive laboratory evolution (ALE) were employed to improve lipid and carotenoids production. To our knowledge, this is the first work combining such approaches for this strain, and the data obtained here can be used to establish future bioprocesses in biorefineries.

MATERIALS AND METHODS

Strain and Inoculum

Rhodotorula toruloides (previously known as *Rhodosporidium toruloides*) CCT 7815 was obtained from "Coleção de Culturas Tropicais" (Fundação André Tosello, Campinas, Brazil) and stored at −80°C in 10% (v/v) glycerol. This strain was derived from *R. toruloides* CCT 0783 (synonym IFO10076) after short-term adaptation in sugarcane bagasse hemicellulosic hydrolysate. Adaptation process resulted in no physiological changes regarding growth and substrate consumption profile between

both strains, but CCT 7815 produces more lipids and has higher induction of some genes related to hydrolysate-tolerance and lipid accumulation (Bonturi et al., 2017). Cultivation inoculum was prepared in YPX medium at 200 rpm and 30°C for 24 h. The cells were washed twice with 0.9% (v/v) NaCl before inoculation.

Adaptive Laboratory Evolution

Adaptive laboratory evolution was performed by successive shake flask cultivations of *R. toruloides* in the presence of H_2O_2 in rich medium (30.0 g xylose, 2.5 g glucose, 0.9 g yeast extract, 0.2 g $(NH_4)_2SO_4$, 1.5 g KH_2PO_4, and 0.9 g $MgSO_4{\cdot}7\ H_2O$). ALE was performed in two cycles with the aim of improving the performance of this yeast under this oxidative environment. The initial H_2O_2 concentration was 10 mmol/L (the highest concentration in which the cells grew in successive cultivations) in the first cycle and 20 mmol/L in the second cycle (the highest concentration that the parental strain tolerated). In each cycle, the cells were harvested at the exponential growth phase and transferred to a fresh medium with the same H_2O_2 concentration. The initial OD (at 600 nm) of every passage was 0.5. The end of the cycle was determined by the stabilisation of the maximum specific growth rate and the length of lag phase.

Yeast Cultivation

Batch cultivations were performed in 1-L bioreactors (Applikon Biotechnology, Delft, Netherlands) with a working volume of 600 mL at pH 6.0 and controlled with the addition of 2 mol/L KOH. Dissolved oxygen was maintained at greater than 25% using 1-vvm airflow and stirring speeds between 400 and 600 rpm. The cultivation started with 1% (v/v) of overnight culture inoculum. Oxidative stress was induced by the addition of H_2O_2 (20 mmol/L) at the start of the cultivation or by light irradiation (white LED light, 40,000 lux) throughout the experiment. The composition of CO_2 and O_2 in the gas outflow was measured using an online gas analyser (BlueSens gas sensor GmbH, Herten, Germany), and optical density was monitored online using a Bug Lab BE3000 Biomass Monitor (Bug Lab, Concord, CA, United States) at 1,300 nm. Data were collected and processed with BioXpert V2 software v. 2.95 (Applikon Biotechnology, Delft, Netherlands). All experiments were performed in triplicate.

The medium composition used in the bioreactor experiments was, per litre, 70.0 g xylose, 1.95 g $(NH_4)_2SO_4$, 3.0 g KH_2PO_4, 0.5 g $MgSO_4{\cdot}7\ H_2O$, 1.0 mL vitamins solution and 1.0 mL trace metal solution (Lahtvee et al., 2017), supplemented with 100 μl antifoam 204 (Sigma-Aldrich, St. Louis, MO, United States). Samples for dry cellular mass, carotenoid and extracellular metabolite analyses were collected every 24 h. Samples were withdrawn from bioreactors, transferred into precooled 2-mL Eppendorf tubes and centrifuged for 20 s at 4°C and 18,000 × *g*. The supernatant was collected and stored at −20°C for extracellular metabolite quantification. Cell pellets was snap-frozen in liquid nitrogen, stored at −80°C, and further used for proteomics analysis.

Quantification of Dry Cell Mass, Extracellular Metabolites, Carotenoids, Total Lipids and Proteins

Online turbidity measurements were calibrated by gravimetrically measuring the dry cellular mass (DCW) every 24 h. Extracellular metabolites in the broth were measured using HPLC (LC-2030C Plus, Shimadzu, Kyoto, Japan) equipped with a refractive index detector (RID-20A, Shimadzu, Kyoto, Japan). Concentrations of xylose, organic acids and glycerol were measured using an HPX-87H column (Bio-Rad, Hercules, CA, United States) at 45°C, and 5 mmol/L H_2SO_4 served as the mobile phase with isocratic elution at 0.6 mL/min. Xylitol and arabitol were quantified using Rezex RPM-Monosaccharide column (Phenomenex, Torrance, CA, United States) at 85°C, and purified water (Milli-Q Ultrapure Water System, Merck, Darmstadt, Germany) used as the mobile phase with isocratic elution at 0.6 mL/min.

For quantification of carotenoids (modified from Lee et al., 2014), 2 mL of cells were harvested by centrifugation, washed twice in distilled water and resuspended in 1.0 mL of acetone. The cells were lysed with acid-washed glass beads (400–650 μm) using the FastPrep homogeniser for three cycles (4 m/s for 20 s) (MP Biomedicals, Irvine, CA, United States). After centrifugation at 15,000 × *g* for 5 min, the acetone solution containing carotenoids was collected and stored at 4°C. These steps were repeated until the cell debris was colourless. Then, the solvent was evaporated in Concentrator Plus (Eppendorf, Hamburg, Germany), and the remaining extracts were resuspended in a known volume of acetone. Carotenoids were measured using Acquity UPLC (Waters, Franklin, MA, United States) equipped with a TUV detector (Waters, Franklin, MA, United States) and C18 column (BEH130, 1.7 μm, 2.1 × 100 mm, Waters, Franklin, MA, United States). The mobile phase was a gradient from 80 to 100% of acetone in purified water at a flow rate of 0.2 mL/min. Detection was performed at 450 nm (modified from Weber et al., 2007). All identified peaks were quantified using the β-carotene standard (Alfa Aesar, Tewksbury, MA, United States). Detected peaks were identified according to the known carotenoid retention time profile (Weber et al., 2007; Lee et al., 2014).

Lipids were extracted according to an adaptation of the Folch method (Folch et al., 1957) described by Bonturi et al. (2015). Briefly, a mixture of chloroform and methanol (2:1 v/v) was added to dried cells. After 24 h, the solvent was evaporated in a rotary evaporator (Buchi, Flawil, Switzerland), and the total lipid content was determined gravimetrically.

Total proteins were extracted from 600 μg of cells resuspended in Y-PER solution (Thermo Fisher Scientific, Waltham, MA, United States) in a 2-mL Eppendorf tube. This suspension was incubated at 30°C for 45 min. Then, glass beads were added in the tube, and cell lysis was performed in a FastPrep homogeniser during ten cycles (4 m/s for 20 s). Between the cycles, the tubes were placed on ice for 3 min. After centrifugation at 18,000 × *g* and 4°C for 10 min, the supernatant was carefully removed and stored at 4°C for further protein quantification. Y-PER reagent was added to the remaining cells in the tube, and the cell

disruption process was repeated. This step was performed until no protein was detected. Total protein was quantified using a commercially available assay (Micro BCA^{TM} Protein Assay Kit, Thermo Fisher Scientific), and samples were diluted in the linear range of BSA protein standard (0.5–20 μg/ml).

Proteome Analysis

Fully labelled cellular biomass was used as the internal standard in absolute proteome analysis, and produced by cultivating *R. toruloides* in minimal mineral medium containing labelled heavy 15N, 13C-lysine (Silantes, Munich, Germany). Heavy labelling of the proteinogenic lysine was measured at 96.6% (data not shown). Absolute proteomics and internal heavy-labelled standard preparation and analyses were performed similarly as described in Lahtvee et al. (2017) and Kumar and Lahtvee (2020). Briefly, cells were resuspended in the lysis buffer (6 mol/L guanidine HCl, 100 mmol/L Tris–HCl pH 8.5 and 100 mmol/L dithiothreitol) and homogenised with glass beads using the FastPrep24 device. The supernatant was removed by centrifugation (17,000 × *g* for 10 min at 4°C) and precipitated overnight with 10% trichloroacetic acid at 4°C. Pellets from the previous precipitation step were spiked in 1:1 ratio with the heavy-labelled standard. This mixture was further precipitated with 10% TCA. The pellet was resuspended in a buffer containing 7 mol/L urea and 2 mol/L thiourea in 100 mmol/L ammonium bicarbonate (ABC) followed by reduction using 5 mmol/L DTT and alkylation with 10 mmol/L chloroacetamide. Peptides were digested at room temperature with *Achromobacter lyticus* Lys-C (Wako Pure Chemical Industries, Osaka, Japan) for 4 h at the ratio of 1:50 (enzyme:protein) followed by overnight digestion of the previous solution diluted five times in 100 mmol/L ABC buffer. Peptides were desalted using in-house prepared C18 (3M Empore, Maplewood, MO, United States) tips and were reconstituted in 0.5% trifluoroacetic acid (TFA). For separation, 2 μg of peptides was injected on an Ultimate 3000 RSLCnano system (Dionex, Sunnyvale, CA, United States) coupled to a C18 cartridge trap-column in a backflush configuration and an analytical 50 cm × 75 μm emitter-column (New Objective, Woburn, MA, United States) in-house packed with 3 μm C18 particles (Dr. Maisch, Ammerbuch, Germany). Eluted peptides were sprayed to a quadrupole-orbitrap Q Exactive Plus (Thermo Fisher Scientific, Waltham, MA, United States) tandem mass spectrometer. MaxQuant 1.4.0.8 software package was used for raw data identification and identification (Cox and Mann, 2008). *R. toruloides* NP11 served as a reference proteome database in UniProt[1]. Protein quantification was performed following the total protein approach described in Sánchez et al. (2020) and assuming 90% coverage from the total protein abundance.

LC-MS/MS proteomics data were deposited in the ProteomeXchange Consortium[2] via the PRIDE partner repository (Vizcaíno et al., 2016) and can be retrieved using the dataset identifier PRIDE: PXD019305. Processed quantitative data are presented in **Supplementary Table S7**. Triplicated quantitative proteomics data were used for differential expression analysis. *p*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). Additional data analysis included gene set analysis (GSA) (carried out using PIANO platform; Väremo et al., 2013) and gene enrichment analysis (g:Profiler; Raudvere et al., 2019).

Genome-Scale Modelling

The intracellular flux patterns were predicted using the *R. toruloides* genome-scale metabolic network rhto-GEM version 1.2.1 and flux balance analysis (FBA) (Tiukova et al., 2019a; Lopes et al., 2020a). Models and experimental rates used for the reference condition (REF_P1, REF_P2, and REF_P3) were uploaded on GitHub repository[3]. Random sampling simulations, quantitative proteomics data and the reference file (gmt format) used for the gene enrichment analysis (g:Profiler) are hosted on the same repository. The model was improved by adding carotenoids (β-carotene, γ-carotene, torulene and torularhodin) into the biomass pseudo-reaction, and adding D-arabitol production pathway (Jagtap and Rao, 2018). To allow the model to predict the production of either D- or L-arabitol, *in silico* pseudo-reactions were added, converting, both, D- and L-arabitol into an "artificial" arabitol without a specific isoform. FBA was performed to calculate intracellular flux patterns using RAVEN Toolbox (Wang H.et al., 2018) on MATLAB (The MathWorks Inc., Natick, MA, United States), Gurobi solver (Gurobi Optimization Inc., Houston, TX, United States) and by optimising for non-growth related ATP maintenance (NGAM). The latter was followed by flux variability analysis (random sampling at $n = 5000$) at 95% from the maximal NGAM value. Experimental data obtained from the yeast cultivations were used to constrain the model if not stated otherwise. Cellular biomass composition was adjusted to the measured total protein, lipid and carotenoid content.

RESULTS

Three Distinct Phases of *R. toruloides* Growth on Xylose

Rhodotorula toruloides growth was characterised with xylose as a sole carbon source under aerobic batch conditions on a mineral medium. The initial xylose concentration of 70 g/L was chosen, and the amount of nitrogen was adjusted to result in a C/N ratio of 80 mol/mol. On-line monitoring of culture turbidity, CO_2 production and O_2 consumption was complemented by off-line analysis of sugars, alcohols and cell mass composition (total lipids, proteins and carotenoids; **Figure 1**). During the batch cultivation of *R. toruloides*, three distinct growth phases were observed based on growth dynamics and substrate consumption patterns. In the first growth phase (P1), cells were growing exponentially without any observable limitation, and xylose was used as the sole carbon source. Arabitol, xylitol and CO_2 were the main fermentation by-products detected. Phase two (P2) started with a sudden decrease in the specific growth rate due to nitrogen limitation. At that point, approximately 23 g/L of xylose

[1]www.uniprot.org

[2]http://proteomecentral.proteomexchange.org

[3]https://github.com/SynBioUniTartu/Rhto_OxidativeStress

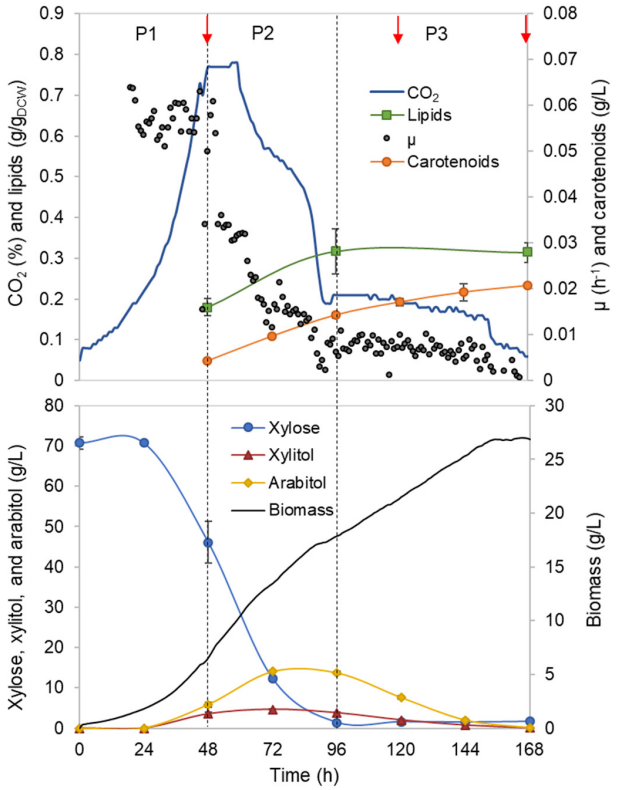


FIGURE 1 | *Rhodotorula toruloides* batch cultivation on xylose under the reference (optimal) environmental conditions for the parental strain. Dashed vertical lines define three observed growth phases. The specific growth rate (μ), cell mass concentration, intracellularly accumulated lipid and carotenoid concentrations, extracellular metabolite profiles and CO_2 production profile in the outflow gas are presented. The values represent an average of three independent cultivation experiments; error bars represent standard deviation. The red arrows represent the proteomic data points.

was consumed, indicating a critical C/N ratio of 26 (mol/mol) for *R. toruloides* to reach nitrogen limitation. P2 lasted until the depletion of the primary carbon source, namely, xylose. Consumption of arabitol and xylitol under nitrogen limitation defined the third growth phase (P3). To our knowledge, no previous study has provided a characterisation of *R. toruloides* physiology on xylose in such detail. These phases were further analysed in this work, aiming to identify the cellular metabolic behaviour in response to the environment changes during the batch growth.

First growth phase (P1) comprised the highest specific growth rate (μ, 0.060 ± 0.001 h^{-1}) and the specific xylose uptake rate (r_{XYL}, 1.74 ± 0.12 mmol/g_{DCW}.h) (**Supplementary Table S1**), while no nutrient-level limitations were detected. During the exponential growth, approximately one-third of the consumed carbon was secreted as arabitol and xylitol. An additional one-third of the consumed carbon was secreted as CO_2. Although the carotenoid content in P1 was low (0.66 ± 0.06 mg/g_{DCW}), the specific production rate (r_{CAR}) was the highest (0.042 ± 0.003 mg/g_{DCW}.h) due to the higher μ (**Figure 2A**, **Supplementary Table S1**). The biomass yield and lipid content were 0.25 ± 0.02 and 0.18 ± 0.03 g/g_{DCW}, respectively. Carbon balance in this phase was estimated at 93%, indicating a small amount of undetected by-products.

Nitrogen limitation under the xylose-excess conditions marked the start of P2. The limitation led to a decrease in μ (average value 0.020 ± 0.002 h^{-1}) and r_{XYL} (average value 0.41 ± 0.06 mmol/g_{DCW}.h). Compared to P1, the carotenoid content did not show a significant difference; however, r_{CAR} decreased four-fold compared with P1 (**Supplementary Table S1**). As expected, lipid accumulation doubled compared to P1 due to the positive influence of nitrogen limitation, reaching 0.38 ± 0.05 g/g_{DCW} (**Figure 2B**). P3 started after the depletion of xylose. Here, xylitol and arabitol were simultaneously consumed by the cells. The average specific growth rate was the lowest of the three growth phases (0.005 ± 0.0003 h^{-1}). The highest accumulation of carotenoids per cell mass was detected in this phase, increasing substantially to 1.87 ± 0.21 mg/g_{DCW}; however, the r_{CAR} was the lowest of all phases (**Figure 2** and **Supplementary Table S1**). The lipid content remained at the same level as that noted in P2. In addition, 50% carbon loss (undetected carbon) was observed in P3, which can be partially explained by the technical uncertainty of measurements for off-gas at the very low growth rate conditions in the P3. Total carotenoid and lipid yields on cell mass for the whole batch cultivation were 0.85 ± 0.01 and 0.33 ± 0.07 g/g_{DCW}, respectively (**Table 1**, column REF).

Understanding Intracellular Flux Patterns Among Three Growth Phases

To understand the changes in intracellular flux patterns between the observed growth phases, simulations using *R. toruloides* GEM were performed (**Supplementary Tables S2**; Tiukova et al., 2019a; Lopes et al., 2020a). GEM is a mathematical reconstruction of the metabolic network based on genome annotations and information, such as gene-protein-reaction relationships. GEM in combination with FBA allows simulation of intracellular flux patterns by using linear programming based optimization under the provided constraints and selects the most ATP-efficient pathways for satisfying the pre-set constraints. However, the model does not take into account metabolic regulation nor limitations due to the capacity or activity of enzymes, which should be taken into account when interpreting the modelling results. For the flux distribution comparison between the observed growth phases, fluxes were normalised by the total carbon uptake rate (**Supplementary Table S3**).

As to the best of our knowledge, there has not been a direct experimental validation of the arabitol isoform which is produced by *R. toruloides*, we used the GEM to predict it based on the provided stoichiometry of the reactions. Our simulations suggested L-arabitol production, as the L-arabitol production pathway regenerates one NADPH and one NADH, while D-arabitol production pathway regenerates only one NADH. NADPH supply is crucial for lipid production given that every elongation step of fatty acid (FA) synthesis requires the oxidation of two NADPH (Wasylenko et al., 2015). In

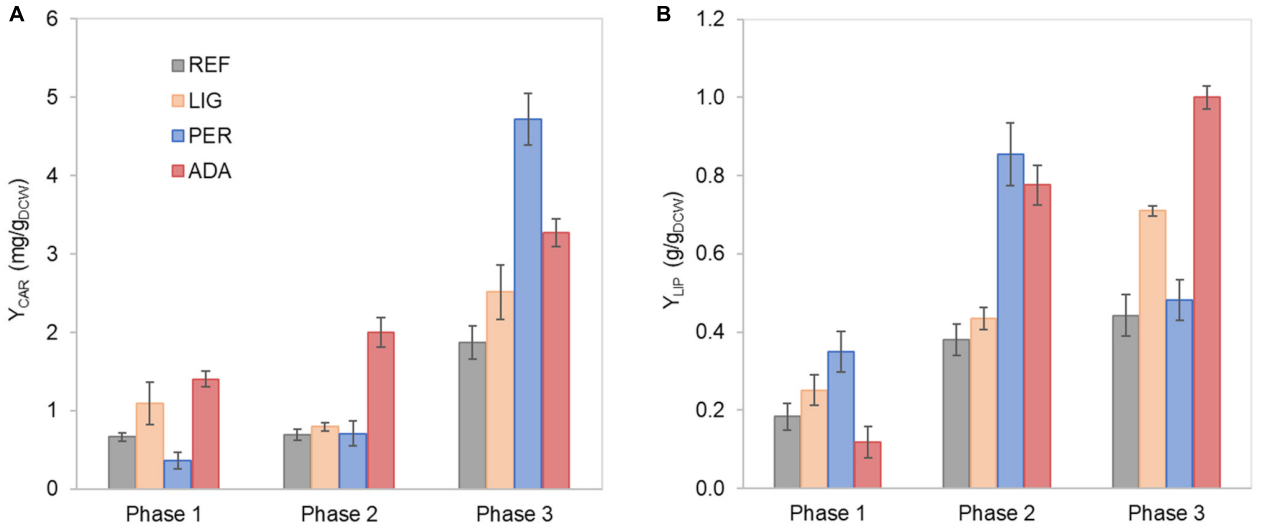


FIGURE 2 | Carotenoid **(A)** and lipid **(B)** yields on cell mass in each growth phase of *R. toruloides* parental strain under the reference condition (REF), light irradiation (LIG) and in the presence of hydrogen peroxide (PER), and the adapted strain under oxidative stress (ADA). Yields were calculated considering the production in each growth phase.

TABLE 1 | Global titres, yields on cell mass (Y) and on substrate (Y_S), and volumetric production rate (q) of carotenoids and lipid production by *R. toruloides* parental strain under the reference condition (REF), light irradiation (LIG), hydrogen peroxide stress (PER), and the adapted strain under oxidative stress (ADA).

		REF	LIG	PER	ADA
Carotenoids	Titre (mg/L)	20.7 ± 0.5	36.2 ± 2.6	19.3 ± 0.2	44.0 ± 2.4
	Y (mg/g_{DCW})	0.85 ± 0.01	1.45 ± 0.14	0.89 ± 0.04	1.90 ± 0.13
	q (mg/L.h)	0.12 ± 0.003	0.22 ± 0.02	0.06 ± 0.02	0.23 ± 0.01
	Y_S (mg/g_{sub})	0.29 ± 0.02	0.54 ± 0.04	0.29 ± 0.01	0.62 ± 0.04
Lipids	Titre (mg/L)	8.1 ± 0.3	11.1 ± 1.7	13.8 ± 0.9	13.3 ± 0.5
	Y (g/g_{DCW})	0.33 ± 0.07	0.46 ± 0.12	0.65 ± 0.06	0.58 ± 0.07
	q (g/L.h)	0.05 ± 0.003	0.07 ± 0.005	0.05 ± 0.002	0.07 ± 0.002
	Y_S (g/g_{sub})	0.12 ± 0.01	0.17 ± 0.03	0.21 ± 0.02	0.19 ± 0.01

These values were based on the final titres and the total fermentation time, representing the mean and standard deviation of triplicate experiments.

addition to FA synthesis, xylose and L-arabitol utilization also require NADPH. For all three growth phases, the highest NADPH usage was observed for substrate uptake: xylose reductase (XR) (r_1093) in P1 and P2 and L-xylulose reductase (t_0882) in P3 (**Supplementary Figure S1** and **Table S5**). During xylose metabolism (P1 and P2), arabitol production via L-xylulose reductase partially regenerated the oxidised NADPH. Once xylose was exhausted (P3), NADPH was required for arabitol catabolism (**Supplementary Figure S1**). Our simulations noted that the oxidative branch of the pentose phosphate pathway (PPP), namely, glucose 6-phosphate dehydrogenase (r_0466) and phosphogluconate dehydrogenase (r_0889), was responsible for 83, 87, and 96% of NADPH regeneration in P1, P2, and P3, respectively. Although NADPH demand in the substrate consumption and amino acid biosynthesis pathways decreased in P2 and P3 compared to P1, fluxes in lipid biosynthesis increased 1.4-fold.

At the xylulose-5P branch point, 91% of carbon entered into the central carbon metabolism via transketolase (r_1049, r_1050) in P1 (**Figure 3B**). The remaining 9% was converted into glyceraldehyde-3-phosphate and acetyl-phosphate by the phosphoketolase reaction (t_0081). Under nitrogen limitation, the activity of phosphoketolase was approximately tripled compared to P1. The phosphoketolase pathway further generates acetyl-CoA, a precursor of FA synthesis, by phosphate transacetylase (t_0082) without losing a carbon compared to the pathway originating from pyruvate.

At the pyruvate branch point, on average, ca. 70% of the pyruvate produced in the cytosol was transported to mitochondria (r_1138 and r_2034) to be converted by pyruvate dehydrogenase (r_0961) into acetyl-CoA, which is used in the tricarboxylic acid (TCA) cycle by citrate synthase (r_0300). The remaining cytosolic pyruvate was either converted to cytosolic oxaloacetate (pyruvate carboxylase, r_0958) or into acetyl-CoA by three enzymatic steps (pyruvate decarboxylase, r_0959; acetaldehyde dehydrogenase, r_2116; acetyl-CoA synthase, r_0112). Acetyl-CoA can also be produced from citrate by ATP-citrate lyase (ACL, y200003) or from acetyl-P by phosphate transacetylase. Under excess nitrogen conditions (P1), acetyl-CoA synthase was responsible for 60% of the flux. However,

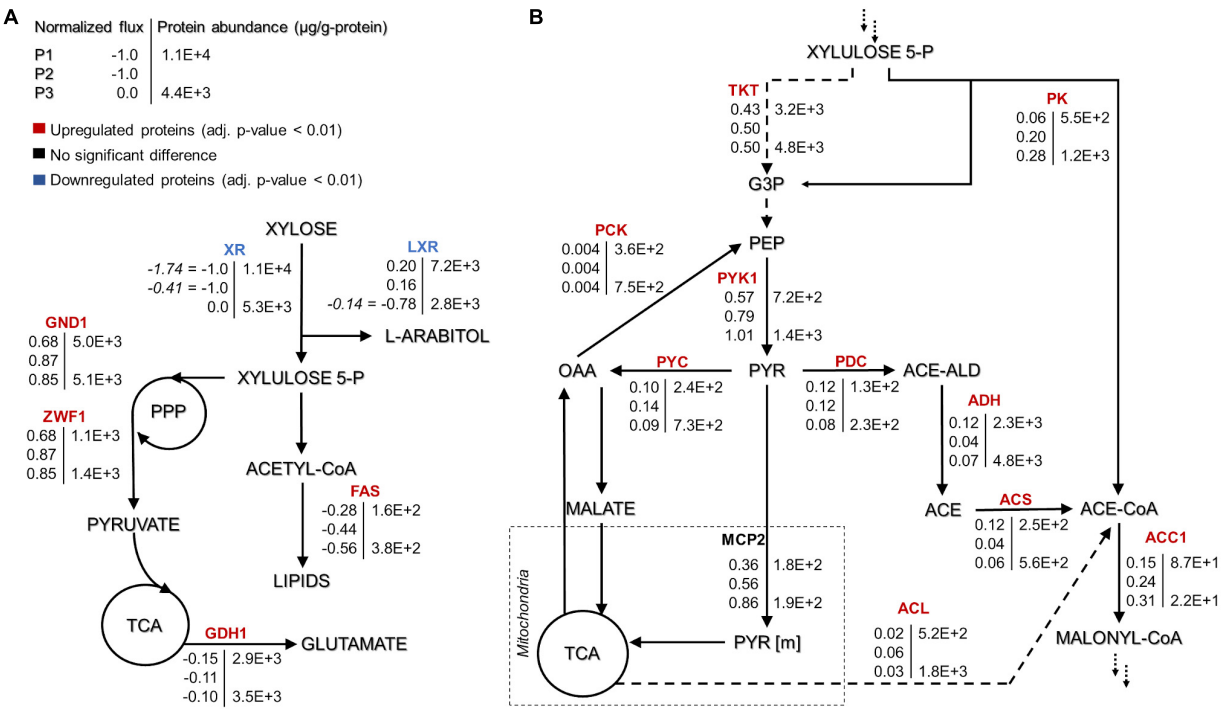


FIGURE 3 | *Rhodotorula toruloides* GEM-based fluxes (normalized to the substrate uptake) and measured protein abundances (µg/g-protein) under the three growth phases of reference cultivation on xylose (P1–P3). The main NADPH regeneration and utilization fluxes are presented with positive and negative values, respectively **(A)**. Representation of the central carbon metabolism illustrating xylulose 5-phosphate, pyruvate and acetyl-CoA branchpoints **(B)**. The fluxes were normalized by the substrate uptake rate in each phase (the absolute substrate uptake values (mmol/g_{DCW}.h) are represented in italic in panel **(A)** [note that in P3, arabitol and xylitol were co-consumed)]. The colour of enzymes name indicates the significant protein allocation change under nitrogen limitation (P3) compared to nitrogen excess condition (P1); and the dashed arrow on panel **(B)** represent multiple reactions. All the fluxes and protein abundance are represented in the **Supplementary Tables S3, S7**, respectively. G3P: glyceraldehyde 3-phosphate, PEP: phosphoenolpyruvate, PYR: pyruvate, OAA: oxaloacetate, ACE-ALD: acetaldehyde, TKT: transketolase, PK: phosphoketolase, PCK: phosphoenolpyruvate carboxykinase, PYK1: pyruvate kinase, PYC: pyruvate carboxylase, PDC: pyruvate decarboxylase, MCP2: mitochondria pyruvate carrier, ADH: alcohol dehydrogenase, ACS: acetyl-CoA synthase, ACL: ATP citrate lyase, and ACC1: acetyl-CoA carboxylase. More abbreviations can be found in the **Supplementary Table S4**.

under nitrogen limitation (P2 and P3), ca. 71% of acetyl-CoA originated via phosphate transacetylase (**Figure 3B**). Although flux variability analysis demonstrated rather high flexibility in the ACL flux, phosphate transacetylase was still predicted as the main source of acetyl-CoA under nitrogen limitation (**Supplementary Table S2**).

Our physiological data showed that the lipid content was higher under nitrogen-limitation phases (P2 and P3), and the carotenoids content was higher in P3. These results can be explained by the higher predicted fluxes through reactions involving phosphoketolase, FA and carotenoids synthesis, and NADPH regeneration.

Impact of Oxidative Stress via Light Irradiation or the Presence of H_2O_2

Further, we were interested in how oxidative stress created by either 20 mmol/L H_2O_2 or light irradiation (40,000 lux white light) affects cellular growth and lipid and carotenoid accumulation. Cultivation of *R. toruloides* under these oxidative stress conditions presented the same three growth phases described for the reference condition (REF) and a similar growth and substrate consumption profile (**Supplementary Figures S2A,B**). The specific growth rate differences compared to the reference condition were insignificant under the light irradiation (LIG) condition, but a significant 50% decrease was observed under the H_2O_2 stress (PER) in P1 (**Supplementary Table S1**). The most significant difference was detected in the longer lag phase (approximately 90 h) shown in the PER condition (**Supplementary Figure S2B**). Moreover, the accumulation profiles of carotenoids and lipids showed altered behaviour compared to the reference (**Figure 2**). The stress caused by H_2O_2 negatively affected carotenoid production in P1. However, in P3, the total carotenoid content was the highest of all conditions (4.72 ± 0.47 mg/g_{DCW}); thus, the highest r_{CAR} (0.026 ± 0.012 mg/g_{DCW}.h) was reached in the third phase (**Figure 2A**, **Supplementary Table S1**). Lipid production exhibited a different behaviour, presenting a higher content in P1 and P2 compared to the reference condition. The specific lipid production rate (r_{LIP}) in P2 was the highest of all conditions in all phases (0.020 ± 0.004 g/g_{DCW}.h), albeit the specific growth rate was not amongst the highest obtained in this study.

Regarding the overall results, for the entire cultivation under light irradiation, the cells presented 70% increased carotenoid

content (1.45 ± 0.14 mg/g_{DCW}) and 40% increased lipid content (0.46 ± 0.12 g/g_{DCW}) compared to the reference condition. H_2O_2 stress in the parental strain (PER) did not affect carotenoid production (compared to REF), and the production yield on cell mass was maintained at 0.89 ± 0.04 mg/g_{DCW}. Surprisingly, this condition showed the highest lipid content (0.65 ± 0.06 g/g_{DCW}), which was increased by two-fold compared with the reference condition (**Table 1**). The achieved lipid content was only slightly lower than the highest lipid content reported in the literature for *R. toruloides*; specifically, 0.68 g/g_{DCW} has been reported in fed-batch cultivation on a rich, glucose-based medium (Li et al., 2007).

Increasing Carotenoid Accumulation via Adaptive Laboratory Evolution

Adaptive laboratory evolution is a strategy to improve the fitness of the microorganism in a challenging environment. H_2O_2 stress improved lipid accumulation but increased the lag phase and lowered the growth rate. Therefore, ALE was performed to improve those parameters. The first cycle of ALE started with 10 mmol/L of H_2O_2. After 16 passages (ca. 30 generations), the lag phase decreased 11-fold (from 46 to 4 h), and μ_{max} stabilised at 0.045 ± 0.003 h^{-1}, resulting in a 66% increase compared to the parental strain under the same conditions (**Supplementary Figure S3A**). Once the μ_{max} plateaued, the second cycle of ALE was started by increasing the selective pressure to 20 mmol/L of H_2O_2 in the medium (**Supplementary Figure S3B**). As a result, after 15 passages (ca. 20 generations), the lag phase of yeast growth decreased from 30 to 5 h, and the μ_{max} improved to 0.055 ± 0.001 h^{-1}, representing a 22% increase compared to the first cycle. Although μ_{max} did not improve remarkably during the second cycle of ALE, the cells presented stronger pink colouration compared to the parental strain, indicating increased carotenoid accumulation. Therefore, the ALE experiment was halted, and the strains were characterised.

The adapted strain in presence of H_2O_2 (20 mmol/L) showed a remarkable ca. three-fold increase in carotenoid content and titre, respectively, compared to the parental strain under the same condition during the initial screening experiments (**Supplementary Table S6**). Therefore, the adapted strain was further characterised under controlled environmental conditions in bioreactors and studied on the proteomics level.

Adapted Strain Under H_2O_2 Stress

The performance of adapted strain under oxidative stress (by the presence of 20 mmol/L H_2O_2, ADA) was evaluated and compared to the parental strain under the same stress condition (PER). ADA exhibited a 70-h shorter lag phase compared with the PER (**Supplementary Figure S2C**). Aeration and agitation in the bioreactor may increase the oxidative stress, which could explain the longer lag phase compared to initial screening experiments in shake flasks (mentioned above).

Some differences were noticed between the ADA and other conditions. Although the μ in P1 did not show a significant difference, the r_{CAR} was increased by four-fold in ADA compared with PER (**Supplementary Table S1**). However, in P2, the μ in ADA was two-fold reduced compared with the other conditions, but r_{CAR} remained 20% increased. The lower r_{XYL} in P2 could have resulted the lower production of xylitol and arabitol in ADA compared to the other conditions probably due to a softer redox imbalance during xylose catabolism. The ADA showed ca. two-fold increased content of carotenoids and lipids under nitrogen-limiting phases (P2 and P3) compared to REF. In P3, the lipid content was 1.0 ± 0.03 g/g_{DCW}, indicating that the gain of cell mass noted during this phase was mainly related to lipid accumulation.

The whole batch growth of ADA exhibited a 2.3-fold increase in carotenoid yield on cell mass compared to PER; however, the lipid yield did not show a significant difference (**Table 1**).

Composition of Carotenoids in Cell Mass

Rhodotorula toruloides mainly produces four carotenoids: γ-carotene, β-carotene, torulene and torularhodin (Mata-Gómez et al., 2014). The carotenoid profile was very similar in all the studied conditions. The β-carotene fraction decreased over time, whereas the opposite was observed for torulene (**Supplementary Figure S3**). Torularhodin was the most abundant fraction of carotenoids under all studied conditions. Growth of the parental strain under light irradiation and adapted strain cultivations showed torularhodin fraction higher than 50% during P2, which can be related to a stronger antioxidative property of this carotenoid, attributed to the presence of more double bonds in its chemical structure (Kot et al., 2018).

Proteomic Results Revealed the Highest Difference Between Nitrogen Excess and – Limiting Conditions

Total protein measurements were combined with the absolute proteome analysis for the most relevant conditions. Therefore, samples from three different growth phases (P1, P3, and SP- an early stationary phase) from the reference cultivation were analysed, illustrating conditions under the nitrogen excess and two samples from nitrogen limitation conditions. Additionally, P3 of light-induced oxidative stress (LIG P3) and P1 and P3 of H_2O_2-induced oxidative stress for the adapted strain (ADA P1 and P3, respectively). For LIG condition, P3 represents the nitrogen-limitation phase with a higher carotenoid titre and specific production rate than REF P3. In ADA P1, the adapted strain presented a higher carotenoid yield on cell mass (Y_{CAR}) compared to parental strain under reference condition (REF P1). While P3 (ADA P3) showed the highest lipid content (g/g_{DCW}) compared to all conditions, indicating that during the phase, almost all carbon taken up by the cells ended in the lipid composition (**Figure 2B**).

Almost doubled total protein content was measured for the nitrogen-excess condition during the P1 of a reference culture. All other conditions showed no significant differences in protein content with an average of 0.24 g/g_{DCW} (**Figure 4A**). In differential expression analysis, proteome data were normalised

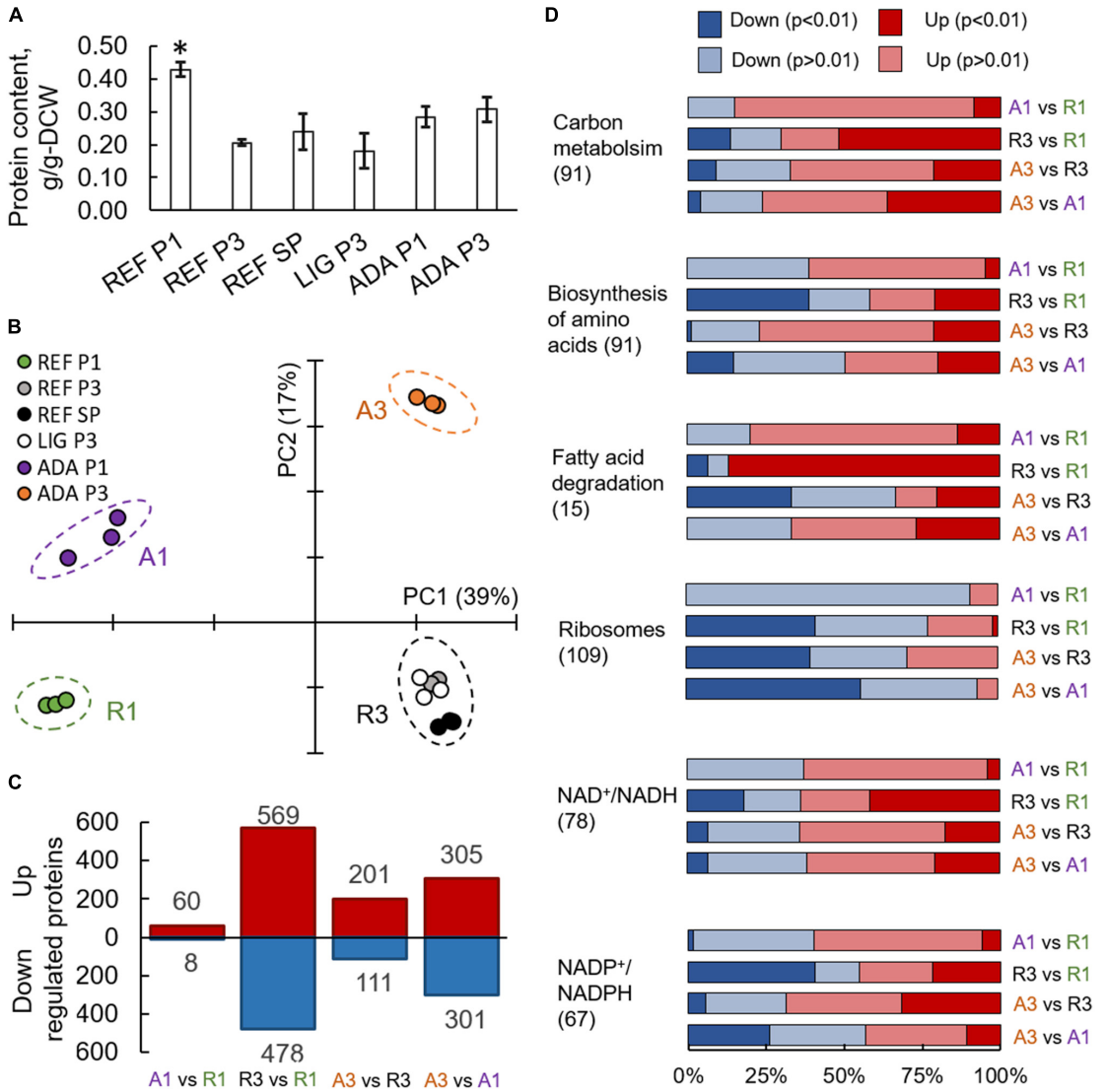


FIGURE 4 | Proteomics results from *R. toruloides* studies on xylose under different environmental conditions (REF P1, P3, and SP; LIG P3; ADA P1 and P3). Total protein content in cells **(A)**. Asterisk "*" indicates significant difference in comparison to all the other studied conditions (Student's *t*-test *p*-value < 0.01). Principal component analysis **(B)**. Significantly (adj. *p*-value < 0.01) up- and down-regulated proteins **(C)**. Gene set enrichment analysis based on the proteomics data, where GO groups were received from UniProt, enzyme-metabolite interactions and subgroups from the rhto-GEM **(D)**. All previewed categories show significant difference at least in two comparisons (adj. *p*-value < 0.01). Number in brackets indicates proteins in each category. Panels **(C,D)** present comparisons between samples showing clear separation in PCA represented in panel **(B)**.

to a constant protein mass, representing allocation differences for the individual proteins.

On average, more than 3,000 individual proteins were quantified under every condition studied (**Supplementary Table S7**). Principal component (PC) analysis clearly identified the biggest differences in the data set, which were determined by the switch into nitrogen limitation as indicated by the clear separation of the samples on the first PC, characterising 39% of the changes (**Figure 4B**). The second PC separated samples based on the use of the adapted strain under the oxidative stress environment (17% of the difference in the data). Altogether, 1,518 proteins showed significant (adj. *p*-value < 0.01) allocation changes under at least one of the environmental conditions (**Figure 4C**).

Translation and NADPH Metabolism Were Most Affected Under Nitrogen Limitation

To understand the main differences in the dataset, GSA was used to identify classes of proteins that are significantly over-represented among the measured proteins and may have an association with a specific phenotype. A variation of GSA-based analysis was conducted. First, protein-Gene Ontology

(GO) group relations were received from UniProt database. Second, protein-subsystem relationships were obtained from rhto-GEM, providing more specific information on various metabolic pathways present in *R. toruloides*. Given that PCA divided samples into four separate quadrants based on nitrogen availability and oxidative stress, we focussed on the comparison of these sample clusters throughout the study. Using UniProt-provided GO groups in GSA, 14 groups exhibited significant over-representation with an adj. p-value < 0.01 under the studied conditions (**Supplementary Table S8**). Most of these groups were related to protein translation, which were down-regulated under nitrogen limitation and correlated with the lower specific growth rates under these conditions (**Figure 4D**). A clear correlation between ribosome abundances and specific growth rate has been demonstrated previously for other organisms (Scott et al., 2014; Metzl-raz et al., 2017). When subsystems from rhto-GEM were considered, significant up-regulation was detected among carbon metabolism and its subgroups (glycolysis, gluconeogenesis, TCA cycle, glyoxylate and dicarboxylate metabolism; **Supplementary Table S9**). Interestingly, only the parental strain (under nitrogen-limiting conditions) showed overexpression in fatty acid degradation pathways and down-regulation in amino acid biosynthesis pathways (with the exception of the tryptophan pathway, which was up-regulated). Differences in the regulation of fatty acid degradation pathways could be responsible for the significantly increased lipid accumulation under the oxidative stress condition.

Based on the enzyme-metabolite relationships present in the rhto-GEM, reporter metabolites were analysed as the third variation of GSA, illustrating metabolites showing significant alterations among enzymes they interact with (**Supplementary Table S10**). When samples under nitrogen-limitation were compared to samples cultured under excess nitrogen, the most significant up-regulation was detected among proteins in proximity to NAD^+/NADH (**Figure 4D**). More than 65% of the proteins associated with NAD^+/NADH showed increased allocation under nitrogen limitation (REF P3 and SP, LIG P3, ADA P3). In contrast, protein allocation decreased significantly for proteins related to $NADP^+$/NADPH metabolism. Down-regulation was predominantly noted in amino acid biosynthesis pathways, while NADPH consumption in lipid metabolism and the glutamate production pathway showed up-regulation. Up-regulation of enzymes in glutamate biosynthesis in response to nitrogen starvation have been demonstrated previously (Zhu et al., 2012; Tiukova et al., 2019b). Additionally, metabolites related to lipid synthesis (CoA, acetyl-CoA, acetate, and pyruvate) showed increased protein allocation during nitrogen limitation, while xylitol related proteins were down-regulated (in all cases adj. p-value < 0.01).

Protein Changes Demonstrate Similar Trends With the Simulated Fluxes

The global comparison of differentially expressed proteins with their corresponding fluxes between the nitrogen excess and limitation conditions demonstrated similar trends for the 62% of cases (**Supplementary Table S11**). In response to nitrogen limitation, proteins involved in central nitrogen metabolism, such as glutamate dehydrogenase (GDH, RHTO_04650, RHTO_07718) and glutamine synthetase (GLN, RHTO_00673, RHTO_00401) were up-regulated (**Figure 3A**). This response has been previously reported under nitrogen limitation for *R. toruloides* grown in both glucose and xylose (Zhu et al., 2012; Tiukova et al., 2019b). Additionally, activation of autophagy process has been described as a direct response via TOR activation to recycle nitrogenous compounds (Zhu et al., 2012; Tiukova et al., 2019b). Although initially expressed at a low level, up-regulation of autophagy-related proteins (RHTO_05541, RHTO_06526) was detected.

Proteins related to oxidative stress response showed up-regulation for P3 and SP (nitrogen limitation) compared to P1 (nitrogen excess). Catalase (CAT, RHTO_01370), which breaks down hydrogen peroxide in the peroxisomal matrix, was the most up-regulated protein with a six-fold increase under nitrogen limitation (**Supplementary Table S7**). Recent reports in oleaginous microorganisms showed that ROS is an important signalling molecules in response to various stresses (Shi et al., 2017). Nitrogen depletion is an example of such stress, leading to the accumulation of ROS (Liu et al., 2012; Fan et al., 2014; Chokshi et al., 2017) and higher activities of catalase and other antioxidant enzymes, suggesting that lipid accumulation under nitrogen depletion is mediated by oxidative stress (Yilancioglu et al., 2014).

The highest carbon fluxes detected with GEM analysis were further assessed to understand the level of their regulation. In P3 and SP, there is no longer xylose in the system, which was reflected in the down-regulation of proteins involved in xylose assimilation (**Supplementary Table S1**). Proteins belonging to arabitol metabolism, such as L-xylulose reductase (LXR, RHTO_00373), and D-arabitol dehydrogenase (DAD, RHTO_07844), and L-iditol 2-dehydrogenase, identified by Bommareddy et al. (2015) as L-arabitol dehydrogenase (LAD, RHTO_01629) were all down-regulated under nitrogen- limiting conditions.

Approximately four-fold increased transketolase (TKT, RHTO_03248) abundance compared to phosphoketolase (PK, RHTO_04463) was consistent with the simulated increased flux through the transketolase reaction. However, PK levels increased more than two-fold under nitrogen limitation, which was consistent with the increased flux levels under the mentioned conditions. Furthermore, the magnitude of the PK increase in this condition was 50–60% higher than TKT (**Figure 3B**). Carbon was mainly channelled via TKT because it leans toward glycolysis and the oxidative branch of the PPP, which have been identified as the preferred pathway to regenerate NADPH (Lopes et al., 2020a, this study). GEM simulations revealed that XR, fatty acid synthase and glutamate dehydrogenase consume most of the NADPH, which was regenerated by 6-phosphogluconate dehydrogenase (GND1), glucose-6-phosphate dehydrogenase (ZWF1), and L-xylulose reductase (LXR) (**Figure 3A**). With the exception of fatty acid synthase, these enzymes were also among the most abundant NADPH-dependent enzymes quantified.

In the oleaginous microorganism, cytosolic ATP-citrate lyase (ACL, RHTO_03915) is considered an important enzyme

as a source of acetyl-CoA (Ratledge and Wynn, 2002; Koutinas and Papanikolaou, 2011). Under nitrogen limitation, this enzyme was found in higher levels compared to P1 (four- and two-fold increase for P3 and SP, respectively, **Supplementary Table S7**). Cytosolic acetyl-CoA can also be supplied by acetyl-CoA synthase (ACS1, RHTO_08027), which was up-regulated two-fold in nitrogen-limiting phases, and by xylose metabolism via PK and phosphate acetyltransferase. FBA predicted that the majority of cytosolic acetyl-CoA originating from pyruvate via ACS1 (in nitrogen-excess phase) and from PK and phosphate transacetylase (in nitrogen-limiting phase) (**Figure 3B**).

Acetyl-CoA can be transformed into malonyl-CoA, a substrate for FA synthesis, by the enzyme acetyl-CoA carboxylase (ACC1, RHTO_02004), which was 3.4- and 1.6-fold up-regulated in P2 and P3, respectively. Then, FAs are produced by fatty acid synthases FAS1 (RHTO_02032) and FAS2 (RHTO_02139), which were significantly up-regulated in the nitrogen-limiting phases. Such findings were in accordance with the predicted higher fluxes from acetyl-CoA toward FA synthesis in P3. Additionally, acetyl-CoA can enter the mevalonate pathway (MEV) through acetyl-CoA C-acetyltransferase (ERG10, RHTO_02048) to produce sterols and carotenoids. Despite the highest carotenoid content under nitrogen limitation, ERG10 and other proteins related to the MEV pathway, hydroxymethylglutaryl-CoA synthase (ERG13, RHTO_02305) and reductase (HMG1, RHTO_04045) were down-regulated. Proteins directly related to the carotenogenesis pathway were either not detected or did not show any significant difference. FBA predicted very low fluxes throughout carotenoid production.

Comparative Proteomics at Different Growth Phases Under Oxidative Stresses

Although P3 of cultivation under light exposure (LIG_P3) showed higher carotenoid production rates than REF_P3 (**Figure 2** and **Supplementary Table S1**), only two proteins (RHTO_06480 and RHTO_01160) were differentially expressed between those conditions (adj. p-value < 0.01), suggesting a post-translational regulation of the carotenoid production pathways under the studied environmental conditions.

Under oxidative stress, the adapted strain showed many similar changes to the reference condition while entering into nitrogen limitation but was also clearly differentiated as described by PC analysis (**Figure 4B**). Clustering of the significantly differentially expressed proteins was performed to understand the main changes compared to REF. Cluster showing significant differences among ADA and other conditions was enriched by the proteins from amino acid biosynthesis and the TCA cycle. Although TCA cycle was already up-regulated under nitrogen-limiting conditions, more pronounced up-regulation was detected in ADA, including the up-regulation of citrate synthase (CIT1, RHTO_06406), malate dehydrogenase (MDH, RHTO_04363), and $NADP^+$-dependent isocitrate dehydrogenase (IDH, RHTO_04315). NADPH regeneration by GND1 (RHTO_ 02788) in PPP was also up-regulated. Similar to the REF under nitrogen limitation, FA biosynthesis was up-regulated; however, β-oxidation, which is responsible for lipid degradation, remained lower and could at least partially explain the higher lipid accumulation under oxidative stress. Additionally, the carotenogenesis pathway was more activated since phytoene dehydrogenase (CRTI, RHTO_04602) were three-fold up-regulated (**Figure 5A**).

As expected, enzymes involved in the oxidative stress response were up-regulated in the ADA compared to REF; CAT exhibits a 16-fold increase for P1 (**Supplementary Table S7**).

Translation Processes Also Play a Crucial Role in Proteome Allocation Analysis

Absolute protein analysis allows comparisons of protein abundance levels between conditions and quantification of condition-dependent protein allocation patterns (**Figure 5B**). The top 100 of the most abundant proteins represented greater than 50% of the total proteome. Ribosomal proteins were the largest protein group, forming almost one-third of the total proteome in REF P1. However, these levels decreased to approximately 10% under the lower growth rate conditions in P3. For ADA, ribosomal protein allocation was already reduced to a lower level and decreased further, forming only 21% under the nutrient-limiting conditions in P3. As ribosomes are essential for achieving faster cell growth, the trade-off between allocation toward ribosomal proteins or energy generation pathways has been demonstrated previously (Nilsson and Nielsen, 2016; Sánchez et al., 2017; Kumar and Lahtvee, 2020). Interestingly, glycolysis was increased to the same extent under nitrogen-limiting conditions in the presence and absence of oxidative stress, but mitochondria and amino acid metabolism demonstrated significantly increased allocation for the ADA. The latter changes seemed to be responsible for the more efficient metabolism, providing higher yields on cell mass for adapted cells (**Figure 5B** and **Table 1**).

DISCUSSION

Efficient microbial production of chemicals from sustainable resources is essential for the transition toward bioeconomy. *R. toruloides* has been considered as a potential microorganism to produce high-value products from biological resources, including hemicellulosic material, mainly composed of xylose. However, xylose metabolism in *R. toruloides* is still not completely understood. Only few studies have focussed on the metabolism of xylose assimilation in this oleaginous yeast (Jagtap and Rao, 2018; Tiukova et al., 2019b; Lopes et al., 2020a). Therefore, in our study, detailed physiology characterisation was combined with genome-scale modelling and quantitative proteomics with the goal to investigate xylose metabolism in *R. toruloides* and use oxidative stress as a strategy to improve the production of lipids and carotenoids.

In this study, we demonstrated how *R. toruloides* growth on xylose exhibited three distinct phases, where most metabolic changes occurred after the transition into nitrogen limitation. Approximately 30% of consumed xylose by the parental strain accumulated into xylitol and arabitol during the first two growth

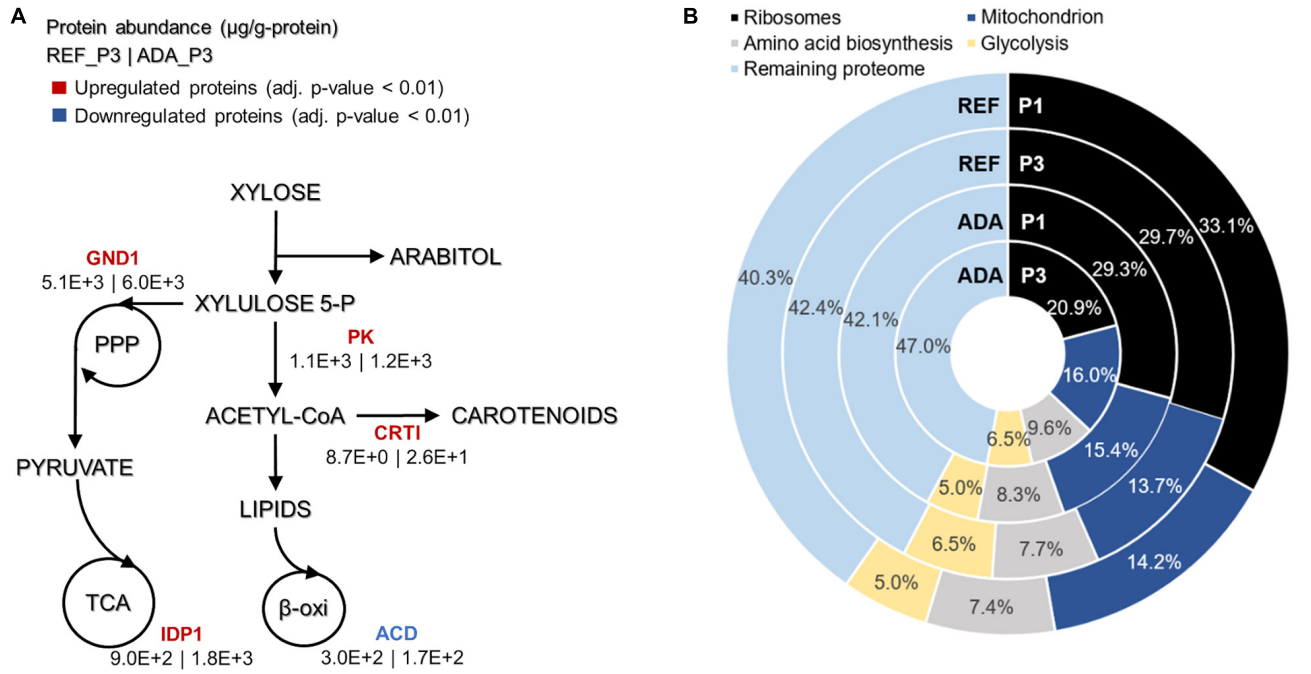


FIGURE 5 | Main differences in protein allocation between the adapted strain under oxidative stress (ADA) and the parental strain under reference condition (REF) studied under the nitrogen limitation (P3). The colour of enzymes name indicates the change in protein allocation in ADA compared to REF **(A)**. Proteome allocation into the most abundant metabolic groups for ADA and REF in P1 (nitrogen excess) and P3 (nitrogen limitation) **(B)**. GND1, 6-phosphogluconate dehydrogenase; IDP1, isocitrate dehydrogenase; PK, phophoketolase; CRTI, phytoene dehydrogenase, ACD, acyl-CoA dehydrogenase.

phases probably to balance NADPH required for the growth or due to limitations in the abundance of xylulokinase. According to Fernandes and Murray (2010), fungi can produce both L- and D-arabitol. Jagtap and Rao (2018) reported that D-xylulose is converted to D-arabitol by DAD, although the methodology employed by the authors is not able to discriminated between the L and D isoforms. Same for the methodology employed in this work. The UniProt protein database for *R. toruloides* contains only sequences for LXR and for DAD, lacking the levogyrous version of the latter. Bommareddy et al. (2015) assembled the metabolic network for *R. toruloides* based in previous omics studies and identified that the enzyme LAD, identified as L-iditol dehydrogenase by UniProt, is responsible for arabitol production in this species. The GEM constructed by Tiukova et al. (2019b) contains only the L-arabitol production pathway. In this work, we updated the model including also the pathway described in Jagtap and Rao (2018) and our simulations showed that the L-isoform was preferred. Forcing the model for the production of only D-arabitol did not reflect our experimental data, unless the cofactor for DAD was NADPH instead of NADH. An *in silico* analysis of the amino acids sequences of the LXR, LAD, and DAD by the NCBI conserved domain search tool[4] showed that those enzymes could use both or either NADH and NADPH. Our proteomics analysis was able to detect enzymes for both pathways, LXR, LAD, and DAD (**Supplementary Figure S1**). Seiboth et al. (2003) reported that the LAD can partially compensate for the xylose metabolism in mutants knocked out for xylitol dehydrogenase in *Hypocrea jecorina*. Curiously, xylulokinase, an enzyme responsible for the last step in xylose metabolism, was not detected in our proteomic analysis, suggesting a low expression level that could limit xylose catabolism. The available knowledge on the xylose metabolism in *R. toruloides* is poor and further studies aiming at revealing the involved enzymes along with its co-factors and reactions are needed.

Under nitrogen-limiting conditions, increased flux via phosphoketolase was observed, corresponding well with the detected enzyme up-regulation. The phosphoketolase reaction saves carbon by producing an acetyl residue together with G3P directly from D-xylulose 5-P instead of requiring several reactions until pyruvate decarboxylation, potentially resulting in an increased product yield from substrate. However, cells seem to prefer the alternative pathway of transketolase as it provides more carbon toward oxidative PPP, where the majority of NADPH is regenerated. The overexpression of phosphoketolase and phosphotransacetylase enzymes in *Yarrowia lipolytica* Po1d resulted in 20–50% increased lipid production (Niehus et al., 2018). Therefore, the up-regulation of phosphoketolase could be partially related to the increased accumulation of lipids and carotenoids at the aforementioned phases.

NADPH regeneration is an important mechanism for xylose assimilation and lipid synthesis. ME is considered a key enzyme in the recycling of NADPH for lipid biosynthesis (Ratledge and Wynn, 2002). However, according to our proteomic data and

[4]https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

simulations, the majority of the reducing power is generated in the oxidative branch of the PPP. These results are consistent with previous 13C-labelling experiments with *Y. lipolytica* performed using glucose as the carbon source (Wasylenko et al., 2015). However, ME overexpression in *Rhodotorula glutinis* has led to a two-fold increase in lipid accumulation (Li et al., 2013).

Induction of oxidative stress in *R. toruloides* cultivation was employed as a strategy to identify the metabolic changes that result in the higher levels of carotenoids and lipids. Surprisingly, under PER conditions, no changes in carotenoid production compared to REF were observed. However, the lipid content was two-fold increased, reaching a content of 0.65 g/g_{DCW}. The highest reported lipid production in *R. toruloides* is 0.675 g/g_{DCW} (Li et al., 2007) using a rich medium with glucose as a carbon source. The review by Shi et al. (2017) reports that different authors found evidence for ROS stress enhancing the formation of lipid droplets as an important signalling molecule in response to nitrogen starvation in oleaginous microorganisms. Cells under oxidative stress might possess an even greater supply of NADPH by channelling sugar catabolism to PPP to stabilise the redox balance and ROS clearance (Kuehne et al., 2015). Enzymes related to oxidative stress were up-regulated not only in ADA but also in REF under nitrogen limitation. Tiukova et al. (2019b) reported not only the up-regulation of oxidative stress related enzymes during the lipid accumulation phases but also the up-regulation of the β-oxidation representing an ATP sink. According to Xu et al. (2017), lipid oxidation leads to accumulation of oxidative and aldehyde species in *Y. lipolytica*, reducing production performance. The same authors showed that overexpressing enzymes from the oxidative stress defence pathway resulted in industrially relevant lipid production (lipid titre of 72.7 g/L and content of 81.4%) by synchronising lipogenesis to cell growth and mitigating lipotoxicity.

The successive application of H_2O_2 in the cell through the ALE potentially overwhelmed the cellular antioxidant defence system, boosting carotenoid production. Increased accumulation of carotenoids and lipids was noted in all the phases of ADA compared to REF. In general, improved production can result from the up-regulation of phosphoketolase, which led to a slightly more efficient production process. The carotenoid biosynthesis pathway was more activated by higher levels of CRTI what could explain the higher carotenoid levels. IDP1 and GND1 up-regulation improved the capacity of NADPH regeneration. The up-regulation of the oxidative stress defence pathway potentially diminished lipotoxicity. All the aforementioned factors combined with the down-regulation of the β-oxidation could explain the higher lipid production by the adapted cells under oxidative stress (**Figure 5A**). No reasonable significant modifications in proteomics levels were observed under the light condition in contrast to the findings of Gong et al. (2019) for *R. glutinis*. However, irradiation increased final titres of carotenoids and lipids (75 and 40% improvement, respectively) likely due to post-translational mechanisms. Some fungi, such as *Mucor circinelloides*, harbour genes for carotenoid synthesis that are based on light-induced expression mechanisms (Quiles-Rosillo et al., 2005).

CONCLUSION

In this study, the detailed physiological characterisation of *R. toruloides* growth revealed that a considerable amount of xylose was converted into by-products, such as arabitol and xylitol. The accumulation of these by-products can be considered an overflow metabolism, contributing to the redox balancing during xylose catabolism. According to the model simulations, which does not take into account metabolic regulation or enzyme capacity constraints, the highest NADPH demand was related to substrate uptake and was about five-fold higher compared with the NADPH levels required for lipid production. The main NADPH regeneration reactions were derived from the oxidative branch of PPP, and ME was underused.

Under nitrogen limitation, the parental strain showed some increased protein expression related to lipid synthesis (FAS complex) and oxidative stress response mechanisms. Interestingly, the adapted strain down-regulated the FA degradation pathway, which combined with the up-regulation of NADPH regeneration mechanisms and FA and carotenoid synthesis, led to better production performance. Additionally, this strain demonstrated increased allocation of proteins related to mitochondria and amino acid metabolism, potentially explaining the more efficient metabolism.

In general, a good correlation was noted between the predicted fluxes and determined protein abundances. Using data obtained in this study, we can design strategies of metabolic engineering to make the process economically viable by improving cell factory performance.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MP, NB, EM, and P-JL designed the experiments. MP performed the experiments. IB performed the metabolic flux simulations. MP, NB, IB, and P-JL analysed the data. All authors wrote and revised the manuscript.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 668997 and the Estonian Research Council (grant PUT1488P). MP would additionally like to acknowledge Coordination for the Improvement of Higher Education Personnel (Capes), São Paulo Research Foundation (FAPESP, grant 2016/10636-8) and DORA Plus.

ACKNOWLEDGMENTS

We thank the Proteomics Core Laboratory at University of Tartu for the proteome quantification. A preprint (Pinheiro et al., 2020) has been deposited at bioRxiv doi: 10.1101/2020.05.28.121012.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.01008/full#supplementary-material

REFERENCES

Adrio, J. L. (2017). Oleaginous yeasts: promising platforms for the production of oleochemicals and biofuels. *Biotechnol. Bioeng.* 114, 1915–1920. doi: 10.1002/bit.26337

BBC Research (2018). *The Global Market for Carotenoids*. Available online at: https://www.globenewswire.com/news-release/2018/05/21/1509412/0/en/Global-Carotenoids-Market-to-Reach-2-0-Billion-by-2022.html (accessed 20th February 2020).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a pratical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57, 298–300.

Blomqvist, J., Pickova, J., Tilami, S. K., and Sabine, S. (2018). Oleaginous yeast as a component in fish feed. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-34232-x

Bommareddy, R. R., Sabra, W., Maheshwari, G., and Zeng, A.-P. (2015). Metabolic network analysis and experimental study of lipid production in *Rhodosporidium toruloides* grown on single and mixed substrates. *Microb. Cell Fact.* 14:36. doi: 10.1186/s12934-015-0217-215

Bonturi, N., Crucello, A., Viana, A. J. C., and Miranda, E. A. (2017). Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a *Rhodosporidium toruloides* adapted strain. *Process Biochem.* 57, 16–25. doi: 10.1016/j.procbio.2017.03.007

Bonturi, N., Matsakas, L., Nilsson, R., Christakopoulos, P., Miranda, E. A., and Berglund, K. A. (2015). Single cell oil producing yeasts Lipomyces starkeyi and *Rhodosporidium toruloides*: selection of extraction strategies and biodiesel property prediction. *Energies* 8, 5040–5052. doi: 10.3390/en8065040

Castañeda, M. T., Nuñez, S., Garelli, F., Voget, C., and De Battista, H. (2018). Comprehensive analysis of a metabolic model for lipid production in *Rhodosporidium toruloides*. *J. Biotechnol.* 280, 11–18. doi: 10.1016/j.jbiotec.2018.05.010

Chokshi, K., Pancha, I., Ghosh, A., and Mishra, S. (2017). Nitrogen starvation - induced cellular crosstalk of ROS - scavenging antioxidants and phytohormone enhanced the biofuel potential of green microalga *Acutodesmus dimorphus*. *Biotechnol. Biofuels* 10:60. doi: 10.1186/s13068-017-0747-747

Coradetti, S. T., Pinel, D., Geiselman, G. M., Ito, M., Mondo, S. J., and Reilly, M. C. (2018). Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*. *eLife* 7:e32110. doi: 10.7554/eLife.32110

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechonol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Dinh, H. V., Suthers, P. F., Chan, S. H. J., Shen, Y., Xiao, T., and Deewan, A. (2019). A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metab. Eng. Commun.* 9:e00101. doi: 10.1016/j.mec.2019.e00101

Du, C., Li, Y., Guo, Y., Han, M., Zhang, W., and Qian, H. (2016). The suppression of torulene and torularhodin treatment on the growth of PC-3 xenograft prostate tumors. *Biochem. Biophys. Res. Commun.* 469, 1146–1152. doi: 10.1016/j.bbrc.2015.12.112

Fan, J., Cui, Y., Wan, M., Wang, W., and Li, Y. (2014). Lipid accumulation and biosynthesis genes response of the oleaginous *Chlorella pyrenoidosa* under three nutrition stressors. *Biotechnol. Biofuels* 7, 1–14. doi: 10.1186/1754-6834-7-17

Fernandes, S., and Murray, P. (2010). Metabolic engineering for improved microbial pentose fermentation. *Bioeng Bugs.* 1, 424–428. doi: 10.4161/bbug.1.6.12724

Folch, J., Lees, M., and Sloane Stanley, G. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* 226, 497–509.

Gong, G., Liu, L., Zhang, X., and Tan, T. (2019). Multi-omics metabolism analysis on irradiation-induced oxidative stress to *Rhodotorula glutinis*. *Appl. Microbiol. Biotechnol.* 103, 361–374. doi: 10.1007/s00253-018-9448-9449

Gong, G., Zhang, X., and Tan, T. (2020). Simultaneously enhanced intracellular lipogenesis and β -carotene biosynthesis of *Rhodotorula glutinis* by light exposure with sodium acetate as the substrate. *Bioresour. Technol.* 295, 122274. doi: 10.1016/j.biortech.2019.122274

Hu, J., and Ji, L. (2016). Draft genome sequences of *Rhodosporidium toruloides* Strains ATCC 10788 and ATCC 10657 with compatible mating types. *Genome Announc.* 4:e0098-16. doi: 10.1128/genomeA.00098-16

Irazusta, V., Nieto-Peñalver, C. G., Cabral, M. E., Amoroso, M. J., and De Figueroa, L. I. C. (2013). Relationship among carotenoid production, copper bioremediation and oxidative stress in *Rhodotorula mucilaginosa* RCL-11. *Process Biochem.* 48, 803–809. doi: 10.1016/j.procbio.2013.04.006

Jagtap, S. S., and Rao, C. V. (2018). Production of D -arabitol from D -xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Appl. Microbiol. Biotechnol.* 102, 143–151. doi: 10.1007/s00253-017-8581-1

Kerkhoven, E. J., Lahtvee, P., and Nielsen, J. (2015). Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Res.* 15, 1–13. doi: 10.1111/1567-1364.12199

Kot, A. M., Błazejak, S., Gientka, I., Kieliszek, M., and Bryś, J. (2018). Torulene and torularhodin: "New" fungal carotenoids for industry. *Microb. Cell Fact.* 17, 1–14. doi: 10.1186/s12934-018-0893-z

Kot, A. M., Błażejak, S., Kieliszek, M., Gientka, I., Bryś, J., and Reczek, L. (2019). Effect of exogenous stress factors on the biosynthesis of carotenoids and lipids by *Rhodotorula yeast* strains in media containing agro-industrial waste. *World J. Microbiol. Biotechnol.* 35, 1–10. doi: 10.1007/s11274-019-2732-2738

Koutinas, A. A., and Papanikolaou, S. (2011). *Biodiesel Production from Microbial Oil*. Sawston: Woodhead Publishing Limited, 177–198. doi: 10.1533/9780857090492.2.177

Kuehne, A., Emmert, H., Lucius, R., Zamboni, N., Kuehne, A., and Emmert, H. (2015). Acute activation of oxidative pentose phosphate pathway as first-line response to oxidative stress in human skin cells. *Mol. Cell* 59, 359–371. doi: 10.1016/j.molcel.2015.06.017

Kumar, R., and Lahtvee, P. (2020). Proteome overabundance enables respiration but limitation onsets carbon overflow. *bioRxiv [Preprint]* doi: 10.1101/2020.02.20.957662

Kumar, S., Kushwaha, H., Bachhawat, A. K., Raghava, G. P. S., and Ganesan, K. (2012). Genome sequence of the oleaginous red yeast *Rhodosporidium toruloides* MTCC 457. *Eukaryot. Cell* 11, 1083–1084. doi: 10.1128/EC.00156-112

Lahtvee, P. J., Sánchez, B. J., Smialowska, A., Kasvandik, S., Elsemman, I. E., and Gatto, F. (2017). Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Syst.* 4, 495–504. doi: 10.1016/j.cels.2017.03.003

Lee, J. J. L., Chen, L., Shi, J., Trzcinski, A., and Chen, W. N. (2014). Metabolomic profiling of *Rhodosporidium toruloides* grown on glycerol for carotenoid production during different growth phases. *J. Agric. Food Chem.* 62, 10203–10209. doi: 10.1021/jf502987q

Li, Y., Zhao, Z. K., and Bai, F. (2007). High-density cultivation of oleaginous yeast *Rhodosporidium toruloides* Y4 in fed-batch culture. *Enzyme Microb. Technol.* 41, 312–317. doi: 10.1016/j.enzmictec.2007.02.008

Li, Z., Sun, H., Mo, X., Li, X., Xu, B., and Tian, P. (2013). Overexpression of malic enzyme (ME) of Mucor circinelloides improved lipid accumulation in engineered *Rhodotorula glutinis*. *Appl. Microbiol. Biotechnol.* 97, 4927–4936. doi: 10.1007/s00253-012-4571-4575

Liu, H., Zhao, X., Wang, F., Li, Y., Jiang, X., Ye, M., et al. (2009). Comparative proteomic analysis of *Rhodosporidium toruloides* during lipid accumulation. *Yeast* 26, 553–566. doi: 10.1002/yea.1706

Liu, W., Huang, Z., Li, P., Xia, J., and Chen, B. (2012). Formation of triacylglycerol in *Nitzschia closterium* f. minutissima under nitrogen limitation and possible physiological and biochemical mechanisms. *J. Exp. Mar. Bio. Ecol.* 418–419, 24–29. doi: 10.1016/j.jembe.2012.03.005

Lopes, H. J. S., Bonturi, N., Kerkhoven, E. J., Miranda, E. A., and Lahtvee, P. (2020a). C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*. *Appl. Microb. Cell Physiol.* 104, 2639–2649. doi: 10.1007/s00253-020-10386-5

Lopes, H. J. S., Bonturi, N., and Miranda, E. A. (2020b). *Rhodotorula toruloides* single cell oil production using eucalyptus urograndis hemicellulose hydrolysate as a carbon source. *Energies* 13:795. doi: 10.3390/en13040795

Mata-Gómez, L. C., Montañez, J. C., Méndez-Zavala, A., and Aguilar, C. N. (2014). Biotechnological production of carotenoids by yeasts: an overview. *Microb. Cell Fact.* 13:12. doi: 10.1186/1475-2859-13-12

Metzl-raz, E., Kafri, M., Yaakov, G., Soifer, I., and Gurvich, Y. (2017). Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife* 6:e28034. doi: 10.7554/eLife.28034

Morin, N., Calcas, X., Devillers, H., Durrens, P., Sherman, J., and Nicaud, J. (2014). Draft genome sequence of *Rhodosporidium toruloides* CECT1137, an oleaginous yeast of biotechnological interest. 2, 578–579. doi: 10.1128/genomeA.00641-614

Niehus, X., Marie, A., Le, C., Sandoval, G., Nicaud, J. M., and Amaro, R. L. (2018). Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. *Biotechnol. Biofuels* 11:11. doi: 10.1186/s13068-018-1010-1016

Nilsson, A., and Nielsen, J. (2016). Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. *Nat. Publ. Gr.* 6:22264. doi: 10.1038/srep22264

Park, Y., Nicaud, J., and Ledesma-amaro, R. (2017). The engineering potential of *Rhodosporidium toruloides* as a workhorse for biotechnological applications. *Trends Biotechnol.* 36, 304–317. doi: 10.1016/j.tibtech.2017.10.013

Pinheiro, M. J., Bonturi, N., Belouah, I., and Miranda, E. A. (2020). Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in *Rhodotorula toruloides*: insights for future biorefinery. *bioRxiv [Preprint]* doi: 10.1101/2020.05.28.121012

Quiles-Rosillo, M. D., Ruiz-Vázquez, R. M., Torres-Martínez, S., and Garre, V. (2005). Light induction of the carotenoid biosynthesis pathway in Blakeslea trispora. *Fungal Genet. Biol.* 42, 141–153. doi: 10.1016/j.fgb.2004.10.008

Ratledge, C., and Wynn, J. P. (2002). The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv. Appl. Microbiol.* 51, 1–51. doi: 10.1016/s0065-2164(02)51000-5

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, 191–198. doi: 10.1093/nar/gkz369

Sambles, C., Middelhaufe, S., Soanes, D., Kolak, D., Lux, T., and Moore, K. (2017). Genomics Data Genome sequence of the oleaginous yeast *Rhodotorula toruloides* strain. *Genomics Data* 13, 1–2. doi: 10.1016/j.gdata.2017.05.009

Sánchez, B. J., Lahtvee, P., Campbell, K., Kasvandik, S., and Yu, R. (2020). Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in *Saccharomyces cerevisiae*. *bioRxiv [Preprint]* doi: 10.1101/2020.03.23.998237

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P. A., Kerkhoven, E. J., and Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* 13:935. doi: 10.15252/msb.20167411

Scott, M., Klumpp, S., Mateescu, E. M., and Hwa, T. (2014). Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.* 10, 747. doi: 10.15252/msb.20145379

Seiboth, B., Hartl, L., Pail, M., and Kubicek, C. P. (2003). D-xylose metabolism in *Hypocrea jecorina*: loss of the xylitol dehydrogenase step can be partially compensated for by lad1-encoded L-arabinitol-4-dehydrogenase. *Eukaryot. Cell* 2, 867–875. doi: 10.1128/EC.2.5.867-875.2003

Shi, K., Gao, Z., Shi, T., Song, P., Ren, L., and Huang, H. (2017). Reactive oxygen species-mediated cellular stress response and lipid accumulation in oleaginous microorganisms: the state of the art and future perspectives. *Front. Microbiol.* 8:793. doi: 10.3389/fmicb.2017.00793

Stahl, W., and Sies, ÆH. (2007). Carotenoids and flavonoids contribute to nutritional protection against skin damage from sunlight. *Mol. Biotechnol.* 37, 26–30. doi: 10.1007/s12033-007-0051-z

Tiukova, I. A., Brandenburg, J., Blomqvist, J., Sampels, S., Mikkelsen, N., Skaugen, M., et al. (2019a). Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. *Biotechnol. Biofuels.* 12:137. doi: 10.1186/s13068-019-1478-8

Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M., and Kerkhoven, E. J. (2019b). Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol. Bioeng.* 116, 3396–3408. doi: 10.1002/bit.27162

Unrean, P., Khajeeram, S., and Champreda, V. (2017). Combining metabolic evolution and systematic fed-batch optimization for efficient single-cell oil production from sugarcane bagasse. *Renew. Energy* 111, 295–306. doi: 10.1016/j.renene.2017.04.018

Väremo, L., Nielsen, J., and Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* 41, 4378–4391. doi: 10.1093/nar/gkt111

Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkv1145

Wang, H., Marcis, S., Hermansson, D., Agren, R., Nielsen, J., and Kerkhoven, E. J. (2018). RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput. Biol.* 14:1–17. doi: 10.1371/journal.pcbi.1006541

Wang, Y., Zhang, S., Zhu, Z., Shen, H., Lin, X., and Jin, X. (2018). Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast *Rhodosporidium toruloides*. *Biotechnol. Biofuels* 11, 1–15. doi: 10.1186/s13068-018-1134-1138

Wasylenko, T. M., Ahn, W. S., and Stephanopoulos, G. (2015). The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. *Metab. Eng.* 30, 27–39. doi: 10.1016/j.ymben.2015.02.007

Weber, R. W. S., Anke, H., and Davoli, P. (2007). Simple method for the extraction and reversed-phase high-performance liquid chromatographic analysis of carotenoid pigments from red yeasts (*Basidiomycota*. Fungi). *J. Chromatogr. A* 1145, 118–122. doi: 10.1016/j.chroma.2007.01.052

Xu, P., Qiao, K., and Stephanopoulos, G. (2017). Engineering oxidative stress defense pathways to build a robust lipid production platform in *Yarrowia lipolytica*. *Biotechnol. Bioeng.* 114, 1521–1530. doi: 10.1002/bit.26285

Yang, X., Sun, W., Shen, H., Zhang, S., Jiao, X., and Zhao, Z. K. (2018). Expression of phosphotransacetylase in: *Rhodosporidium toruloides* leading to improved cell growth and lipid production. *RSC Adv.* 8, 24673–24678. doi: 10.1039/c8ra03028f

Yilancioglu, K., Cokol, M., Pastirmaci, I., Erman, B., and Cetiner, S. (2014). Oxidative stress is a mediator for increased lipid accumulation in a newly isolated *Dunaliella salina* strain. *PLoS One* 9:e91957. doi: 10.1371/journal.pone.0091957

Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., and Yang, F. (2012). A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nat. Commun.* 3:1112. doi: 10.1038/ncomms2112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 21 August 2020
doi: 10.3389/fbioe.2020.00772

Development of a Genome-Scale Metabolic Model of *Clostridium thermocellum* and Its Applications for Integration of Multi-Omics Datasets and Computational Strain Design

Sergio Garcia[1,2], *R. Adam Thompson*[2,3,4†], *Richard J. Giannone*[2,5], *Satyakam Dash*[2,6], *Costas D. Maranas*[2,6] *and Cong T. Trinh*[1,2,3,4*]

[1] Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, United States, [2] Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, United States, [3] Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, TN, United States, [4] Oak Ridge National Laboratory, Oak Ridge, TN, United States, [5] Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, [6] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, United States

Solving environmental and social challenges such as climate change requires a shift from our current non-renewable manufacturing model to a sustainable bioeconomy. To lower carbon emissions in the production of fuels and chemicals, plant biomass feedstocks can replace petroleum using microorganisms as biocatalysts. The anaerobic thermophile *Clostridium thermocellum* is a promising bacterium for bioconversion due to its capability to efficiently degrade lignocellulosic biomass. However, the complex metabolism of *C. thermocellum* is not fully understood, hindering metabolic engineering to achieve high titers, rates, and yields of targeted molecules. In this study, we developed an updated genome-scale metabolic model of *C. thermocellum* that accounts for recent metabolic findings, has improved prediction accuracy, and is standard-conformant to ensure easy reproducibility. We illustrated two applications of the developed model. We first formulated a multi-omics integration protocol and used it to understand redox metabolism and potential bottlenecks in biofuel (e.g., ethanol) production in *C. thermocellum*. Second, we used the metabolic model to design modular cells for efficient production of alcohols and esters with broad applications as flavors, fragrances, solvents, and fuels. The proposed designs not only feature intuitive push-and-pull metabolic engineering strategies, but also present novel manipulations around important central metabolic branch-points. We anticipate the developed genome-scale metabolic model will provide a useful tool for system analysis of *C. thermocellum* metabolism to fundamentally understand its physiology and guide metabolic engineering strategies to rapidly generate modular production strains for effective biosynthesis of biofuels and biochemicals from lignocellulosic biomass.

Keywords: *Clostridium thermocellum*, biofuels, genome-scale model, metabolic model, omics integration, modular cell design, ModCell

OPEN ACCESS

Edited by:
Young-Mo Kim, Pacific Northwest National Laboratory (DOE), United States

Reviewed by:
Karsten Zengler, University of California, San Diego, United States
Esteban Marcellin, The University of Queensland, Australia

***Correspondence:**
Cong T. Trinh
ctrinh@utk.edu

†Present address:
R. Adam Thompson, Quantitative Translational Pharmacology, DMPK-BA, Abbvie Inc., North Chicago, IL, United States

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *02 April 2020*
Accepted: *18 June 2020*
Published: *21 August 2020*

Citation:
Garcia S, Thompson RA, Giannone RJ, Dash S, Maranas CD and Trinh CT (2020) Development of a Genome-Scale Metabolic Model of Clostridium thermocellum and Its Applications for Integration of Multi-Omics Datasets and Computational Strain Design. Front. Bioeng. Biotechnol. 8:772. doi: 10.3389/fbioe.2020.00772

1. INTRODUCTION

Global oil reserves will be soon depleted (Shafiee and Topal, 2009), and climate change could become a major driver of civil conflict (Hsiang et al., 2011). These challenges to security and the environment need to be addressed by replacing our current non-renewable production of energy and materials for a renewable and carbon neutral approach (Ragauskas et al., 2006). The gram-positive, thermophilic, cellulolytic, strict anaerobe *C. thermocellum* is capable of efficient degradation of lignocellulosic biomass to produce biofuels and biomaterial precursors, making this organism an ideal candidate for consolidated bioprocessing (CBP), where production of lignocellulosic enzymes, saccharification, and fermentation take place in a single step (Olson et al., 2012). However, its complex and poorly understood metabolism remains the main roadblock to achieve industrially competitive titers, rates, and yields of biofuels such as ethanol (Tian et al., 2016) and isobutanol (Lin et al., 2015).

For the past decade, significant efforts have been dedicated to characterize and manipulate the central metabolism of *C. thermocellum*, due to increasing interest in developing this organism as a CBP manufacturing platform for biofuels production (Akinosho et al., 2014). *C. thermocellum* possesses atypical central metabolism, characterized by the important roles of pyrophosphate and ferredoxin (Zhou et al., 2013), which makes redirection of both carbon and electron flows for biofuel production challenging to achieve. Specifically, the metabolic network of *C. thermocellum* contains various reactions to regulate intracellular concentration levels of NADH, NADPH, and reduced ferredoxin. These cofactors are used as electron donors with high specificity throughout metabolism. To maintain redox balance, *C. thermocellum* also possesses several hydrogenases to oxidize these reduced cofactors to molecular hydrogen that is secreted by the cell. Removal of these hydrogenases through deletion of *ech* (encoding the ferredoxin-dependent hydrogenase, ECH) and *hydG* (associated with the bifurcating hydrogenase, BIF, and bidirectional hydrogenase, H2ASE_syn) was successfully applied to increase ethanol yield by electron rerouting (Biswas et al., 2015). Thompson et al. (2015) characterized the Δ*hydG*Δ*ech* strain in depth by flux analysis of its core metabolism, concluding that the major driver for ethanol production was redox rather than carbon balancing. In particular, the conversion of reduced ferredoxin to NAD(P)H is likely the most rate limiting step. In a subsequent study, Lo et al. (2017) over-expressed *rnf* (encoding the ferredoxin-NAD oxidoreductase, RNF) in the Δ*hydG*Δ*ech* strain that is expected to enhance NADH supply, but did not achieve improved ethanol yield.

In an attempt to redirect carbon and electron flows for enhanced ethanol production, Deng et al. (2013) manipulated the pyruvate node and malate shunt of *C. thermocellum*. By converting phosphoenolpyruvate (*pep*) to oxaloacetate (*oaa*) and then to pyruvate (*pyr*), this shunt can interchange one mole of NADPH with one mol of NADH generated from glycolysis. Interestingly, the authors noted that replacement of the malate shunt by alternative pathways not linked to NADPH increased ethanol production and carbon recovery but reduced amino acid formation, confirming the role of the malate shunt as an NADPH source in *C. thermocellum*.

Sulfur metabolism also plays a key role in redox metabolism of *C. thermocellum* and has been investigated for its role in ethanol production. Sulfate, a component of *C. thermocellum* media, serves as an electron acceptor, which is capable of oxidizing sulfate to sulfite and then sulfide. Thompson et al. (2015) demonstrated that the strain Δ*hydG*Δ*ech*Δ*pfl* could not grow in a conventional defined medium due to its inability to secrete hydrogen or formate, but was able to rescue growth by sulfate supplementation to the culture medium. More recently, Biswas et al. (2017) reported an increase in final sulfide concentration and over-expression of the associated sulfate uptake and reduction pathway in the Δ*hydG* strain, but did not observe a significant difference in final sulfide concentration in Δ*hydG*Δ*ech*. Remarkably, neither of the strains consumed cysteine from the medium, unlike the wild-type. Sulfide can be converted to cysteine by CYSS (cysteine synthase) or homocysteine and then methionine by SHSL2 (succinyl-homoserine succinate lyase) and METS (methionine synthase), but the connection between the cessation of cysteine uptake and sulfate metabolism remains unclear.

Overall, the complex interactions of *C. thermocellum* metabolic pathways remain challenging to understand and engineer with conventional methods, and hence require a quantitative systems biology approach to decipher. To this end, several genome-scale metabolic models (GSMs) of *C. thermocellum* have been developed. The first GSM, named iSR432, was constructed for the strain ATCC27405 and applied to identify gene deletion strategies for high ethanol yield (Roberts et al., 2010). This model was then further curated into iCth446 (Dash et al., 2017). More recently, Thompson et al. developed the iAT601 genome-scale model (Thompson et al., 2016) for the strain DSM1313, which is genetically tractable (Argyros et al., 2011). The iAT601 model was used to identify genetic manipulations for high ethanol, isobutanol, and hydrogen production (Thompson et al., 2016), and to understand growth cessation prior to substrate depletion observed under high-substrate loading fermentations that simulate industrial conditions (Thompson and Trinh, 2017). In addition to these core and genome-scale steady-state metabolic models, a kinetic model of central metabolism, k-ctherm118, was recently developed and used to elucidate the mechanisms of nitrogen limitation and ethanol stress (Dash et al., 2017). Due to the biotechnological relevance of the *Clostridium* genus, GSMs have also been developed for other species (Dash et al., 2016), including *C. acetobutylicum* (Senger and Papoutsakis, 2008; Salimi et al., 2010; McAnulty et al., 2012; Wallenius et al., 2013; Dash et al., 2014; Yoo et al., 2015; Lee and Trinh, 2019), *C. beijerinckii* (Milne et al., 2011), *C. butyricum* (Serrano-Bermúdez et al., 2017), *C. cellulolyticum* (Salimi et al., 2010), and *C. ljungdahlii* (Nagarajan et al., 2013).

In this study, we developed an updated genome-scale metabolic model of *C. thermocellum*, named iCBI655, with more comprehensive and precise metabolic coverage, enhanced prediction accuracy, and extensive documentation. This model

is a human-curated database that coherently represents all the available genetic, genomic, and metabolic knowledge of *C. thermocellum* from both experimental literature and bioinformatic predictions. Furthermore, the model can be applied not only to enable metabolic flux simulation but also to provide a framework to contextualize disparate datasets at the system level. As a demonstration for the model application, we first developed a quantitative multi-omics integration protocol and used it to fundamentally study redox metabolism and potential redox bottlenecks critical for production of biofuels (e.g., ethanol) in *C. thermocellum*. Furthermore, we used the model, in combination with the previously developed ModCell tool (Garcia and Trinh, 2019c), to design modular (chassis) cells (Garcia and Trinh, 2019b) for alcohol and ester production.

2. RESULTS

2.1. Development of an Upgraded *C. thermocellum* Genome-Scale Model Named iCBI655

The iCBI655 model was developed using the published iAT601 model (Thompson et al., 2016) as a starting point. The model improvements include updated metabolic pathways, new annotation, and new extensive documentation. A detailed account of these changes can be found in the **Supplementary Datasheet 1**. Here, we highlight the most relevant modifications.

2.1.1. Modeling Updates

To facilitate model usage and reduce human error, the identifiers of reactions and metabolites were converted from KEGG into BiGG human-readable form (King et al., 2015). Additionally, reaction and metabolite identifiers were linked to the modelSEED database (Henry et al., 2010) that enables analysis through the KBase web interface (Arkin et al., 2018). The gene identifiers and functional descriptions were updated to the most current annotation (NCBI Reference Sequence: NC_017304.1). Metabolite formulas and charges from the modelSEED database (Henry et al., 2010) were included in the model and reactions were systematically corrected for charge and mass balance by the addition of protons and water.

2.1.2. Metabolic Updates

The automated construction process used in the previous model introduced several inconsistencies that were corrected in the current model. We removed reactions that were blocked and non-gene-associated, apparently introduced during automated gap-filling. Two notable examples are (i) the blocked selenate pathway which lacks experimental evidence (e.g., selenoproteins have not been found in *C. thermocellum*), and (ii) blocked reactions involving molecular oxygen (e.g., oxidation of Fe^{2+} to Fe^{3+}) that are not possible in strict anaerobes like *C. thermocellum*. Furthermore, tRNA cycling reactions were unblocked by including tRNA in the biomass reaction (Reimers et al., 2017). Metabolite isomers were examined and consolidated under the same metabolite identifier when possible, leading to the removal of duplicated reactions and the elimination of gaps. Transport and exchange reactions were updated to reflect the export of amino acids and uptake of pyruvate as observed during fermentation experiments (Holwerda et al., 2014).

In terms of specific reactions, oxaloaceate decarboxylase was eliminated from the model in accordance with recent findings (Olson et al., 2017). The stoichiometries of pentose-phospate reactions, including sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase (FBA3) and sedoheptulose 1,7-bisphosphate ppi-dependent phosphofructokinase (PFK3_ppi), were corrected (according to experimental evidence, Rydzak et al., 2012) from the previous model by ensuring mass balance and avoiding lumping multiple steps into one reaction. Transaldolase (TALA) was removed from the model due to lack of annotation for this gene in *C. thermocellum*.

Several modifications were also performed in key bioenergetic reactions. The reactions catalyzed by membrane-bound enzymes, including inorganic diphosphatase (PPA) (Zhou et al., 2013) and membrane-bound ferredoxin-dependent hydrogenase (ECH) (Calusinska et al., 2010), were corrected to capture proton translocation. Furthermore, hydrogenase reactions were updated to ensure ferredoxin association for all cases and remove those reactions that do not involve ferredoxin and only use NAD(P)H as a cofactor, based on our recent understanding of *C. thermocellum* metabolism (Biswas et al., 2017). Gene-protein-reaction associations were updated to represent experimental knowledge. For instance, the hydrogenases BIF (CLO11313_RS09060-09070) and H2ASE (CLO1313_RS12830, CLO1313_RS02840) require the maturase Hyd (CLO1313_RS07925, CLO1313_RS11095, CLO1313_RS12830) to be functional, and the maturase itself requires all of its subunits to operate, which enables accurate representations of *hydG* deletion genotypes (Biswas et al., 2015).

Two hypothetical reaction modifications were introduced to ensure consistency with reported phenotypes. First, to enable growth without the need for succinate secretion, as observed in experimental data (**Supplementary Datasheet 2**), the reaction homoserine-O-trans-acetylase (HSERTA) was added to enable methionine biosynthesis (essential for growth). Although this reaction is not currently known to be associated with any gene in *C. thermocellum*, it is present as a gene-associated reaction in other *Clostridium* GSMs (Nagarajan et al., 2013). Next, the reaction deoxyribose-phosphate aldolase (DRPA) was removed based on a systematic analysis (section 4.4) to ensure correct lethality prediction of the Δ*hydG*Δ*ech*Δ*pfl* mutant strain as well as the correct prediction of growth recovery in this mutant by addition of external electron sinks such as sulfate or ketoisovalerate (**Table 1**). The correct prediction of Δ*hydG*Δ*ech*Δ*pfl*-associated phenotypes is critical to successfully use the model for computational strain design (Long et al., 2015; Ng et al., 2015; Maranas and Zomorrodi, 2016; Wang and Maranas, 2018; Garcia and Trinh, 2019a,b,c).

2.2. Comparison of iCBI655 Against Other Genome-Scale Models

We compared iCBI655 with the previous GSMs of *C. thermocellum* and the highly-curated GSM iML1515 of

TABLE 1 | Comparison of mutant growth rates predicted by iAT601 and iCBI655.

Gene deletions	Medium	Percent of W.T. growth rate (%)		
		iAT601	***iCBI655***	**Experiment**
hydg	MTC	100	100	73
hydg-ech	MTC	85	85	67
hydg-pta-ack	MTC	100	100	48
hydG-ech-pfl	MTC	58	0	0
hydG-ech-pfl	MTC + fumarate	377	726	0
hydG-ech-pfl	MTC + sulfate	58	65	+
hydG-ech-pfl	MTC + ketoisovalerate	97	101	+

Experimental values are taken from Thompson et al. (2015); for some mutants whose growth recovery, not growth rate, was reported, they are presented with "+". W.T., wildtype; MTC, Medium for Thermophilic Clostridia.

TABLE 2 | Comparison of all genome-scale metabolic models of *C. thermocellum* and the latest *E. coli* model.

	iSR432	iCth446	iAT601	iCBI665	iML1515
Strain	ATCC27405	ATCC27405	DSM1313	DSM1313	MG1655
Genes	432	446	601	665	1515
Metabolites	583	599	903	795	1877
Reactions	632	660	872	854	2712
Blocked reactions	39.2%	32.1%	40.8%	35.1%	9.8%
Reference	Roberts et al., 2010	Dash et al., 2017	Thompson et al., 2016	This study	Monk et al., 2017

the extensively studied bacterium *Escherichia coli* (**Table 2**). The increased number of genes in iCBI655 with respect to iAT601 cover a variety of functions, including hydrogenase chaperones, cellulosome and cellulase, ATP synthase, and transporters. Remarkably, iCBI655 has a smaller percentage of blocked reactions than iAT601, indicating higher biochemical consistency. The number of metabolites in iCBI655 is smaller than those in iAT601 mainly due to the removal of metabolites that did not appear in any reaction, duplicated metabolites (e.g., certain isomers), and blocked pathways added automatically during gap-filling without any gene association. *C. thermocellum* DSM1313 has 2911 protein coding genes, 22% of which is captured by iCB655, while *E. coli* MG1655 has 4240 genes, 35% of which is included in iML1515. Overall, iCBI655 has the increased coverage of the metabolic functionality of *C. thermocellum* but remains far from the well-studied *E. coli*.

2.3. Training of Model Parameters Under Diverse Conditions

Growth and non-growth associated maintenance (GAM and NGAM) are parameters that capture the consumption of ATP toward cell division and homeostasis, respectively. These are known to be condition-specific; however, genome-scale models do not include a mechanistic description that allows to determine these ATP consumption rates as part of the simulation. Instead, GAM is incorporated into the biomass pseudo-reaction and NGAM has its own pseudo-reaction that hydrolyzes ATP at a rate tuned by the constraint parameters.

To increase model prediction accuracy for various conditions, we trained GAM and NGAM parameters of iCBI655 using an extensive dataset of 28 extracellular fluxes (**Supplementary Datasheet 2**) measured during the growth phase under different reactor configurations, carbon sources, and gene deletion mutants. This approach is based on the method used to train the iML1655 *E. coli* model (Monk et al., 2017). Remarkably, we observed highly linear trends under three different conditions, including chemostat reactor with cellobiose as a carbon source, chemostat reactor with cellulose as a carbon source, and batch reactor with either cellobiose or cellulose as a carbon source (**Figure 1A**). This model training has led to improved growth rate prediction of iCBI655 as compared to iAT601 that has previously been trained with only a smaller dataset (**Figure 1B**). Specifically, the iAT601 training dataset was limited to batch conditions; hence, the inaccurate predictions of iAT601 were observed for chemostat conditions (**Figure 1C**).

2.4. Assessment of Model Quality and Standard Compliance With Memote

The field of metabolic network modeling suffers from a lack of standard enforcement and quality control metrics that limit model reproducibility and applicability. To address this issue, Lieven et al. (2020) recently developed the Memote framework that systematically tests for standards and best practices in GSMs. We applied Memote to both the iCBI655 and *E. coli* iML1515 models for comparison (**Figure 1D**). This analysis produced five independent scores that assess model quality. The *consistency score* measures basic biochemical requirements, such as mass and charge balance of metabolic reactions, and it was near 100% for both models. Additionally, the different annotation scores quantify how many elements in the model contain relevant metadata. More specifically, the *systems biology ontology (SBO) annotation* indicates if an object in the model refers to a metabolite, reaction, or gene, while the respective *annotation scores* of these elements correspond to properties (e.g., name, chemical formula, etc.) and identifiers linking them to relevant databases (e.g., KEGG Kanehisa and Goto, 2000 or modelSEED Henry et al., 2010). The *overall score* is computed as a weighted average of all the individual scores with additional emphasis on the *consistency score*. In summary, the high scores obtained by iCBI655 indicate the quality of the model and ensure its applicability for future studies.

2.5. Model-Guided Analysis of Proteomics and Flux Datasets Sheds Light on Redox Metabolism Critical for Biofuel Production in *C. thermocellum*

For the first application of the genome-scale metabolic model, we aimed to understand the complex redox metabolism and potential redox bottlenecks critical for enhanced biofuel production in *C. thermocellum*. We used the model as a scaffold to analyze proteomics and metabolic flux data collected for the *C. thermocellum* wild-type and Δ*hydG*Δ*ech* strains. The

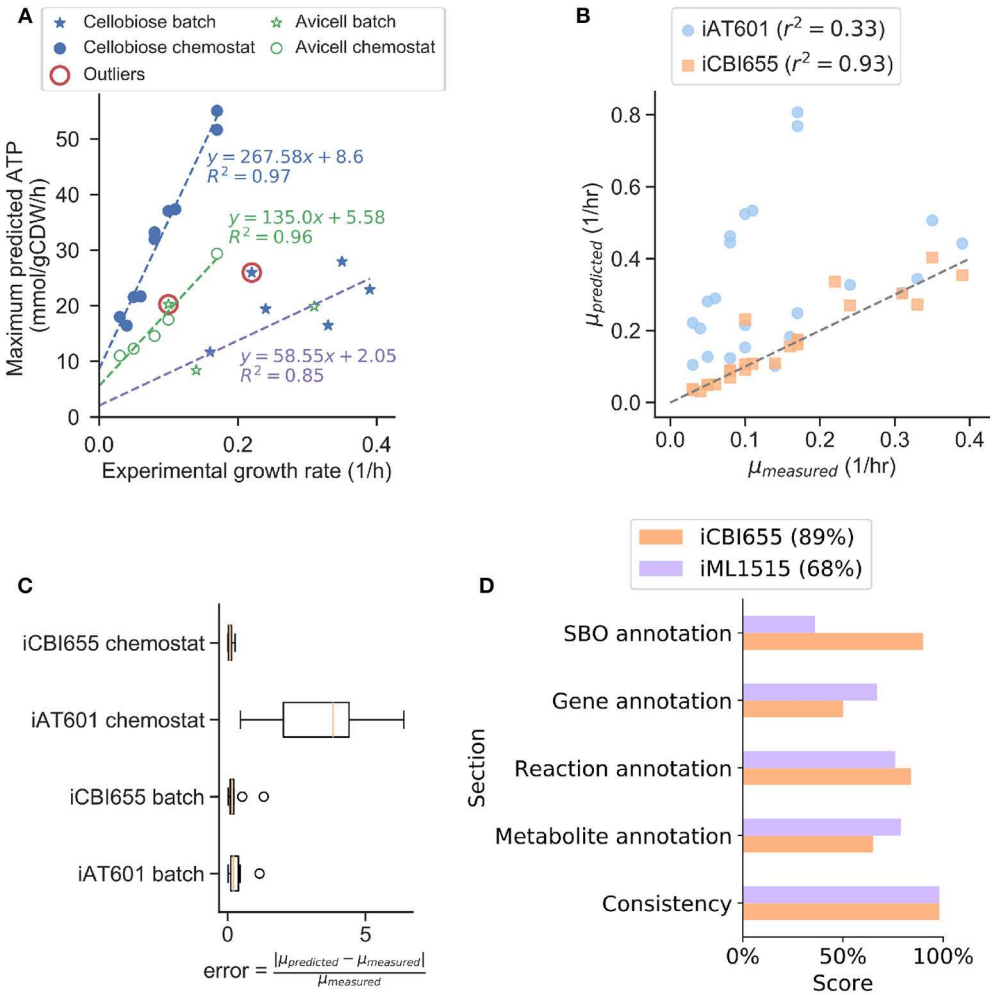


FIGURE 1 | Training and validation of the iCBI655 model. **(A)** Training of GAM and NGAM parameters. Discrete points correspond to experimental data. The slope of the linear regression function corresponds to GAM, while the intercept corresponds to NGAM. The data points circled as outliers were not included in any of the linear regression calculations. **(B)** Comparison of growth prediction error between iCBI655 and iAT601. Each maximum growth rate was predicted by constraining the models with the measured substrate uptake and product secretion fluxes (**Supplementary Datasheet 2**). r^2 corresponds to the Pearson correlation coefficient. **(C)** Error in growth predictions under batch and chemostat conditions. Predicted and measured growth rates correspond to the values included in **(B)**. **(D)** Scores provided by the quality control tool Memote (Lieven et al., 2020) for iCBI655 and iML1515. "Overall score" is shown in the legend.

Δ*hydG*Δ*ech* mutant was engineered to redirect electron flow from hydrogen to ethanol by removal of primary hydrogenases (Biswas et al., 2015; Thompson et al., 2015). Previous studies of Δ*hydG*Δ*ech* based on analysis of secretion fluxes (Thompson et al., 2015) or omics (Biswas et al., 2017) suggested the presence of redox bottlenecks in this mutant but did not identify which specific pathways and cofactors (i.e., NADH vs. NADPH) are responsible. We aim to solve this problem through integrated and quantitative analysis of omics and fluxes at the genome scale.

2.5.1. Development of Fold Change-Based Omics Integration Protocol

To perform the analysis, we formulated an omics integration protocol anchored at the quantification of fold change (FC) between case and control samples (**Figure 2A**). In this approach, we first compared FCs between simulated intracellular fluxes and measured omics data. Next, we identified *consistent reactions* with FCs of the same sign and different from zero in both measured proteomics and simulated fluxes for further analysis (section 4.6).

To start the FC-based omics integration, we obtained measured FCs by mapping the measured proteomics data to 510 out of the 856 reactions in the model through the gene-protein-reaction (GPR) associations (**Figure 2B**). Then, we identified 70 consistent reactions by comparing measured FCs with two types of simulated FCs: (i) parsimonious flux balance analysis (pFBA) that determines the most efficient flux distribution (assuming all enzymes are roughly as efficient) and (ii) flux variability analysis (FVA) that identifies the feasible flux range of each reaction.

The Pearson correlation coefficients between simulated and measured FCs for the consistent reactions were 0.26 and 0.09 for pFBA and FVA, respectively (**Figure 2C**). In general,

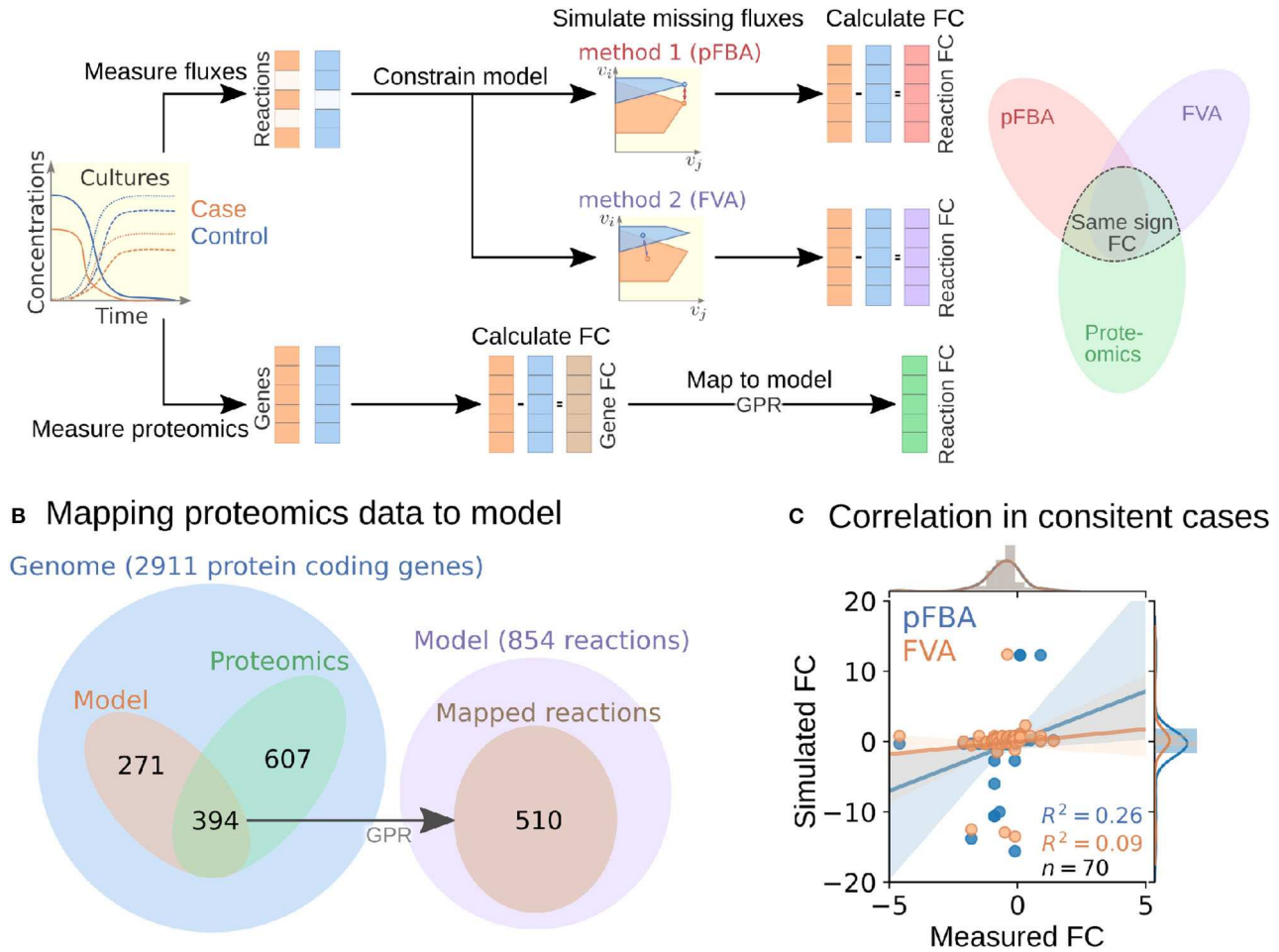


FIGURE 2 | Multi-omics data integration procedure. **(A)** Fold change-based multi-omics data integration and analysis protocol. **(B)** Mapping of proteomic data for the Δ*hydG*Δ*ech* case study to model reactions. **(C)** Correlation between measured and simulated fold changes (pFBA in blue and FVA in orange) for all 70 consistent reactions of the Δ*hydG*Δ *ech* case study.

the FVA reaction flux ranges remained mostly unchanged, suggesting that pFBA is a better representation of actual metabolic fluxes as previously observed (Machado and Herrgård, 2014). The top consistent reactions with the highest proteomics FCs (**Table S1**) belong primarily to the central metabolism of *C. thermocellum* (**Figure 3**). Interestingly, discrepancies in magnitude between flux and protein FCs for consistent reactions could be used to identify bottlenecks. For example, for a given enzyme, a small increase in flux combined with a large increase in translation could be an indicator of low catalytic efficiency; alternatively, such discrepancy could also point at an upstream thermodynamic bottleneck. Similar comparisons between simulated flux and omics has previously been used to identify regulatory mechanisms (Bordel et al., 2010). Overall, the identification and analysis of consistent reactions is an effective approach to gain certainty on the activity changes of metabolic pathways between conditions.

2.5.2. FC-Based Omics Integration Reveals Redirection of Electron Flow for NADPH Supply in Δ*hydG*Δ*ech* Strain

Our identification of the consistent reactions by using the FC-based omics integration protocol revealed coherent indications of increased NADPH biosynthesis in the Δ*hydG*Δ*ech* mutant with respect to the wild-type across three major metabolic areas: (i) an increased protein level of FRNDPR2r (also known as NFN) that converts one mol of reduced ferredoxin (*fdxr_42*) and one mole of NADH into two moles of NADPH (**Figure 3A**), (ii) an increased protein level of all three malate shunt enzymes and a decreased protein level of the alternative route PPDK (**Figure 3B**), and (iii) a decreased protein level of sulfur transporter and of HSOR that oxidizes sulfite into sulfide consuming NADPH (**Figure 3C**). These observations are consistent with the failure of *rnf* over-expression to enhance ethanol production (Lo et al., 2017), since RNF

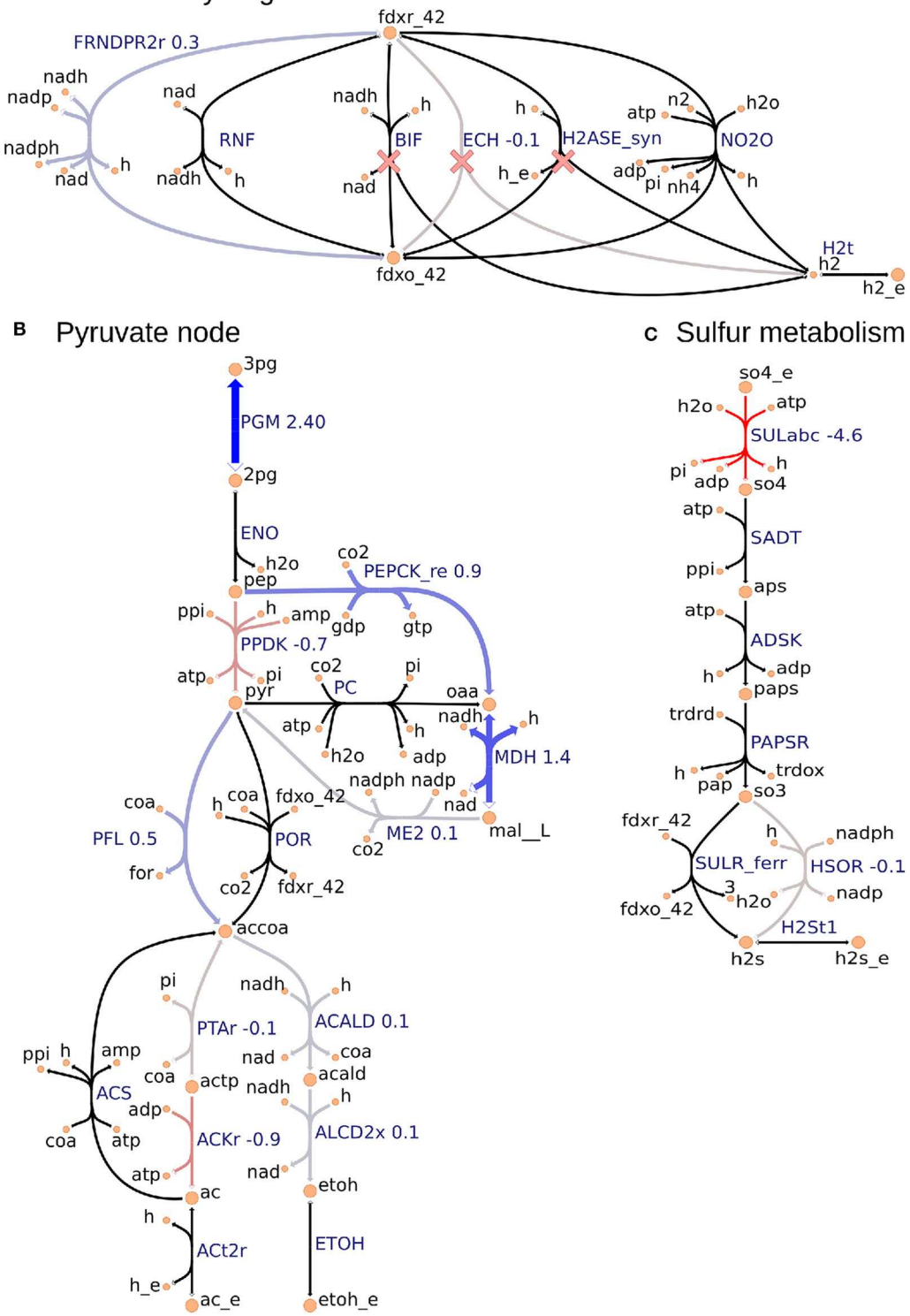


FIGURE 3 | Metabolic map visualization for **(A)** redox and hydrogenases pathway, **(B)** pyruvate node that links glycolysis, incomplete Krebs cycle, anapleurotic pathway, and fermentative pathway, and **(C)** sulfur metabolism using the Escher tool. Values next to reaction labels correspond to proteomics fold change between the Δ*hydG*Δ*ech* and wild-type strains only for the 70 consistent reactions identified by using the FC-based omics integration protocol (section 2.5). The labels of extracellular metabolites are appended with "_e." Reactions marked with a red cross are deleted in Δ*hydG*Δ*ech*.

produces NADH but the key cofactor bottleneck seems to be NADPH. Furthermore, a direct look at the proteomics data revealed that RNF subunits (Clo1313_0061-Clo1313_0066) had a statistically significant decrease in protein levels of the mutant (**Supplementary Datasheet 2**). The preference of Δ*hydG*Δ*ech* toward NADPH could be due to the cofactor specificity of the remaining redox balancing pathways (e.g., isobutanol), thermodynamics and protein cost constraints, or a combination of both. While the contribution of thermodynamic constraints is beyond the scope of this study, a recent analysis (Dash et al., 2019) of the ethanol production pathway in *C. thermocellum* highlighted the importance of engineering strategies to increase NADPH, for instance, introduction of NADPH-linked GAPDH that converts glyceraldehyde-3-phosphate to 3-phospho-D-glyceroyl phosphate in glycolysis or overexpression of NADPH-FNOR that transfers electrons from reduced ferredoxin to NADPH. The contribution of alternative redox balancing pathway toward the increased NADPH biosynthesis will be examined next.

2.5.3. Analysis of Simulated Fluxes Reveals the Role of NADPH in Redox Balancing

The analysis based on consistent reactions strongly indicates that NADPH production is important in the Δ*hydG*Δ*ech* mutant to achieve redox balance. However, the pathways oxidizing NADPH remain unknown since not all reactions in the model could be mapped to proteomics measurements and carbon recovery was lower in the mutant strain (Thompson et al., 2015). To identify these pathways, we examined the simulated fluxes of all reactions (instead of only consistent reactions) that differed in value between wild-type and mutant, and limited this search to exchange reactions and reactions that involve NADPH (**Table S2**). These simulated fluxes predicted an increase in the isobutanol pathway, including keto-acid reductoisomerase (KARA1) that consumes NADPH and isobutanol secretion (EX_ibutoh_e). The isobutanol pathway can consume NADPH through several enzymes (Lin et al., 2015) and has increased flux during overflow metabolism at high-substrate loading (Holwerda et al., 2014; Thompson and Trinh, 2017). The model also predicted a decrease in valine secretion (EX_val__L_e), since the isobutanol pathway competes with the valine pathway after KARA1. Remarkably, this prediction is consistent with the lower valine secretion measured in Δ*hydG*Δ*ech* (Biswas et al., 2017). A certain amount of NADPH is likely oxidized by the mutated alcohol-dehydrogenase enzyme observed after short adaptation in Δ*hydG* that is compatible with both NADH and NADPH (Biswas et al., 2015). However, this feature is not captured by the model since in general gene knock-outs are simulated by blocking the associated reactions. Overall this analysis indicates that Δ*hydG*Δ*ech* likely increases isobutanol secretion to alleviate redox imbalance.

Taken altogether, model-guided data analysis illustrates the power of the model as contextualization tool and provides new insights into the redox bottlenecks present in *C. thermocellum* that are critical in the production of reduced molecules. The integration of omics and fluxes led to the resolution of NADPH as the key cofactor in redox bottleneck of Δ*hydG*Δ*ech*. It helped identify specific pathways that undergo major changes in protein levels, providing interesting target reactions for further engineering. Generally, the developed FC-based omics integration protocol can be applied to different omics data types due to its simplicity. The method does not require one to formulate or assume a quantitative relationship between omics measurements and simulated fluxes. Furthermore, fold change in biomolecule concentrations implemented in the method is currently much easier to measure in a quantitatively reliable manner for many molecules than case-specific absolute concentrations.

2.6. Model-Guided Design of Modular Production Strains for Biofuel Synthesis

Another common application of genome-scale models is strain design (Long et al., 2015; Ng et al., 2015; Maranas and Zomorrodi, 2016; Wang and Maranas, 2018; Garcia and Trinh, 2019a,b,c). We used the iCBI655 model combined with the ModCell tool (section 4.9) to design *C. thermocellum* modular production strains for efficient biosynthesis of alcohols and esters. Briefly, with ModCell we aim to design a modular (chassis) cell that can be rapidly combined with exchangeable pathway modules in a plug-and-play fashion to obtain modular production strains exhibiting target phenotypes with minimal strain optimization cycles (Trinh, 2012; Trinh et al., 2015; Garcia and Trinh, 2019b,c). In this study, the target phenotype for modular production strains is growth-coupled to product synthesis (*wGCP*), that corresponds to the minimum product synthesis rate at the maximum growth rate. The ModCell mathematical formulation, computational algorithm, and implementation were described in details previously (Garcia and Trinh, 2019a,c, 2020). The design variables to attain the target phenotypes involve genetic manipulations of two types: (i) reaction deletions, constrained by the parameter α, that corresponds to gene knock-outs; and (ii) module reactions, constrained by the parameter β, that corresponds to reactions deleted in the modular cell but added back to specific modules to enhance the compatibility of the modular cell. Once these two parameters are specified, the solution to the problem is a set of Pareto optimal designs named Pareto front. In a Pareto optimal design, the performance (i.e., objective value) of a given module can only be increased at the expense of lowering the performance of another module. To characterize the practicality of each design, we say a modular cell is compatible with certain modules if the design objective is above a specific threshold (e.g., 0.5 in this study). Hence, the *compatibility* of a design corresponds to the number of compatible modules.

To design *C. thermocellum* modular cells, we first evaluated a range of design parameters α and β with an increasing number of genetic manipulations (**Figure 4A**). As expected, increasing the number of deletions leads to more compatible designs, at the expense of more complexity in the implementation. We selected an intermediate point of $\alpha = 6$, $\beta = 0$ for further analysis. This Pareto front is composed of 12 designs that can be clustered into two groups (**Figure 4B**). The first group (e.g., designs 3, 8, and 9) are compatible with all products except butanol and its derived

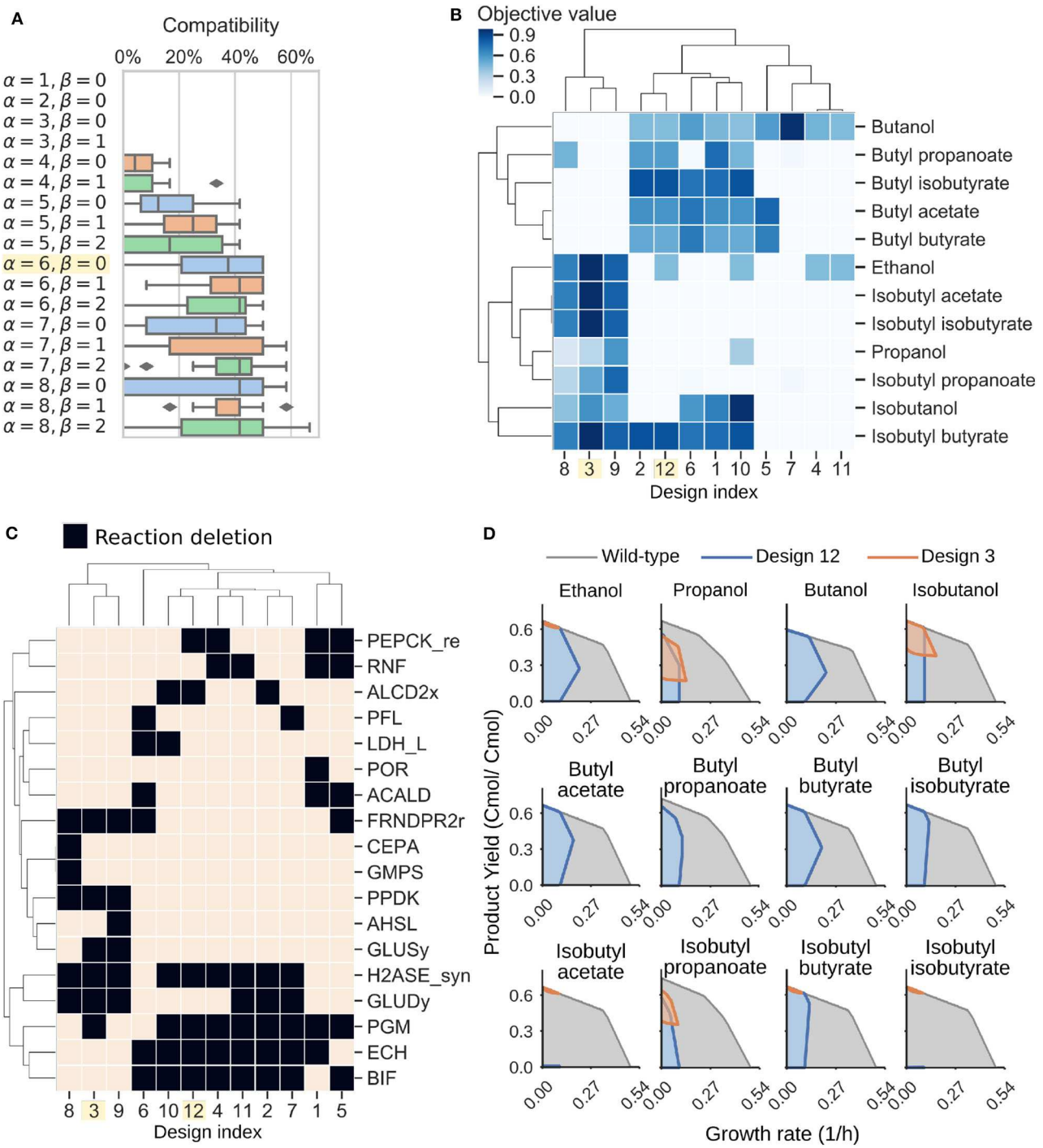


FIGURE 4 | Modular cell designs for biosynthesis of 12 alcohols and esters. **(A)** Module compatibility for various design parameters. **(B)** Pareto front for parameters $\alpha = 6, \beta = 0$. **(C)** Pareto set for parameters $\alpha = 6, \beta = 0$. Reaction names and formulas are included in **Table 3**. **(D)** Feasible phenotypic spaces for select designs.

esters, whereas the second group (e.g., designs 1, 2, 10, and 12) have high objective values for butanol and its derived esters.

To understand the characteristics of each group, we can inspect the deletions of each design (**Figure 4C**, **Table 3**). Designs 3, 8, and 9 all have in common H2ASE_syn, GLUDy, PPDK, and FRNDPR2r deletion, while the last two deletions never appear in design 1, 2, 10, or 12. The majority of deletion targets are central metabolic reactions (**Table 3**). The common targets include deletion of hydrogenases that appear in the cluster of designs 2, 4, 7, 10, 11, and 12 with the Δ*hydG*Δ*ech* genotype discussed

TABLE 3 | Reaction deletions sorted by appearance frequency (counts) in the designs of the Pareto front for $\alpha = 6$, $\beta = 0$.

ID	Name	Formula	Counts (%)
PGM	Phosphoglycerate mutase	2pg_c ↔ 3pg_c	75
H2ASE_syn	Bidirectional [NiFe] Hydrogenase (Fe-H2)	h2_c + nadp_c ↔ h_c + nadph_c	75
ECH	(FeFe)-hydrogenase, ferredoxin dependent, membrane-bound	2.0 fdxr_42_c + 3.0 h_c ↔ 2.0 fdxo_42_c + h2_c + h_e	66.7
BIF	Bifurcating Hydrogenase	2.0 fdxr_42_c + 3.0 h_c + nadh_c ↔ 2.0 fdxo_42_c + 2.0 h2_c + nad_c	66.7
GLUDy	Glutamate dehydrogenase (NADP)	glu__L_c + h2o_c + nadp_c ↔ akg_c + h_c + nadph_c + nh4_c	50
FRNDPR2r	Ferredoxin: nadp reductase (NFN)	2.0 fdxr_42_c + h_c + nadh_c + 2.0 nadp_c ↔ 2.0 fdxo_42_c + nad_c + 2.0 nadph_c	41.7
RNF	Ferredoxin:NAD oxidoreductase (membrane bound)	2.0 fdxr_42_c + 2.0 h_c + nad_c ↔ 2.0 fdxo_42_c + h_e + nadh_c	33.3
PEPCK_re	Phosphoenolpyruvate carboxykinase (GTP)	co2_c + gdp_c + pep_c → gtp_c + oaa_c	33.3
ALCD2x	Alcohol dehydrogenase (ethanol)	acald_c + h_c + nadh_c → etoh_c + nad_c	25
ACALD	Acetaldehyde dehydrogenase (acetylating)	accoa_c + h_c + nadh_c → acald_c + coa_c + nad_c	25
PPDK	Pyruvate, phosphate dikinase	amp_c + 2.0 h_c + pep_c + ppi_c → atp_c + pi_c + pyr_c	25
GLUSy	Glutamate synthase (NADPH)	akg_c + gln__L_c + h_c + nadph_c → 2.0 glu__L_c + nadp_c	16.7
PFL	Pyruvate formate lyase	coa_c + pyr_c → accoa_c + for_c	16.7
LDH_L	L-lactate dehydrogenase	h_c + nadh_c + pyr_c → lac__L_c + nad_c	16.7
POR	Pyruvate-ferredoxin oxidoreductase	coa_c + 2.0 fdxo_42_c + pyr_c → accoa_c + co2_c + 2.0 fdxr_42_c + h_c	8.3
CEPA	Cellobiose phosphorylase	cellb_c + pi_c → g1p_c + glc__D_c	8.3
GMPS	GMP synthase	atp_c + nh4_c + xmp_c → amp_c + gmp_c + 3.0 h_c + ppi_c	8.3
AHSL	O-Acetyl-L-homoserine succinate-lyase	achms_c + cys__L_c ↔ ac_c + cyst_L_c + h_c	8.3

earlier or removal of reactions that form fermentative byproducts such as ALCD2x and ACALD (ethanol), PFL (formate), LDH_L (lactate). Interestingly, ACKr or PTA (acetate) does not appear in this list, likely because acetate production can serve as a regulatory valve for redox metabolism, especially in a modular cell that must be compatible with products of diverse degrees of reduction.

More interestingly, we also found important branch-point deletion reactions (Stephanopoulos and Vallino, 1991) in central metabolism that have not yet been explored for strain design. Most prominently, these reactions include GLUDy, PEPCK_re, and PPDK, which appear with percentage frequencies of 50%, 33.3%, and 25%, respectively (**Table 3**). Both PEPCK_re and PPDK present two alternative routes that influence the ratio of NADPH to NADH, which is relevant to control metabolic fluxes though the specific dependencies of certain enzymes toward each redox cofactor. Since GLUDy consumes NADPH and is a key reaction in amino-acid metabolism, this enzyme and related ones (e.g., GLUSy) are interesting targets for practical implementation. We speculate the two product groups emerge likely because the butanol pathway strictly depends on NADH due to the reactions ACOAD1z (acyl-CoA dehydrogenase) and HACD1 (3-hydroxyacyl-CoA dehydrogenase), while the ethanol, propanol, and isobutanol pathways are more flexible in their use of NADH or NADPH. The designs 3, 8, and 9 perform poorly with butanol, and are also the only ones containing PPDK deletion. This deletion forces pep to pyruvate flux through the malate shunt that converts NADH to NADPH. Engineering of the cofactor specificities of the butanol pathway can be used to build one modular cell compatible with all products under consideration.

Two representative designs from the groups mentioned earlier are 3 and 12. Their feasible growth and production phenotypes reveal a tight coupling between product formation and growth rate (**Figure 4D**). This phenotype enables pathway optimization through adaptive laboratory evolution, as previously done for ethanol (Tian et al., 2016), overcoming one of the main challenges of *C. thermocellum* engineering that is optimization of enzyme expression levels. Hence, the proposed modular cells can also serve as platforms for pathway selection and optimization. In summary, this analysis demonstrates the potential of the model to identify non-intuitive metabolic engineering strategies that can be key to build effective modular platform strains for the production of biofuels and biochemicals in *C. thermocellum*.

3. CONCLUSIONS

In this study, we developed a genome-scale metabolic model of the biotechnologically relevant organism *C. thermocellum*. Model development followed standards and best practices to ensure reproducibility and accessibility. We demonstrated the enhanced predictions of the model for diverse fermentation conditions and gene lethality. Genome-scale models have a broad range of applications in systems biology, including metabolic engineering, physiological discovery, phenotype interpretation, and studies of evolutionary processes (Feist and Palsson, 2008; Palsson, 2015). To illustrate the model applications, we chose to tackle the challenge of disparate data integration and interpretation at the systems level. We developed a fold-change-based omics integration method for this purpose, and used it to identify routes in central metabolism that were selected to increase NADPH generation in the Δ*hydG*Δ*ech* strain. This analysis revealed the importance of NADPH cofactor over its alternatives and provided new engineering targets for enhanced biosynthesis of

reduced products in *C. thermocellum*. We also illustrated the use of the model to design *C. thermocellum* modular cells, using the ModCell tool (Garcia and Trinh, 2019b). The proposed designs cover C2 through C4 alcohols and their derived esters, which are key target molecules for renewable production with *C. thermocellum* (Peters, 2018). The proposed designs feature a combination of previously-explored and novel strategies to couple target metabolite production with cellular growth. Like the well-developed genome-scale models (Monk et al., 2017; Lu et al., 2019) of the important organisms *Escherichia coli* and *Saccharomyces cerevisiae* broadly used for strain engineering both in academia (Blazeck and Alper, 2010) and industry (Yim et al., 2011), we anticipate the iCBI655 genome-scale model will also provide a versatile tool for systems metabolic engineering of *C. thermocellum*.

4. METHODS

4.1. Model Curation

The genome scale model iCBI655 was constructed from iAT601 (Thompson et al., 2016) by following the standard GSM development protocol (Thiele and Palsson, 2010). Reaction and metabolite identifiers were mapped from KEGG to BiGG using the BiGG API (King et al., 2015). Metabolite charges were obtained from modelSEED when available, and otherwise calculated using the Chemaxon pKa plugin (Szegezdi and Csizmadia, 2007) for a pH of 7.2 (Thiele and Palsson, 2010). The biomass objective function was consolidated into one pseudo-reaction avoiding the use of intermediate pseudo-metabolites present in iAT601. Reactions were assigned with a confidence level based on a standard genome-scale model annotations (Thiele and Palsson, 2010).

4.2. Metabolic Flux Simulations

Constraint-based metabolic network modeling (Palsson, 2015) is based on the feasible flux space, Ω_k, defined by network stoichiometry and flux bounds that represent thermodynamic constraints and measured values:

$$\Omega_k := \{v_{jk} \in \mathbb{R} : \sum_{j \in \mathcal{J}} S_{ij} v_{jk} = 0 \ \forall i \in \mathcal{I} \quad (1)$$

$$l_{jk} \leq v_{jk} \leq u_{jk} \ \forall j \in \mathcal{J}\} \quad (2)$$

Here $\mathcal{I}$ and $\mathcal{J}$ are the sets of metabolites and reactions in the model, respectively, and v_{jk} is the metabolic flux (mmol/gCDW/h) through reaction j in the simulation condition k. Constraint (1) enforces mass balance for all metabolites in the network, where S_{ij} represents the stoichiometric coefficient of metabolite i in reaction j. Constraint (2) enforces lower and upper bounds l_{jk} and u_{jk}, respectively, for each reaction j in the network.

In different simulation conditions, k, S_{ij} remains fixed given the structure of the network for all $i, j \in \mathcal{I}, \mathcal{J}$. However, certain bounds u_{jk} and l_{jk} are modified to represent specific metabolic constraints. For example, to apply measured reaction fluxes such as in the case of GAM and NGAM calculation or the omics integration protocol (section 4.6), l_{jk} and u_{jk} are specified using the experimentally measured average (μ_{jk}) and standard deviation (σ_{jk}), which for normally distributed samples with 3 replicates produces an interval with a confidence level above 90% (3-4). Similarly, to represent a certain gene deletion mutant k, the bounds are set to be $u_{jk} = l_{jk} = 0$ for the associated reaction j.

$$l_{jk} = \mu_{jk} - \sigma_{jk} \ \forall j \in \text{Measured}_k \quad (3)$$

$$u_{jk} = \mu_{jk} + \sigma_{jk} \ \forall j \in \text{Measured}_k \quad (4)$$

The feasible flux space Ω_k can be explored in different ways; (Trinh et al., 2009; Palsson, 2015) for instance, an optimization objective is often defined to identify specific flux distributions $v_{jk}^{\text{sim}} \ \forall j \in \mathcal{J}$:

$$v_{jk}^{\text{sim}} \in \arg\max \left\{ \sum_{j \in \mathcal{J}} c_j v_{jk} : v_{jk} \in \Omega_k \right\} \forall j \in \mathcal{J} \quad (5)$$

Here c_j is the coefficient of reaction j in the linear objective function, which is changed according to the simulation context. For example, to train GAM and NGAM (**Figure 1A**) the objective was set to maximize flux through the ATP hydrolysis reaction, i.e., $c_j = 1$ for j corresponding to ATP hydrolysis reaction, and 0 otherwise. To evaluate growth prediction accuracy (**Figures 1B,C**), the objective was set to maximize growth, i.e., $c_j = 1$ for j corresponding to growth pseudo-reaction and 0 otherwise.

4.3. Simulation of Different Growth Environments

The model is configured to generally represent different medium and reactor conditions by modifying three features. The first feature involves model boundaries specifying which metabolites may enter the intracellular environment (i.e., present in the growth medium) or may exit the intracellular environment (i.e., secreted by *C. thermocellum*). This feature can be adjusted through u_{jk} and l_{jk} for exchange reactions. In our simulations, only essential metabolites required for *in silico* growth may be consumed and only commonly observed metabolites may be produced, unless otherwise noted. The second feature involves biomass objective function. iCBI655 contains 3 possible biomass reactions: (i) BIOMASS_CELLOBIOSE used for growth in cellobiose with cellulosan constituting 2% of cell dry weight (CDW) (Zhang and Lynd, 2005), (ii) BIOMASS_CELLULOSE used for growth on cellulose with cellulosan constituting 20% of CDW (Zhang and Lynd, 2005), and (iii) BIOMASS_NO_CELLULOSOME, a biomass function that does not consider cellulosan production and only used as a control. The combination of cellulosome and protein fractions accounts for 52.85% of the CDW in all cases (Roberts et al., 2010; Thompson et al., 2015). Cellobiose conditions were used in all simulations unless otherwise noted. The third feature involves GAM/NGAM. Three sets of these parameters are considered including batch, chemostat-cellulose, and chemostat-cellobiose, based on fitting the model to experimental data. Batch conditions were used in all simulations unless otherwise noted.

For growth on cellulose, the experimentally measured glucose-equivalent uptake was represented in the model through the following pseudo-reactions: *3 glceq_e → cell3_e*; *4 glceq_e → cell4_e*; *5 glceq_e → cell5_e*; and *6 glceq_e → cell6_e*. Here, *cell3_e*, *cell4_e*, *cell5_e*, and *cell6_e* are cellodextrin polymers with 3, 4, 5, and 6 glucose monomers, respectively. These polymers can be imported inside the cell through the oligo-cellulose transport ABC system. The model is free to use any cellodextrin length, although utilization of longer cellodextrins results in higher ATP yield (Zhang and Lynd, 2005; Thompson et al., 2016).

4.4. Single-Reaction Deletion Analysis to Match Experimentally Observed Phenotype

A core model of *C. thermocellum* (Thompson et al., 2015) correctly predicted the experimentally observed lethality of Δ*hydG*Δ*ech*Δ*pfl*; however, the iAT601 genome-scale model built by extension of this core model failed, suggesting that the genome-scale model has alternative active pathways leading to the false growth prediction *in silico*. To resolve this false positive prediction in iCBI655, we calculated the maximum growth rates for all possible additional single reaction deletions in the Δ*hydG*Δ*ech*Δ*pfl* mutant. This analysis resulted in three possible additional reaction deletions that are predicted to be lethal (i.e., maximum growth rate prediction below 20% of the simulated wild-type value Palsson, 2015), including the removal of (i) glycine secretion (EX_gly_e), (ii) 5,10-methylenetetrahydrofolate oxidoreductase (MTHFC), and (iii) deoxyribose-phosphate aldolase (DRPA). For the first removal, addition of sulfate or ketoisovalerate in the growth medium of Δ*hydG*Δ*ech*Δ*pfl* fails to predict growth recovery as observed experimentally (Thompson et al., 2015), making this option invalid. Likewise, the second removal is invalid because it makes PFL an essential reaction in the wild-type strain; however, experimental evidence demonstrates that Δ*pfl* mutant is able to grow (Papanek et al., 2015). The last option was chosen since it correctly predicts growth recovery of Δ*hydG*Δ*ech*Δ*pfl* by sulfate or ketoisovalerate addition in the growth medium, and does not make PFL essential in the wild-type strain.

4.5. Model Comparison

The *C. thermocellum* and *E. coli* models were obtained from their respective publications in SBML format. Blocked reactions were calculated by allowing all exchange reactions to have an unconstrained flux (i.e., $lb_{jk} = -1,000$, $ub_{jk} = 1,000 \quad \forall j \in Exchange$). This procedure enables the most general scenario which produces the smallest number of blocked reactions in each model. Additional details can be found in **Supplementary Datasheet 1**.

4.6. Omics Integration Protocol

The omics integration protocol developed in this study consists of three steps: (i) simulation of fold changes, (ii) mapping of measured gene fold changes to reactions, and (iii) comparison of measured and simulated fold changes.

4.6.1. Calculation of Simulated Fold Changes

To simulate metabolic fluxes, lower and upper bounds (2) are constrained according to experimental data as described in section 4.2. Then, for the pFBA method, a quadratic optimization problem (6) is solved, leading to a unique flux distribution $v_{jk}^{\text{pFBA}} \forall j \in \mathcal{J}$.

$$v_{jk}^{\text{pFBA}} \in \arg\min \left\{ \sum_{j \in \mathcal{J}} v_{jk}^2 : v_{jk} \in \Omega_k \right\} \forall j \in \mathcal{J} \tag{6}$$

For the FVA method, a sequence of linear programming problems is solved where each flux is minimized (7) and maximized (8):

$$v_{jk}^{\min} \in \arg\min \left\{ v_{jk} : v_{jk} \in \Omega_k \right\} \ \forall j \in \mathcal{J} \tag{7}$$

$$v_{jk}^{\max} \in \arg\max \left\{ v_{jk} : v_{jk} \in \Omega_k \right\} \ \forall j \in \mathcal{J} \tag{8}$$

Note that for computation we applied the loop-less FVA method (Schellenberger et al., 2011; Chan et al., 2018), as implemented in cobrapy (Ebrahim et al., 2013), that introduces additional constraints in Ω_k to remove thermodynamically infeasible cycles from all feasible flux distributions.

FVA produces a flux range $[v_{jk}^{\min}, v_{jk}^{\max}]$ for each reaction $j \in \mathcal{J}$. To compare between states k (e.g., wild-type and mutant), we define the *FVA center*, a scalar variable that generally indicates a change in this range (9).

$$v_{jk}^{\text{FVA}} = \frac{v_{jk}^{\max} + v_{jk}^{\min}}{2} \tag{9}$$

The FVA center is a heuristic analysis with the main purpose of determining whether a reaction exhibits an upward shift (center increase) or a downward shift (center decrease) between two conditions k. It should be emphasized that the FVA center, v_{jk}^{FVA}, does not attempt to quantify the fraction of overlap between ranges nor to identify what type of shift might occur from all possible permutations. Unlike v_{jk}^{pFBA}, v_{jk}^{FVA} does not necessarily represent a feasible flux distribution of Ω_k. Furthermore, the FVA center could potentially fail to capture hypothetical permutations of fluxes. Despite these considerations, the FVA center remains a useful heuristic to analyze simulated fold changes.

Finally, to determine the fold change for either pFBA or FVA simulated fluxes, the conventional procedure for fold change calculation in omics data is emulated. First, values are floored to avoid very large (or infinite) fold changes in cases with very small magnitude change. This is accomplished through a flooring piece-wise function (10), where $\epsilon = 0.0001$ is the minimum value and x is an arbitrary scalar variable.

$$floor(x) = \begin{cases} x + \epsilon & \text{if } 0 < x < \epsilon \\ x - \epsilon & \text{if } 0 > x > -\epsilon \\ x & \text{otherwise} \end{cases} \tag{10}$$

Then, the fluxes are normalized to the substrate uptake rate $v_{\text{uptake},k}$ and fold change is calculated in $\log_2$ space (11).

$$FC_j^{\text{sim}}(v_{j,\text{mut}}, v_{j,\text{wt}}) = \log_2\left[floor\left(\frac{v_{j,\text{mut}}}{|v_{\text{uptake,mut}}|}\right)\right] - \log_2\left[floor\left(\frac{v_{j,\text{wt}}}{|v_{\text{uptake,wt}}|}\right)\right] \quad (11)$$

4.6.2. Calculation of Measured Fold Changes

Fold change between case and control samples, FC_l, is calculated in $\log_2$ space for each gene $l \in \mathcal{L}$, where $\mathcal{L}$ is the set of genes in the model. These gene fold changes can be mapped to metabolic reaction fold changes using the gene-protein reaction associations (GPR), given $\mathcal{G}_j$ as the set of genes with $FC_l \neq 0$ in the GPR of reaction j:

$$FC_j^{\text{meas}} = \frac{1}{\text{card}(\mathcal{G}_j)} \sum_{l \in \mathcal{G}_j} FC_l \quad (12)$$

4.6.3. Identification of Consistent Fold Changes

A reaction j is said to have a consistent fold change if the measured fold change has the same sign of at least one of the simulated fold changes, more formally:

$$\mathcal{M} := \left\{ j \in \mathcal{J} : \left(\left[(FC_j^{\text{sim,pFBA}} < 0) \vee (FC_j^{\text{sim,FVA}} < 0)\right] \wedge (FC_j^{\text{meas}} < 0)\right) \vee \left(\left[(FC_j^{\text{sim,pFBA}} > 0) \vee (FC_j^{\text{sim,FVA}} > 0)\right] \wedge (FC_j^{\text{meas}} > 0)\right) \right\} \quad (13)$$

where $\mathcal{M} \subseteq \mathcal{J}$ is the set of consistent reactions which is considered for further analysis and the simulated fold changes are re-defined for brevity (14-15).

$$FC_j^{\text{sim,pFBA}} := FC_j^{\text{sim}}(v_{j,\text{mut}}^{\text{pFBA}}, v_{j,\text{wt}}^{\text{pFBA}}) \quad (14)$$

$$FC_j^{\text{sim,FVA}} := FC_j^{\text{sim}}(v_{j,\text{mut}}^{\text{FVA}}, v_{j,\text{wt}}^{\text{FVA}}) \quad (15)$$

4.7. Software Implementation

Model development was performed using Python and Jupyter notebooks with open-source Python libraries including cobrapy (Ebrahim et al., 2016). The sequence of upgrades and improvements can be seen in the Git version control records. The repository is available online through Github (https://github.com/trinhlab/ctherm-gem) and in **Supplementary Datasheet 1**.

4.8. Proteomics Data Collection

C. thermocellum wild-type and Δ*hydG*Δ*ech* strains were cultured in batch reactors and metabolic fluxes were calculated as previously described (Thompson et al., 2015). For proteomics measurements, the wild-type and mutant strains were cultured in MNM and MTC media (Kridelbaugh et al., 2013), respectively. While both wild-type and mutant were originally cultured in MTC (Thompson et al., 2015), the wild-type had to be cultured separately in MNM medium due to insufficient volume for proteomics sampling in the MTC culture. MTC has higher nitrogen and trace mineral concentrations, but previous studies have shown no effect on growth rates (Kridelbaugh et al., 2013). During the mid-exponential growth phase 10 mL samples were collected, centrifuged, and the resulting pellet was stored at −20 °C. Cell pellets were then prepared for LC-MS/MS-based proteomic analysis. Briefly, proteins extracted via SDS, boiling, and sonic disruption were precipitated with trichloroacetic acid (Giannone et al., 2015b). The precipitated protein was then resolubilized in urea and treated with dithiothreitol and iodoacetamide to reduce and block disulfide bonds prior to digestion with sequencing-grade trypsin (Sigma-Aldrich). Following two-rounds of proteolysis, tryptic peptides were salted, acidified, and filtered through a 10 kDa MWCO spin column (Vivaspin 2; GE Healthcare) and quantified by BCA assay (Pierce).

For each LC-MS/MS run, 25 µg of peptides were loaded via pressure cell onto a biphasic MudPIT column for online 2D HPLC separation and concurrent analysis via nanospray MS/MS using a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) operating in data-dependent acquisition (one full scan at 15 k resolution followed by 10 MS/MS scans in the LTQ, all one µscan; monoisotopic precursor selection; rejection of analytes with an undecipherable charge; dynamic exclusion = 30 s) (Giannone et al., 2015a).

Eleven salt cuts (25, 30, 35, 40, 45, 50, 65, 80, 100, 175, and 500 mM ammonium acetate) were performed per sample run with each followed by 120 min organic gradient to separate peptides.

Resultant peptide fragmentation spectra (MS/MS) were searched against the *C. thermocellum* DSM1313 proteome database concatenated with common contaminants and reversed sequences to control false-discovery rates using MyriMatch v.2.1. (Tabb et al., 2007). Peptide spectrum matches (PSM) were filtered by IDPicker v.3 (Ma et al., 2009) to achieve a peptide-level FDR of <1 % per sample run and assigned matched-ion intensities (MIT) based on observed peptide fragment peaks. PSM MITs were summed on a per-peptide basis and those uniquely mapping to their respective proteins were imported into InfernoRDN (Taverner et al., 2012). Peptide intensities were log2-transformed, normalized across replicates by LOESS, standardized by median absolute deviation, and median centered across all samples. Peptide abundance data were then assembled to proteins via RRollup and further filtered to maintain at least two values in at least one replicate set. Protein abundances were then used for the modeling efforts describe herein.

All raw and database-searched LC-MS/MS data pertaining to this study have been deposited into the MassIVE proteomic data repository and have been assigned the following accession numbers: MSV000084488 (MassIVE) and PXD015973 (ProteomeXchange). Data files are available upon publication (ftp://massive.ucsd.edu/MSV000084488/).

4.9. Modular Cell Design

The ModCell formulation, computational algorithm, and implementation followed the previous reports (Garcia and Trinh, 2019a,c, 2020). The iCBI655 model with cellobiose as a carbon source in the batch reactors (**Supplementary Datasheet 4**) was used as an input for modular cell design. The alcohol pathways

were curated from recent literature (Holwerda et al., 2014; Lin et al., 2015; Loder et al., 2015), where adapted Adh can use either NADH or NADPH as an electron donor to synthesize the target alcohol (Biswas et al., 2015). The esters-producing pathways require an alcohol acetyltransferase (AAT) reaction to condense an alcohol and acyl-CoA that are already present in the alcohols-producing pathways. Even though a thermostable AAT has not yet been reported in literature to function at high temperature, an engineered chloramphenicol acetyl transferase (CAT) can be repurposed as a thermostable AAT (Seo et al., 2019, 2020). The ModCell software is available online at https://github.com/TrinhLab/ModCell2.

DATA AVAILABILITY STATEMENT

All raw and database-searched LC-MS/MS data pertaining to this study have been deposited into the MassIVE proteomic data repository and have been assigned the following accession numbers: MSV000084488 (MassIVE) and PXD015973 (ProteomeXchange). Data files are available upon publication (ftp://massive.ucsd.edu/MSV000084488/).

AUTHOR CONTRIBUTIONS

CT managed the project. SG and CT conceived the study, designed experiments, and analyzed the data. SG, RT, RG, and SD performed the experiments. SG prepared the draft with co-authors' inputs. All read, wrote, and approved the final draft.

FUNDING

This research was financially supported in part by the NSF CAREER award (NSF#1553250 to CT) and by The Center of Bioenergy Innovation (CBI), U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science (to CT and CM). The funders had no role in this study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00772/full#supplementary-material

Supplementary Datasheet 1 | Software used to develop, configure, and analyze iCBI655.

Supplementary Datasheet 2 | Flux dataset used to train the iCBI655 model and proteomics dataset for the wild-type and Δ*hydG*Δ*ech* strains.

Supplementary Datasheet 3 | Supplementary tables.

Supplementary Datasheet 4 | iCBI655 model in various formats for cellobiose growth conditions and map of central metabolic pathways in Escher format.

REFERENCES

Akinosho, H., Yee, K., Close, D., and Ragauskas, A. (2014). The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Front. Chem.* 2:66. doi: 10.3389/fchem.2014.00066

Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., et al. (2011). High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl. Environ. Microbiol.* 77, 8288–8294. doi: 10.1128/AEM.00646-11

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Biswas, R., Wilson, C. M., Giannone, R. J., Klingeman, D. M., Rydzak, T., Shah, M. B., et al. (2017). Improved growth rate in *Clostridium thermocellum* hydrogenase mutant via perturbed sulfur metabolism. *Biotechnol. Biofuels* 10:6. doi: 10.1186/s13068-016-0684-x

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., and Guss, A. M. (2015). Elimination of hydrogenase active site assembly blocks h 2 production and increases ethanol yield in *Clostridium thermocellum*. *Biotechnol. Biofuels* 8:20. doi: 10.1186/s13068-015-0204-4

Blazeck, J., and Alper, H. (2010). Systems metabolic engineering: genome-scale models and beyond. *Biotechnol. J.* 5, 647–659. doi: 10.1002/biot.200900247

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.* 6:e1000859. doi: 10.1371/journal.pcbi.1000859

Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156, 1575–1588. doi: 10.1099/mic.0.032771-0

Chan, S. H., Wang, L., Dash, S., and Maranas, C. D. (2018). Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics* 34, 4248–4255. doi: 10.1093/bioinformatics/bty446

Dash, S., Khodayari, A., Zhou, J., Holwerda, E. K., Olson, D. G., Lynd, L. R., et al. (2017). Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnol. Biofuels* 10:108. doi: 10.1186/s13068-017-0792-2

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., and Maranas, C. D. (2014). Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. *Biotechnol. Biofuels* 7:144. doi: 10.1186/s13068-014-0144-4

Dash, S., Ng, C. Y., and Maranas, C. D. (2016). Metabolic modeling of clostridia: current developments and applications. *FEMS Microbiol. Lett.* 363, 1–10. doi: 10.1093/femsle/fnw004

Dash, S., Olson, D. G., Chan, S. H. J., Amador-Noguez, D., Lynd, L. R., and Maranas, C. D. (2019). Thermodynamic analysis of the pathway for ethanol production from cellobiose in *Clostridium thermocellum*. *Metab. Eng.* 55, 161–169. doi: 10.1016/j.ymben.2019.06.006

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Shaw, A. J., and Lynd, L. R. (2013). Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. *Metab. Eng.* 15, 151–158. doi: 10.1016/j.ymben.2012.11.006

Ebrahim, A., Brunk, E., Tan, J., O'brien, E. J., Kim, D., Szubin, R., et al. (2016). Multi-omic data integration enables discovery of hidden biological regularities. *Nat. Commun.* 7:13091. doi: 10.1038/ncomms13091

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Feist, A. M., and Palsson, B. Ø. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26:659. doi: 10.1038/nbt1401

Garcia, S., and Trinh, C. T. (2019a). Comparison of multi-objective evolutionary algorithms to solve the modular cell design problem for novel biocatalysis. *Processes* 7:361. doi: 10.3390/pr7060361

Garcia, S., and Trinh, C. T. (2019b). Modular design: implementing proven engineering principles in biotechnology. *Biotechnol. Adv*. 37:107403. doi: 10.1016/j.biotechadv.2019.06.002

Garcia, S., and Trinh, C. T. (2019c). Multiobjective strain design: a framework for modular cell engineering. *Metab. Eng.* 51, 110–120. doi: 10.1016/j.ymben.2018.09.003

Garcia, S., and Trinh, C. T. (2020). Harnessing natural modularity of cellular metabolism to design a modular chassis cell for a diverse class of products by using goal attainment optimization. *ACS Synth. Biol.* 9, 1665–1681. doi: 10.1021/acssynbio.9b00518

Giannone, R. J., Wurch, L. L., Heimerl, T., Martin, S., Yang, Z., Huber, H., et al. (2015a). Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. *ISME J*. 9:101. doi: 10.1038/ismej.2014.112

Giannone, R. J., Wurch, L. L., Podar, M., and Hettich, R. L. (2015b). Rescuing those left behind: recovering and characterizing underdigested membrane and hydrophobic proteins to enhance proteome measurement depth. *Anal. Chem*. 87, 7720–7728. doi: 10.1021/acs.analchem.5b01187

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol*. 28:977. doi: 10.1038/nbt.1672

Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., et al. (2014). The exometabolome of *Clostridium thermocellum* reveals overflow metabolism at high cellulose loading. *Biotechnol. Biofuels* 7:155. doi: 10.1186/s13068-014-0155-1

Hsiang, S. M., Meng, K. C., and Cane, M. A. (2011). Civil conflicts are associated with the global climate. *Nature* 476:438. doi: 10.1038/nature10311

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28, 27–30. doi: 10.1093/nar/28.1.27

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2015). Bigg models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*. 44, D515-D522. doi: 10.1093/nar/gkv1049

Kridelbaugh, D. M., Nelson, J., Engle, N. L., Tschaplinski, T. J., and Graham, D. E. (2013). Nitrogen and sulfur requirements for *Clostridium thermocellum* and *Caldicellulosiruptor bescii* on cellulosic substrates in minimal nutrient media. *Bioresour. Technol*. 130, 125–135. doi: 10.1016/j.biortech.2012.12.006

Lee, J.-W., and Trinh, C. T. (2019). Microbial biosynthesis of lactate esters. *Biotechnol. Biofuels* 12:226. doi: 10.1186/s13068-019-1563-z

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., et al. (2020). Memote for standardized genome-scale metabolic model testing. *Nat. Biotechnol*. 38, 272–276. doi: 10.1038/s41587-020-0446-y

Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., et al. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metab. Eng*. 31, 44–52. doi: 10.1016/j.ymben.2015.07.001

Lo, J., Olson, D. G., Murphy, S. J.-L., Tian, L., Hon, S., Lanahan, A., et al. (2017). Engineering electron metabolism to increase ethanol production in *Clostridium thermocellum*. *Metab. Eng*. 39, 71–79. doi: 10.1016/j.ymben.2016.10.018

Loder, A. J., Zeldes, B. M., Garrison, G. D., Lipscomb, G. L., Adams, M. W., and Kelly, R. M. (2015). Alcohol selectivity in a synthetic thermophilic n-butanol pathway is driven by biocatalytic and thermostability characteristics of constituent enzymes. *Appl. Environ. Microbiol*. 81, 7187–7200. doi: 10.1128/AEM.02028-15

Long, M. R., Ong, W. K., and Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol*. 34, 135–141. doi: 10.1016/j.copbio.2014.12.019

Lu, H., Li, F., Sánchez, B. J., Zhu, Z., Li, G., Domenzain, I., et al. (2019). A consensus *S. cerevisiae* metabolic model yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun*. 10, 1–13. doi: 10.1038/s41467-019-11581-3

Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., et al. (2009). Idpicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J. Proteome Res*. 8, 3872–3881. doi: 10.1021/pr900360j

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

Maranas, C. D., and Zomorrodi, A. R. (2016). *Optimization Methods in Metabolic Networks*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119188902

McAnulty, M. J., Yen, J. Y., Freedman, B. G., and Senger, R. S. (2012). Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism *in silico*. *BMC Syst. Biol*. 6:42. doi: 10.1186/1752-0509-6-42

Milne, C. B., Eddy, J. A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R. S., et al. (2011). Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* ncimb 8052. *BMC Syst. Biol*. 5:130. doi: 10.1186/1752-0509-5-130

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. doi: 10.1038/nbt.3956

Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., et al. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. Microbial cell factories 12:118. doi: 10.1186/1475-2859-12-118

Ng, C. Y., Khodayari, A., Chowdhury, A., and Maranas, C. D. (2015). Advances in *de novo* strain design using integrated systems and synthetic biology tools. *Curr. Opin. Chem. Biol*. 28, 105–114. doi: 10.1016/j.cbpa.2015.06.026

Olson, D. G., Hörl, M., Fuhrer, T., Cui, J., Zhou, J., Maloney, M. I., et al. (2017). Glycolysis without pyruvate kinase in *Clostridium thermocellum*. *Metab. Eng*. 39, 169–180. doi: 10.1016/j.ymben.2016.11.011

Olson, D. G., McBride, J. E., Shaw, A. J., and Lynd, L. R. (2012). Recent progress in consolidated bioprocessing. *Curr. Opin. Biotechnol*. 23, 396–405. doi: 10.1016/j.copbio.2011.11.026

Palsson, B. Ø. (2015). *Systems Biology: Constraint-Based Reconstruction and Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139854610

Papanek, B., Biswas, R., Rydzak, T., and Guss, A. M. (2015). Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. *Metab. Eng*. 32, 49–54. doi: 10.1016/j.ymben.2015.09.002

Peters, N. K. (2018). *Bioenergy Research Centers*. Technical report, USDOE Office of Science (SC), Washington, DC. doi: 10.2172/1471709

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., et al. (2006). The path forward for biofuels and biomaterials. *Science* 311, 484–489. doi: 10.1126/science.1114736

Reimers, A.-M., Lindhorst, H., and Waldherr, S. (2017). A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites* 7:47. doi: 10.3390/metabo7030047

Roberts, S. B., Gowen, C. M., Brooks, J. P., and Fong, S. S. (2010). Genome-scale metabolic analysis of clostridium thermocellum for bioethanol production. *BMC Syst. Biol*. 4:31. doi: 10.1186/1752-0509-4-31

Rydzak, T., McQueen, P. D., Krokhin, O. V., Spicer, V., Ezzati, P., Dwivedi, R. C., et al. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. *BMC Microbiol*. 12:214. doi: 10.1186/1471-2180-12-214

Salimi, F., Zhuang, K., and Mahadevan, R. (2010). Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnol. J*. 5, 726–738. doi: 10.1002/biot.201000159

Schellenberger, J., Lewis, N. E., and Palsson, B. Ø. (2011). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J*. 100, 544–553. doi: 10.1016/j.bpj.2010.12.3707

Senger, R. S., and Papoutsakis, E. T. (2008). Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng*. 101, 1036–1052. doi: 10.1002/bit.22010

Seo, H., Lee, J.-W., Garcia, S., and Trinh, C. T. (2019). Single mutation at a highly conserved region of chloramphenicol acetyltransferase enables isobutyl acetate production directly from cellulose by *Clostridium thermocellum* at elevated temperatures. *Biotechnol. Biofuels* 12:245. doi: 10.1186/s13068-019-1583-8

Seo, H., Nicely, P. N., and Trinh, C. T. (2020). Endogenous carbohydrate esterases of *Clostridium thermocellum* are identified and disrupted for enhanced

isobutyl acetate production from cellulose. *Biotechnol. Bioeng.* 117, 2223–2236. doi: 10.1002/bit.27360

Serrano-Bermúdez, L. M., Barrios, A. F. G., Maranas, C. D., and Montoya, D. (2017). *Clostridium butyricum* maximizes growth while minimizing enzyme usage and ATP production: metabolic flux distribution of a strain cultured in glycerol. *BMC Syst. Biol.* 11:58. doi: 10.1186/s12918-017-0434-0

Shafiee, S., and Topal, E. (2009). When will fossil fuel reserves be diminished? *Energy Policy* 37, 181–189. doi: 10.1016/j.enpol.2008.08.016

Stephanopoulos, G., and Vallino, J. J. (1991). Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252, 1675–1681. doi: 10.1126/science.1904627

Szegezdi, J., and Csizmadia, F. (2007). "Method for calculating the PKA values of small and large molecules," in *Abstracts of Papers of The American Chemical Society*, Vol. 233 (Washington, DC: Amer Chemical Soc).

Tabb, D. L., Fernando, C. G., and Chambers, M. C. (2007). Myrimatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* 6, 654–661. doi: 10.1021/pr0604054

Taverner, T., Karpievitch, Y. V., Polpitiya, A. D., Brown, J. N., Dabney, A. R., Anderson, G. A., et al. (2012). Danter: an extensible r-based tool for quantitative analysis of-omics data. *Bioinformatics* 28, 2404–2406. doi: 10.1093/bioinformatics/bts449

Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5:93. doi: 10.1038/nprot.2009.203

Thompson, R. A., Dahal, S., Garcia, S., Nookaew, I., and Trinh, C. T. (2016). Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of *Clostridium thermocellum* DSM 1313 implementing an adjustable cellulosome. *Biotechnol. Biofuels* 9:194. doi: 10.1186/s13068-016-0607-x

Thompson, R. A., Layton, D. S., Guss, A. M., Olson, D. G., Lynd, L. R., and Trinh, C. T. (2015). Elucidating central metabolic redox obstacles hindering ethanol production in *Clostridium thermocellum*. *Metab. Eng.* 32, 207–219. doi: 10.1016/j.ymben.2015.10.004

Thompson, R. A., and Trinh, C. T. (2017). Overflow metabolism and growth cessation in *Clostridium thermocellum* DSM1313 during high cellulose loading fermentations. *Biotechnol. Bioeng.* 114, 2592–2604. doi: 10.1002/bit.26374

Tian, L., Papanek, B., Olson, D. G., Rydzak, T., Holwerda, E. K., Zheng, T., et al. (2016). Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. *Biotechnol. Biofuels* 9:116. doi: 10.1186/s13068-016-0528-8

Trinh, C. T. (2012). Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.* 95, 1083–1094. doi: 10.1007/s00253-012-4197-7

Trinh, C. T., Liu, Y., and Conner, D. J. (2015). Rational design of efficient modular cells. *Metab. Eng.* 32, 220–231. doi: 10.1016/j.ymben.2015.10.005

Trinh, C. T., Wlaschin, A., and Srienc, F. (2009). Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.* 81, 813–826. doi: 10.1007/s00253-008-1770-1

Wallenius, J., Viikilä, M., Survase, S., Ojamo, H., and Eerikäinen, T. (2013). Constraint-based genome-scale metabolic modeling of *Clostridium acetobutylicum* behavior in an immobilized column. *Bioresour. Technol.* 142, 603–610. doi: 10.1016/j.biortech.2013.05.085

Wang, L., and Maranas, C. D. (2018). Mingenome: an *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synth. Biol.* 7, 462–473. doi: 10.1021/acssynbio.7b00296

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., et al. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nat. Chem. Biol.* 7:445. doi: 10.1038/nchembio.580

Yoo, M., Bestel-Corre, G., Croux, C., Riviere, A., Meynial-Salles, I., and Soucaille, P. (2015). A quantitative system-scale characterization of the metabolism of *Clostridium acetobutylicum*. *MBio* 6:e01808-15. doi: 10.1128/mBio.01808-15

Zhang, Y.-H. P., and Lynd, L. R. (2005). Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7321–7325. doi: 10.1073/pnas.0408734102

Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., van Gulik, W. M., van Dijken, J. P., et al. (2013). Atypical glycolysis in *Clostridium thermocellum*. *Appl. Environ. Microbiol.* 79, 3000–3008. doi: 10.1128/AEM.04037-12

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

REVIEW
published: 14 October 2020
doi: 10.3389/fbioe.2020.539902

Multi-Faceted Systems Biology Approaches Present a Cellular Landscape of Phenolic Compound Inhibition in *Saccharomyces cerevisiae*

*Eugene Fletcher and Kristin Baetz**

Ottawa Institute of Systems Biology, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada

OPEN ACCESS

Edited by:
Chris Petzold, Lawrence Berkeley National Laboratory, United States

Reviewed by:
Guoqiang Xu, Jiangnan University, China
Sastia Prama Putri, Osaka University, Japan

***Correspondence:**
Kristin Baetz
kbaetz@uottawa.ca

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *02 March 2020*
Accepted: *02 September 2020*
Published: *14 October 2020*

Citation:
Fletcher E and Baetz K (2020) Multi-Faceted Systems Biology Approaches Present a Cellular Landscape of Phenolic Compound Inhibition in Saccharomyces cerevisiae. Front. Bioeng. Biotechnol. 8:539902. doi: 10.3389/fbioe.2020.539902

Synthetic biology has played a major role in engineering microbial cell factories to convert plant biomass (lignocellulose) to fuels and bioproducts by fermentation. However, the final product yield is limited by inhibition of microbial growth and fermentation by toxic phenolic compounds generated during lignocellulosic pre-treatment and hydrolysis. Advances in the development of systems biology technologies (genomics, transcriptomics, proteomics, metabolomics) have rapidly resulted in large datasets which are necessary to obtain a holistic understanding of complex biological processes underlying phenolic compound toxicity. Here, we review and compare different systems biology tools that have been utilized to identify molecular mechanisms that modulate phenolic compound toxicity in *Saccharomyces cerevisiae.* By focusing on and comparing functional genomics and transcriptomics approaches we identify common mechanisms potentially underlying phenolic toxicity. Additionally, we discuss possible ways by which integration of data obtained across multiple unbiased approaches can result in new avenues to develop yeast strains with a significant improvement in tolerance to phenolic fermentation inhibitors.

Keywords: systems biology, synthetic biology, yeast, phenolic inhibitors, fermentation, metabolism, biomanufacturing

INTRODUCTION

Biomanufacturing is transforming how new and existing platform chemicals are made in a way that is environmentally friendly, renewable, and sustainable. To make bio-derived chemicals competitive to fossil-derived chemicals, high productivity and cost reduction are a major consideration. Therefore, there has been a growing interest in using cheap and readily available feedstocks, such as plant material (lignocellulose) obtained from agricultural and forestry wastes.

Lignocellulose is an abundant and ubiquitous biomass feedstock that can be hydrolyzed to yield simple sugars which are fermented by yeast to produce bioethanol, fine chemicals, and other

bioproducts (Abo Bodjui et al., 2019). However, converting lignocellulose to these products during biomanufacturing has its challenges. The sugars in lignocellulose exist as long polysaccharide chains in the form of cellulose and hemicellulose which are held together by lignin (Becker and Wittmann, 2019). In order to make the cellulose and hemicellulose polymers accessible for hydrolysis to release sugars for fermentation, a pre-treatment step is required to dissolve the lignin fibers holding the sugar polymers. While physical (pyrolysis), physicochemical (ammonia fiber explosion) and biological methods exist for pre-treating lignocellulose, chemical pre-treatment methods are commonly used since they are simple and efficient (Becker and Wittmann, 2019). Chemical pre-treatment involves the use of dilute acid or alkali to break down the lignin. As a result, phenolic compounds which are monomeric subunits of lignin are produced during the pre-treatment step (Palmqvist and Hahn-Hägerdal, 2000). Phenolic compounds inhibit enzymes used to hydrolyze cellulose (Qin et al., 2016) and in effect, limit the amount of sugars available for fermentation. Phenolics are also extremely toxic to yeast even in minute quantities and significantly inhibit yeast growth and fermentation (Ando et al., 1986; Adeboye et al., 2014) thus, reducing the product yield and increasing the cost of fermentation.

Phenolic compounds exist in different forms in lignocellulosic hydrolysates as phenolic acids (e.g., ferulic acid), phenolic aldehydes (e.g., vanillin), phenolic ketones (e.g., 4-hydroxyacetophenone), and phenolic alcohols. The concentrations of each of these compounds in hydrolysates vary depending on the plant material and the pre-treatment method used. They appear to have different toxic effects on the cell with phenolic aldehydes being the most toxic and completely inhibit yeast growth at concentrations as low as 1 and 5 mM for coniferyl aldehyde and vanillin, respectively (Adeboye et al., 2014). The different levels of toxicity of multiple phenolic compounds were confirmed in a study which showed that the chemical nature of phenolic compounds determine their toxicity and the physiological impact they have on the cell (Adeboye et al., 2014). This study was backed by another report that demonstrated that ferulic acid and coniferyl aldehyde, though structurally similar with the only difference being the functional group, presented very distinct chemogenomic profiles and inhibited yeast growth using specific mechanisms (Fletcher et al., 2019).

Apart from converting lignocellulosic materials to bioethanol and other chemicals by fermentation, there has been a recent interest in valorizing lignin in lignocellulose to produce precursors and final products for the fine chemicals industry (Becker and Wittmann, 2019; Li et al., 2019; Ponnusamy et al., 2019). Vanillin is an example of a valuable phenolic compound in the fine chemicals industry mainly used as flavor or scent in food, pharmaceuticals and cosmetics (Luziatelli et al., 2019). While a process has been developed for fermenting glucose to vanillin (Brochado et al., 2010), ferulic acid is an important precursor which can be converted to vanillin by engineering microbial cell factories to express feruloyl-CoA synthase and feruloyl-CoA hydratase (Luziatelli et al., 2019). Other phenolic compounds, such as eugenol present in grains and cereals can be converted to ferulic acid and subsequently to vanillin (Overhage et al., 2002; Di Gioia et al., 2009). Again, a major limitation of using engineered yeasts for ferulic acid conversion to vanillin is the issue of toxicity of both vanillin and its ferulic acid precursor. It is possible to remove phenolic compounds from lignocellulosic hydrolysates as they form to prevent toxicity to the yeast cells (Carter et al., 2011; Xue et al., 2018) but this comes at an extra manufacturing cost. Therefore, to cost-effectively achieve high yields of bioethanol and other bioproducts from lignocellulose by fermentation, there is the need to improve tolerance to phenolic fermentation inhibitors in yeast cell factories that are used for the bioconversion.

A thorough understanding of the mechanisms that modulate phenolic compound toxicity is required to engineer yeast strains that are tolerant to individual phenolic compounds and/or a complex mix of phenolics found in hydrolysates. As inhibitor tolerance is a multigenic complex trait (de Witt et al., 2019) global cellular approaches are required to identify key determinants associated with phenolic compound tolerance. Advances in systems biology approaches have revolutionized our ability to assess how cells respond to phenolic toxicity. The use of genome-wide approaches have given insight into how the cell responds to individual phenolics and identified genetic and metabolic targets that can be engineered to improve tolerance to toxic phenolic fermentation inhibitors. However, a comprehensive understanding of the phenolic tolerance pathway remains lacking since data from the individual studies have not been fully integrated.

Here, we review several unbiased functional genomics and transcriptomic approaches to identify general and specific genetic targets that modulate phenolic compound toxicity in *S. cerevisiae*. We also highlight the potential of exploiting proteomics and metabolomics approaches, which remain underutilized in the field. Finally, synthetic biology approaches and future developments that can rapidly be used to generate yeast tolerant to phenolic fermentation inhibitors are discussed.

FUNCTIONAL GENOMIC APPROACHES

Improving yeast tolerance to phenolic compounds first requires the identification of genes and pathways that can be engineered to confer increased tolerance. Therefore, functional genomic tools including chemogenomic screens, adaptive laboratory evolution, genome shuffling, and high content imaging can be exploited to discover genes associated with biological processes underlying phenolic tolerance in yeast.

Chemical Genomics

The availability of both haploid and diploid deletion mutant collections, in which the majority of yeast open reading frames have been systematically deleted has made it possible to conduct chemical profiling or chemogenomic screens (reviewed in Giaever and Nislow, 2014). In agar-based array screens, the deletion mutant library is pinned onto solid media containing sub-lethal concentrations of the compound(s) being tested. Following incubation, colony sizes of the mutant strains on

the selection plates vs. control plates are quantified to obtain fitness scores for all the mutants. Mutants that lack genes required for growth in the presence of the compound show a significant growth defect and are hypersensitive to the compound. Mutants that are sensitive to a compound aid in the identification of proteins and pathways needed for survival upon exposure to inhibitors. On the other hand, mutants that grow better than the wild type in the presence of the compound are called "suppressors." Though it is harder to generalize the mechanism(s) by which suppressors when deleted confer protection to a compound, one possibility is the suppressor protein increases toxicity of the compounds. An example of a suppressor gene is *BNA7* which was found to enable yeast growth, when deleted, in media containing ferulic acid (Fletcher et al., 2019). Interestingly, though Bna7 has a well-established role in the tryptophan catabolic pathway, no other components of this pathway when deleted conferred improved tolerance to ferulic acid.

Presently, four phenolics (coniferyl aldehyde, ferulic acid, 4-hydroxybenzoic acid, and vanillin) (Endo et al., 2008; Fletcher et al., 2019) have been screened through agar based methods to identify their chemogenomic profiles (**Figure 1A** and **Table 1**). The highest overlap in chemogenomic profiles (or shared gene "hits") was found between ferulic acid and vanillin with 17 common deletion mutant genes with hypersensitivity shared between these two compounds, which is not unexpected considering vanillin is derived from ferulic acid. Most of these common genes clustered into biological processes [according to Gene Ontology enrichment analysis (Robinson et al., 2002)] mainly associated with protein transport (*COG6*, *COG7*, and *ARL1*), chromatin modification and transcription (*SWC3, ARP6, YAF9*, *HTZ1*) (**Figure 1B**; Endo et al., 2008; Fletcher et al., 2019). These biological processes serve as interesting targets for engineering a vanillin-producing yeast strain that uses ferulic acid as a precursor since tolerance to both phenolics will be required in such a strain. Also, wheat straw hydrolysate and synthetic miscanthus hydrolysate which contain a complex mixture of several phenolics has been screened for yeast tolerance (Skerker et al., 2013; Pereira et al., 2014; **Table 1**). Genes involved in protein synthesis (*RPL13B, RPL13A*), ergosterol biosynthesis (*ERG2*) and oxidative stress response (*SOD1, SOD2*) were among the top hits that came up in the screen (Skerker et al., 2013; Pereira et al., 2014). Since the yeast libraries contain over 4,000 non-essential mutants, the use of robotics in performing genome-wide screens is now taking center stage as it allows the screens to be performed rapidly and simply.

In addition, agar-based screens have also been used to perform focused screens. For example, a yeast deletion mutant array composed of 30 yeast transcription factor mutants has been screened to provide insight into phenolic-induced transcriptional changes. The study revealed that genes encoding the transcription factors *YAP1*, *DAL81, GZF3, LEU3, PUT3*, and *WAR1* were required by yeast to grow in coniferyl aldehyde (Wu et al., 2017). This approach provides preliminary information on transcription factors that can be engineered to concurrently regulate the expression of several genes associated with phenolic tolerance instead of engineering the individual genes they regulate. However, studies of this nature are limited by the size and composition of the mini-array being screened.

The yeast deletion libraries are barcoded with a unique 20 bp sequence placed upstream (uptag) and downstream (dntag) of the KanMX selection marker gene used to replace the gene of interest (Winzeler et al., 1999). Genetic barcoding is a powerful tool for even more complex fitness profiling of the mutant collection where thousands of yeast mutants are pooled, and grown in liquid media containing an inhibitor and analyzed in parallel (Shoemaker et al., 1996). Coupled to next generation sequencing (NGS), the amount of barcoded PCR product representing each mutant can be quantified to identify mutants with high tolerance to the inhibitor being tested (Smith et al., 2009). Genetic barcoding has the advantage of allowing screening of complex phenolic mixtures and plant hydrolysates (Skerker et al., 2013; **Table 1**). Furthermore, this method is useful for identifying suppressors since tolerant mutants outgrow the other strains in the mutant pool and are selected at the end of the experiment. For example, genetic barcoding was used to identify suppressor genes involved in fatty acid metabolism (*EEB1*) and vesicle trafficking (*SSH4* and *VAM6*) as important for conferring tolerance to a mixture of phenolics in a synthetic hydrolysate (Xue et al., 2018; **Table 1**). Similarly, by screening and sequencing a pooled yeast deletion mutant library, suppressor genes (*SUR1, NBP2, DFG1*) whose deletion resulted in tolerance to poacic acid were identified (Piotrowski et al., 2015; **Table 1**).

Although chemogenomic screens, serve as a powerful tool for identifying genes associated with phenolic compound tolerance (sensitive mutants), it remains a challenge to identify suppressors by either method. This is likely because in most cases the growth improvement of the suppressors is small at the sub-lethal concentrations these screens have been performed at. However, success at identifying suppressors can be improved by performing parallel chemical genomic screens at multiple dosages. Further, chemogenomic screens are limited in that they only screen the impact of loss of open reading frames, hence this type of screen excludes gain or separation of function mutations and mutations in regulatory regions. Again, current screens have only probed the non-essential genes, and have not probed the essential mutant collections. Plus, chemogenomic screens are not ideal for selecting tolerance phenotypes that are as a result of epistatic interactions between multiple genes since it only determines the effect of single-gene deletions or mutations.

To complement the chemogenomic method where yeast deletion libraries are screened, overexpression libraries, such as the MoBY collection (Hou, 2009) can be probed to identify genes whose overexpression result in tolerance to phenolic compounds. Even though this method has not yet been applied to phenolic tolerance, it has been demonstrated that by screening an overexpression library, a multi efflux pump, *SGE1*, was found to improve yeast tolerance to a yeast growth inhibitor used as a pre-treatment solvent (Higgins et al., 2018).

Adaptive Laboratory Evolution

As an alternative approach, other studies have used adaptive laboratory evolution to point out key driver mutations that

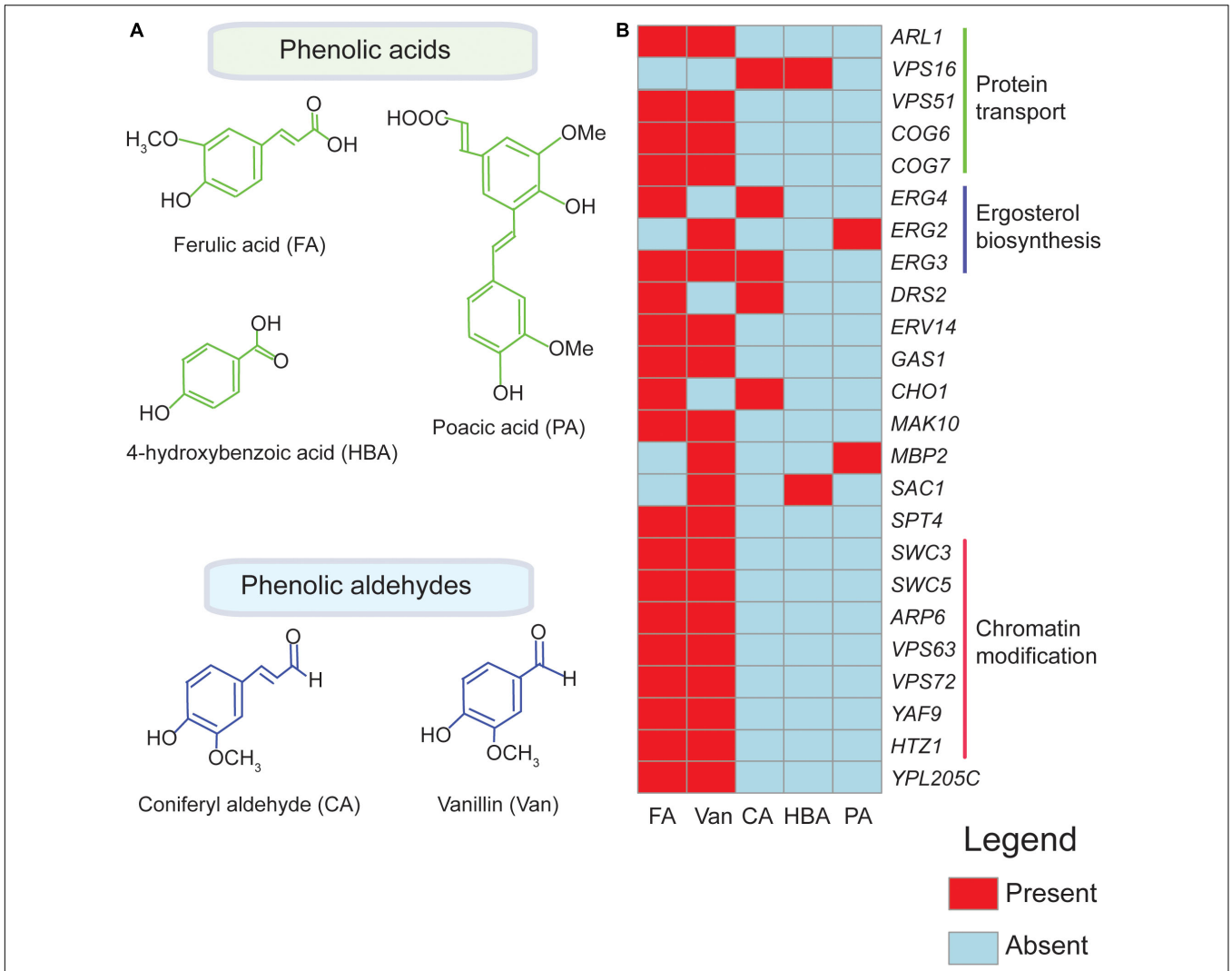


FIGURE 1 | Genetic hits for different phenolics in functional genomics screens overlap. Chemical structures of phenolic compounds ferulic acid (FA), poacic acid (PA), coniferyl aldehyde (CA), 4-hydroxybenzoic acid (HBA), and vanillin (Van.) are shown in **(A)**. Overlapping biological processes of the deletion mutants identified in chemogenomic screens for FA, PA, CA, HBA, and Van are indicated in **(B)** The heatmap shown here is based on published data obtained from Endo et al. (2008), Skerker et al. (2013), Pereira et al. (2014), Piotrowski et al. (2015), Wang X. et al. (2017), Wu et al. (2017), Biot-Pelletier et al. (2018), Sardi et al. (2018), Xue et al. (2018), Fletcher et al. (2019), and Hacısalihoğlu et al. (2019). The genes were clustered according to their associated biological process (Robinson et al., 2002).

are essential to increase tolerance to phenolic compounds. In adaptive laboratory evolution, yeast cultures containing mild concentrations of the phenolic compound are serially transferred into fresh media supplemented with increasing concentrations of the phenolic compound until a significant improvement in growth rate is observed after several generations and serial transfers (Dragosits and Mattanovich, 2013). Advances in whole-genome sequencing technologies with regards to reduction in costs, new and improved sequence analysis tools (Sandmann et al., 2017) and sequencing platforms (Rhoads and Au, 2015; Tyler et al., 2018) have made it possible to identify mutations that lead to improved growth in the presence of phenolic compounds.

Adaptive mutations can occur in the regulatory regions or coding regions of the target gene and result in loss of function, increased activity, or decreased dosage of the gene product (Dragosits and Mattanovich, 2013). For example, the zinc finger transcription factor, *YRR1* acquired a frameshift mutation resulting in a loss of function which improved yeast growth in vanillin (Wang X. et al., 2017; **Table 1**). The role of *YRR1* in vanillin tolerance was confirmed by deleting the gene (Wang X. et al., 2017). In another example, nonsense mutations acquired by *MUK1* and *MRS4* resulted in tolerance to coniferyl aldehyde (Hacısalihoğlu et al., 2019). The challenge with using adaptive laboratory evolution, though, is most times several mutations unrelated to the compound tolerance phenotype will arise making it challenging to pinpoint the actual mutations that are required for tolerance. It is possible to confirm each mutation gained in a laboratory evolution experiment but in instances where several mutations arise, a considerable amount

TABLE 1 | Functional genomic strategies to elucidate yeast tolerance to phenolic inhibitors.

Screening tool	Fermentation inhibitor	Significant biological processes	References
Chemogenomic screen (agar-based array)	Ferulic acid	Protein and vacuolar trafficking Ergosterol biosynthesis	Fletcher et al., 2019
Chemogenomic screen (agar-based array)	4-hydroxybenzoic acid	Ergosterol biosynthesis Protein trafficking	Fletcher et al., 2019
Chemogenomic screen (agar-based array)	Coniferyl aldehyde	Pentose phosphate pathway	Fletcher et al., 2019
Chemogenomic screen (agar-based array)	Vanillin	Ergosterol biosynthesis Histone exchange	Endo et al., 2008
Chemogenomic screen (agar-based array)	Synthetic miscanthus hydrolysate	Ergosterol biosynthesis Oxidative stress response Pentose phosphate pathway	Skerker et al., 2013
Chemogenomic screen (well-by-well array)	Wheat straw hydrolysate	Vacuolar acidification Ribosome biogenesis Mitochondrial and peroxisomal function Ergosterol biosynthesis	Pereira et al., 2014
Chemogenomic screen (barcode sequencing of pooled mutants)	Poacic acid	Cell wall and glycosylation	Piotrowski et al., 2015
Chemogenomic screen (barcode sequencing of pooled mutants)	Synthetic corn stover hydrolysate	Fatty acid biosynthesis Vesicle trafficking	Xue et al., 2018
Adaptive laboratory evolution	Coniferyl aldehyde	Vacuolar transport Mitochondrial function Transcriptional regulation	Hacısalihoğlu et al., 2019
Adaptive laboratory evolution	Vanillin	Transcriptional regulation	Wang H.-Y. et al., 2017; Wang X. et al., 2017
Genome shuffling	Lignocellulosic hydrolysate	Transcriptional regulation	Biot-Pelletier et al., 2018
Genome-wide association studies	Synthetic corn stover hydrolysate	Ergosterol biosynthesis Proteolysis	Sardi et al., 2018
Chemogenomic screen (well-by-well array)	Coniferyl aldehyde	Membrane transport Oxidative stress response	Wu et al., 2017

of time is required to validate the effect of all the mutations in a wild type strain. One method to circumvent this challenge is to perform multiple parallel screens (Fletcher et al., 2017). High throughput evolution of several lines in parallel facilitated by robotics, automation and mutational analysis makes it possible to quickly identify common suppressors over parallel evolutions (Radek et al., 2017) but this has not yet been applied to phenolic screens.

Genome-Shuffling

Laboratory evolution can be extended in another approach called genome shuffling to find novel genes that can be modulated for increased phenolic tolerance. Genome shuffling allows the discovery of positive epistasis and the accumulation of beneficial mutations. The technique involves performing mutagenesis on haploid yeasts of both mating types (a and α), selecting for tolerant haploids and mating them to obtain diploid strains (Hou, 2009; Pinel et al., 2011). The diploids are screened on media containing increasing concentrations of the phenolic compound after which the most tolerant diploids are isolated. The diploids with increased tolerance then undergo a new round of sporulation, mating and selection on increasing concentrations of phenolic compounds. Several cycles of "genome-shuffling" are performed to generate yeast strains with improved tolerance. The resulting strains are sequenced to identify key mutations that can then re-constructed in the wild type to confirm the role of the selected mutations on increasing tolerance. Using genome shuffling, genes including *NRG1*, *GSH1*, and *GDH1* were identified as key determinants required by yeast for improved tolerance to lignocellulosic hydrolysate (Biot-Pelletier et al., 2018; **Table 1**). Although genome shuffling has the advantage of filtering mutations that are unrelated to phenolic tolerance, it can be laborious and time consuming. Currently, the technology is challenged by the lack of high throughput screening methods (Magocha et al., 2018). As such, there is the need to automate the process to make it rapid and more efficient in identifying novel genetic mutations that are linked with improved tolerance to phenolic inhibitors in yeast.

High Content Imaging

Another emerging technology is the use of a high content, image-based profiling to identify biological processes that are targeted by toxic compounds (Ohnuki et al., 2010). Here, it is assumed that changes in yeast morphology as a result of a chemical treatment will resemble the morphology of mutants that lack genes associated with biological processes that are inhibited by the chemical. High content imaging was used to identify genes associated with toxicity to vanillin by comparing the altered morphology of >4,000 yeast mutants to that of wild type yeast treated with vanillin (Iwaki et al., 2013). Using this technique, 18 mutants with an altered morphology that mimic the morphology of vanillin-treated cells were identified. Out of the 18 mutants, three mutants (*rpl8a*Δ, *rpp1b*Δ, and *rpl16a*Δ) that had the closest resemblance to vanillin-treated cells lacked genes belonging to the gene ontology (GO) term "cytoplasmic protein component of the large ribosomal subunit"(Iwaki et al., 2013). The outcome of the study indicates that vanillin toxicity may be due to inhibition of large ribosomal subunit leading to an impairment of protein synthesis.

Though useful in elucidating novel mechanisms of toxicity of phenolic compounds, a limitation of cell imaging is that it is not suitable for screening phenolic compounds that do not induce

morphological changes. An extension of this technology will be the development of high throughput microscopy screening of the yeast GFP collection (Ghaemmaghami et al., 2003) to determine the impact of phenolic treatment on protein subcellular localization and abundance (Koh et al., 2015). For instance, upon coniferyl aldehyde treatment, several enzymes in the pentose phosphate pathway are both induced and partially localize to the mitochondria (Fletcher et al., 2019). It will be interesting to fully screen the yeast GFP collection to identify other proteins that change localization upon exposure to different phenolic compounds. Not only will such a screen provide an idea of what proteins are induced and change localization but it will also bring to light novel and alternative mechanisms of regulation where changes in protein localization render a pathway active or inactive due to unavailability of upstream intermediates.

TRANSCRIPTOMIC APPROACHES

Exploring transcriptomic changes during exposure to phenolics provides another dimension to obtaining a holistic view of the phenolic tolerance landscape of yeast. Historically, microarray technology has been used to observe the expression of thousands of genes simultaneously to obtain a gene expression profile of cells under a given condition (Raghavachari, 2013). More recently, developments in high throughput RNA sequencing (RNA-Seq) has made it possible to quantify transcriptomes by measuring mRNA transcripts (Sardi et al., 2018). Unlike, microarrays which are limited by the genes included on the array, RNA-Seq allows for the identification of novel transcriptomic changes including alternative splice variants, novel genes and small mRNA sequences (Han et al., 2015).

Presently, two distinct transcriptomics approaches have been applied to unravel the biology of phenolic tolerance in yeast (**Table 2**). In one strategy, the transcriptome of phenolic-adapted strains are compared to that of un-adapted strains. Here, the goal is to identify the genes whose expression is modulated in the adapted strain as these genes and their associated biological pathways potentially confers tolerance to the phenolic inhibitor. In the second strategy, the transcriptome profile of yeast exposed to phenolic inhibitors are compared to that of an un-treated yeast. The goal of these experiments is to identify changes in gene expression upon phenolic exposure as these genes and their associated biological pathways may contribute to protecting the cell from phenolic toxicity.

Remarkably, yeast strains evolved for vanillin tolerance in two independent studies displayed a similar transcriptome profile (Endo et al., 2008; Shen et al., 2014). Subunits of the cytochrome b-c1 complex (*QCR2, QCR10, QCR6*, and *CYT1*), which are components of the mitochondrial electron transport chain, together with an electron donor to the mitochondrial electron transport chain (*CYC1*), were upregulated in both vanillin-tolerant yeasts. This suggests an induction in aerobic respiration and energy generation in vanillin adapted strains is critical to confer tolerance (a summary of the transcriptome data comparing published transcriptome studies is provided in **Supplementary Table 1**). Furthermore, more than half of the upregulated genes that overlap in vanillin- (Endo et al., 2008; Shen et al., 2014) and coniferyl aldehyde- (Hacısalihoğlu et al., 2019) adapted strains are involved in oxidation-reduction processes and NADPH production (*BDH2, CTT1, COX5B, SDH1, IDP2, CYB2, NDI1, ALD3, COX7, SPS19, ALD4*, and *ALD6*). Additionally, *FAA1, PRS3*, and *ALD5* were repressed in both vanillin-tolerant and coniferyl aldehyde-tolerant strains under non-stress conditions. The similarity in gene expression profiles in coniferyl aldehyde- and vanillin-adapted yeast from independent studies suggest that yeast utilizes common mechanisms to build tolerance to these compounds. Knowledge of these commonly induced pathways in adapted strains could be exploited to further improve phenolic tolerance.

Given the similarity in transcriptomes from adapted strains, it is somewhat surprising that transcriptome profiles of wild-type yeast exposed to phenolic compounds (ferulic acid, coniferyl aldehyde, vanillin, isoeugenol, and plant hydrolysates composed of a combination of these phenolic compounds) during growth share limited common features (Sundström et al., 2010; Park and Kim, 2014; Thompson et al., 2016). In these studies, apart from the mitochondrial potassium homeostasis gene *YLH47* that was upregulated in ferulic acid, coniferyl aldehyde and isoeugenol-treated cells, no particular set(s) of genes overlapped in all the studies. Though direct comparisons between transcriptomic studies have limitations (Larsson and Sandberg, 2006), these transcriptome studies suggests that the cell's transcriptional response is largely distinct for each phenolic study so far.

Interestingly, four genes that are upregulated in the transcriptome of vanillin-adapted strains obtained under no stress (Shen et al., 2014) are also upregulated in the transcriptome of un-adapted yeasts treated with vanillin (Park and Kim, 2014). All four genes (*CIT1*, *LSC2*, *SDH1*, *SDH2*) encode enzymes in the TCA cycle. Similarly, genes involved in oxidation-reduction (*YML131W, YKL071W*, and *OYE3*), transport (*SNQ2*) and response to oxidative stress (*SRX1*) were upregulated in both coniferyl aldehyde-adapted yeasts (non-stressed conditions) (Hacısalihoğlu et al., 2019) and in un-adapted yeasts treated with coniferyl aldehyde (Sundström et al., 2010).

Taken together, oxidation-reduction, electron transfer chain, and the TCA cycle are enriched in the transcriptome of yeast during phenolic toxicity suggesting an upregulation of mitochondrial activity during phenolic stress (**Supplementary Table 1**). A broader overview of transcriptome changes induced by phenolics in yeast studies are limited by the number of phenolics studied. There is the need to expand these studies to include a wide range of phenolic compounds to ascertain the effect of various phenolic compounds on the yeast transcriptome as a way of identifying potential biological processes that can be targeted to improve phenolic tolerance in yeast.

While transcriptomics can identify gene targets that when overexpressed or downregulated can improve phenolic tolerance, this technology is challenged by the fact that changes in the expression of most genes do not correlate with improved tolerance to yeast stress (Evans, 2015). Hence transcriptional

TABLE 2 | Transcriptomic profiling of yeast to elucidate cellular responses to phenolic inhibitors.

Screening tool	Fermentation inhibitor	Significant biological processes	References
Microarray analysis of a tolerant strain (non-stressed conditions)	Vanillin	Ergosterol biosynthesis Mitochondrial function	Endo et al., 2009
Microarray analysis of a tolerant strain (non-stressed conditions)	Coniferyl aldehyde	Oxido-reductase activity Oxidative stress response	Hacısalihoğlu et al., 2019
Microarray analysis of a tolerant strain (non-stressed conditions)	Vanillin	Oxido-reductase activity Oxidative stress response	Shen et al., 2014
Microarray analysis of a tolerant strain (non-stressed conditions)	Vanillin	Response to stress Phospholipid metabolism	Wang X. et al., 2017
Microarray analysis of evolved strain (stressed conditions)	Softwood hydrolysate	Oxidative stress response Membrane transport	Thompson et al., 2016
Microarray analysis of wild type strain (stressed conditions)	Vanillin	TCA cycle Aerobic respiration	Park and Kim, 2014
Microarray analysis of wild type strain (stressed conditions)	Coniferyl aldehyde	Oxido-reductase activity Mitochondrial function	Sundström et al., 2010
Microarray analysis of wild type strain (stressed conditions)	Ferulic acid	Protein import Mitochondrial function	Sundström et al., 2010
Microarray analysis of wild type strain (stressed conditions)	Isoeugenol	Mitochondrial function	Sundström et al., 2010
Microarray analysis of *yrr1*Δ strain (stressed conditions)	Vanillin	Ribosome biogenesis rRNA processing	Wang X. et al., 2017

profiling, while providing a holistic snapshot of the yeast's response to a phenolic, may not provide a direct entry point into genetic engineering strains for phenolic tolerance improvement.

OTHER SYSTEMS BIOLOGY APPROACHES

Though functional genomics and transcriptomic studies have so far dominated the field of yeast phenolic tolerance, they clearly do not capture all the biological events that occur upon exposure to phenolics. While functional genomics identify proteins and pathways required for survival upon exposure to phenolics, it fails to assess how these proteins are regulated and their biological role in phenolic tolerance. Gene expression modulation during phenolic compound stress serves as a tangible way of quantifying induction and repression of genes associated with phenolic toxicity. However, RNA levels can only be used as a proxy for measuring products of expressed genes within the cell and may not reflect protein levels, protein function or modification of proteins by post-translational modifications. Further, neither functional genomics nor transcriptomics can assess how phenolic exposure modifies a cell's metabolism. Hence, two emerging systems biology approaches worth highlighting that can provide this extra layer of genome-wide information are proteomics and metabolomics.

PROTEOMIC PROFILING OF YEAST PHENOLIC TOLERANCE

Shotgun Proteomics

One approach to capture protein changes in the cells is the shotgun proteomic method which involves digesting total cellular proteins (isolated from cells treated with or without a toxic compound) into peptides which are separated by liquid chromatography followed by identification and quantification using mass spectrometry (Zhang et al., 2013). Beyond identifying differential changes in protein expression, post-translational modification sites can be identified using quantitative methods, such as stable isotope labeling by/with amino acid in cell culture (SILAC) (Ong et al., 2004).

To date, very few studies have probed the proteomic profile of yeast upon phenolic exposure, using shotgun proteomics tools. A proteomic study quantified protein expression in two natural isolates of *S. cerevisiae* that exhibited remarkable tolerance to a synthetic inhibitor cocktail containing ferulic acid, cinnamic acid, and coniferyl aldehyde (de Witt et al., 2018). Their proteomic profile revealed a general tolerance mechanism which mainly included genes associated with oxido-reductase activity (de Witt et al., 2018). In another proteomic study, expression of oxidative stress response proteins (Ahp1 and Grx1) was found to be induced during yeast growth in a combination of three inhibitors which include phenol (Ding et al., 2012a). So far, the application of shotgun proteomics to understanding phenolic tolerance has been limited to quantifying protein expression. Future work should investigate post-translational modification of the most differentially expressed proteins during yeast growth in different phenolic compounds.

METABOLOMICS PROFILING OF PHENOLIC FERMENTATION INHIBITORS

Comprehensive analysis of metabolites during cellular stress is gaining popularity as another strategy to understand tolerance mechanisms (Nugroho et al., 2015). Therefore, metabolomics tools are being developed to obtain a cell's metabolic profile

or metabolome which directly reflects the cell's metabolic state (Zampieri et al., 2017).

Mass Spectrometry-Based Metabolomics

Developments in mass spectrometry-based metabolomics have enabled quantification of metabolites even at low concentrations with high resolution and dynamic range (Marshall and Powers, 2017). Extending this technology to understanding the underlying basis of phenolic toxicity in yeast, a metabolic shift between a parental yeast and an inhibitor-tolerant yeast was observed during growth in a mixture of inhibitors which included phenol (Ding et al., 2012b). The mixed inhibitors induced the production of myo-inositol and phenylamine in the tolerant yeast suggesting regulation of membrane trafficking and cytosolic Ca^{2+} concentration, respectively. Remarkably, glycolysis and TCA cycle intermediates including citrate, succinate, and 2-oxoglutarate were decreased in the tolerant strain during growth in the mixed inhibitors (Ding et al., 2012b). It will be interesting to further explore the effect of a wide range of phenolic compounds on changes in yeast metabolomic profiles using untargeted metabolomic tools followed by a targeted approach to confirm any observed metabolic shifts.

MINING GENOME-WIDE STUDIES TO IDENTIFY COMMON APPROACHES TO IMPROVE TOLERANCE TO PHENOLICS

The different systems biology tools discussed above have highlighted several biological processes associated with phenolic tolerance. While proteomic and metabolomics data is currently limited for phenolics, the functional genomics and transcriptomics data can be used to identify common mechanisms underlying phenolic toxicity. By targeting common mechanisms it may be possible to engineer strains with improved tolerance toward all phenolic compounds. Four common cellular responses to phenolic exposure that have been identified from the functional genomics and transcriptomics data are: oxidative stress response, oxido-reductase and mitochondrial activity, ergosterol biosynthesis, and membrane transport.

Oxidative Stress Response

Functional genomic screens determined that deletion of genes involved in oxidative stress response (*YAP1, STB5, GSH1, SOD1, SOD2*) resulted in hypersensitivity to phenolics (Skerker et al., 2013; Pereira et al., 2014; Wu et al., 2017; Biot-Pelletier et al., 2018). Also, *YAP1, SOD2* and other genes with antioxidant activity (*GRX2, CTT1, CTA1*) are upregulated during phenolic exposure as shown in the transcriptomic studies (Park and Kim, 2014; Shen et al., 2014). Furthermore, Grx1 and Ahp1 which are oxidative stress response genes are differentially enriched in proteomic studies when yeasts are treated with phenolics (Ding et al., 2012a) suggesting the production of reactive oxygen species (ROS). Together, this suggests that phenolic exposure elicits an oxidative stress response in yeast.

Experimental evidence shows that phenolic compounds, particularly those with an aldehyde functional group, such as vanillin and coniferyl aldehyde induce ROS formation (Nguyen et al., 2014; Fletcher et al., 2019). ROS production comes from a combination of aerobic respiration and possibly from the endogenous process of oxidizing the compound to less toxic forms (Adeboye et al., 2015; Hacısalihoğlu et al., 2019). Yap1, a major oxidative stress transcription factor that regulates the expression of several genes responsible for scavenging ROS (Maeta et al., 2004; Rodrigues-Pousada et al., 2010) changes localization from the cytosol to the nucleus upon vanillin treatment where it activates transcription of its target genes (Nguyen et al., 2014). Not surprisingly, some genes regulated by Yap1 including *GPX1* and *SOD2* are significantly upregulated in a transcriptomic screen for vanillin (Shen et al., 2014). Additionally, yeast functional genomic study showed that *YAP1* deletion results in hypersensitivity to synthetic miscanthus hydrolysate which contains a mixture of phenolic compounds including coniferyl aldehyde (Skerker et al., 2013). Although it has been previously established that ferulic acid, 4-hydroxybenzoic acid and coniferyl aldehyde induce ROS formation in yeast, the specific species of ROS induced by different phenolic compounds is yet to be identified.

Catalases and the glutathione pathway which scavenge ROS from the cell require NADPH (Toledano et al., 2013; Gómez et al., 2019). Interestingly, not only are oxidative stress response proteins differentially expressed during treatment with phenolic inhibitors, as revealed in a proteomics screen, but also flux is moved toward the pentose phosphate pathway (Lv et al., 2014), one of the main metabolic pathways that generate cytosolic NADPH (Chen et al., 2019). Indeed, this proteomic study is corroborated by a functional genomic screen which showed that pentose phosphate pathway mutants accumulated ROS and were hypersensitive to coniferyl aldehyde (Fletcher et al., 2019). Moving forward, knowledge of the specific ROS species induced by individual phenolics will be beneficial in tailoring and engineering specific tolerance pathways for individual phenolic compounds.

Oxido-Reductase and Mitochondrial Activity

Coniferyl aldehyde, vanillin, synthetic hydrolysates, and softwood hydrolysate induce an upregulation of genes involved in oxidoreductase activity and mitochondrial function (Endo et al., 2009; Sundström et al., 2010; Shen et al., 2014; Thompson et al., 2016). Also, functional genomic studies identified other proteins with mitochondrial activity (*MRS4* and *AFG3*) that play a role in phenolic tolerance (Fletcher et al., 2019; Hacısalihoğlu et al., 2019). A proteomic screen identified several oxido-reductases including Adh7, Adh4, and Ald6 as the most differentially expressed proteins upon phenolic treatment (de Witt et al., 2018).

In the context of tolerance to phenolic inhibitors, the mitochondria are an important site for detoxification of phenolic compounds as enzymes, such as Ald5 and Pad1 that catabolize phenolic aldehydes are located in the mitochondria

(Adeboye et al., 2017). Detoxification of phenolics is a strategy employed by cells to prevent intracellular accumulation. Conversion of vanillin and coniferyl aldehyde to less toxic compounds catalyzed by enzymes with an oxido-reductase activity is upregulated during exposure to these compounds (Wang H.-Y. et al., 2017; Hacısalihoğlu et al., 2019). Furthermore, a gene encoding an NADH-dependent aldehyde reductase (*YLL056C*) is enriched in yeast exposed to coniferyl aldehyde (Sundström et al., 2010). While degradation of coniferyl aldehyde by Yll056c has not been reported, enzymatic reduction of other toxic aldehydes found in lignocelluloic hydrolysates (furfural and glycoaldehyde) has been demonstrated (Wang H.-Y. et al., 2017). Lastly, the NADPH-dependent alcohol dehydrogenase, Adh6, contributes to vanillin tolerance by converting intracellular amounts of vanillin to the less toxic vanillyl alcohol (Nguyen et al., 2015). Moving forward, it will be interesting to determine if mitochondrial function or content, for example increased mitochondrial volume per cell, can be engineered to improve phenolic tolerance.

Ergosterol Biosynthesis

Ergosterol biosynthesis genes (*ERG5, ERG26, ERG7, HMG1*, *ERG28*) have been reported to upregulated upon phenolic exposure in transcriptomic studies (Endo et al., 2009; Sardi et al., 2016). Confirming the role of ergosterol biosynthesis in phenolic tolerance, several functional genomics studies have shown that deletion of ergosterol biosynthesis genes result in a growth defect in the presence of individual phenolics and plant hydrolysates (Endo et al., 2008; Skerker et al., 2013; Piotrowski et al., 2015; Fletcher et al., 2019).

These data suggest that maintaining the proper levels of ergosterol is essential for growth in the presence of phenolic compounds. Indeed, increased levels of ergosterol are seen in vanillin-tolerant yeasts (Endo et al., 2009; Zheng et al., 2017) further confirming the need for ergosterol in the cell during exposure to phenolic compounds. Cellular ergosterol plays different roles in cells mainly by maintaining membrane integrity (Abe and Hiraki, 2009) and acting as components of lipid rafts (Eisenkolb et al., 2002; Bastos et al., 2012) which are possibly required for proper cellular function during growth in phenolics. Fine-tuning ergosterol levels and spatial localization to organelles may offer a unique way to buffer the toxic effects of phenolics.

Membrane Transport

Integration of systems biology tools reveal membrane efflux as another significant mechanism used by yeast as a survival strategy during growth in phenolic compounds. A functional genomic screen showed that the deletion of genes encoding membrane transporters, *PDR5, YOR1*, and *SNQ2* made *S. cerevisiae* sensitive to coniferyl aldehyde (Hacısalihoğlu et al., 2019). Also, *SNQ2* and another transporter, *MCH2*, were upregulated in transcriptomic studies during yeast growth in vanillin (Park and Kim, 2014; Wang X. et al., 2017).

Enrichment of genes encoding membrane transporters during exposure to vanillin hints that flushing out phenolic compounds from the cell prevents intracellular accumulation (Thompson et al., 2016; Wang X. et al., 2017; Hacısalihoğlu et al., 2019). Not surprisingly, Pdr1 which regulates the transcription of these transporters was identified during a screen of transcription factor mutants for coniferyl aldehyde tolerance (Wu et al., 2017). Export of phenolics from the cell complements efforts used by the cell to detoxify intracellular amounts of the compound. These transporters require ATP for activity (Katzmann et al., 1995; Mamnun et al., 2004). Hence, in order to meet the ATP needs of the cell, aerobic respiration *via* the TCA cycle and the mitochondrial electron transport chain are induced during growth in vanillin, coniferyl aldehyde and lignocellulosic hydrolysates as revealed in multiple studies (Endo et al., 2009; Park and Kim, 2014; Sardi et al., 2016; Thompson et al., 2016). Besides, since the transporters localize to the plasma membrane, changes in membrane composition and integrity have severe consequences on their activity (Kodedová and Sychrová, 2015). Not surprisingly, genes ascribed to fatty acid metabolism (*TES1*), ergosterol biosynthesis (*ERG5, ERG7, ERG26, HMG1, ERG28*) and cell membrane-associated proteins (*HES1, PUN1*) are upregulated during growth in phenolic compounds (Endo et al., 2009; Sardi et al., 2016; Thompson et al., 2016).

Taken together, in dealing with phenolic compound toxicity, it is evident that *S. cerevisiae* upregulates its oxido-reductase machinery to oxidize and/or reduce phenolic compounds into less toxic forms as well as deal with the oxidative stress associated with this conversion. In addition, the mitochondrial function is upregulated to ensure ATP production required for the activity of transporters which potentially extrudes the phenolic compounds and/or the detoxified forms of it.

METABOLIC ENGINEERING CONSIDERATIONS AND SYNTHETIC BIOLOGY TOOLS TO IMPROVE *S. cerevisiae* TOLERANCE TO PHENOLIC FERMENTATION INHIBITORS DURING BIOMANUFACTURING

ROS scavenging, regulation of ergosterol biosynthesis and compound efflux serve as general phenolic tolerance pathways that can be engineered in a yeast production strain. Beyond engineering a general tolerance pathway, more distinct genetic modifications can be incorporated to result in tolerance to particular phenolics. For instance, deleting *YRR1*, *MRS4*, and *BNA7* to specifically increase tolerance to vanillin (Wang X. et al., 2017), coniferyl aldehyde (Hacısalihoğlu et al., 2019), and ferulic acid (Fletcher et al., 2019) respectively. To rapidly facilitate these metabolic engineering strategies in building a phenolic tolerance pathway, genome editing tools, such as the CRISPR/Cas technology will make this possible (Li et al., 2020).

With the CRISPR/Cas technology, stable genetic modifications including introduction of specific mutations can be inserted into both promoter regions and coding regions of genes. This will facilitate modulating the transcription of genes required for phenolic tolerance (Giersch and Finnigan, 2017).

Alternatively, expression of target genes can be regulated using a repurposed CRISPR system referred to as CRISPR interference (CRISPRi) (Qi et al., 2013). Similarly, introduction of mutations into the coding region can target protein modification sites (Giersch and Finnigan, 2017), modify activity (gain or loss of function), or changes in localization. Other modifications possible with the CRISPR/Cas method are gene insertions and deletions (Akhmetov et al., 2018). Insertion of extra copies of genes is done to stably overexpress target genes whereas multiple gene deletions can be used to knock out specific metabolic pathways (Giersch and Finnigan, 2017). Presently, CRISPR/Cas technology has not yet been applied to improving phenolic tolerance in yeast although the technology has been used to improve tolerance to other fermentation-related stresses in yeast (Giersch and Finnigan, 2017).

While growth (biomass yield) and tolerance are closely connected, making certain genetic manipulations associated with phenolic tolerance may lead to unwanted trade-offs in production hosts which can negatively impact product yield and titers. Therefore, to ensure the success of engineering yeast phenolic tolerance, metabolic flux analysis should be performed to assess the effect of the genetic modifications on the general physiology of the cell as well as carbon flux toward product formation. Such metabolic flux analyses should quantify and guide efficient resource allocation to ensure that cellular resources, such as NADPH and ATP are not diverted to phenolic tolerance at the expense of biosynthesis of bioproducts in the production strain.

Lastly, in the context of producing phenolic compounds, such as vanillin by fermentation, metabolic regulation of pathways that result in the catabolism of the phenolic as a detoxification strategy should be eliminated to improve yields.

FUTURE OUTLOOK AND CONCLUDING REMARKS

Overall, different systems biology approaches have been used to track global phenolic stress responses in *S. cerevisiae*. While common themes or mechanisms coincide among multiple studies, the different approaches provide alternative pathways and biological processes that can be exploited for strain improvement. Moving forward, since most of the long list of genetic hits reported in the various studies has not been validated, significant effort is required to confirm their actual role in tolerance or sensitivity to phenolics. This is particularly crucial for the transcriptomics data because the fact that a gene is upregulated or enriched during stress does not necessarily mean an overexpression of that gene will result in tolerance (Evans, 2015). Gene enrichment could merely be a stress response and not a tolerance mechanism. If possible, the role of enriched genes associated with phenolic tolerance should be confirmed by deleting and/or overexpressing target genes in cells grown in the presence phenolic inhibitors. Such confirmed genes should be cataloged in a "phenolic stressome" database similar to the yStreX (Wanichthanarak et al., 2014) as a repository where synthetic biologists can search for genetic targets to engineer tolerance to different phenolic compounds. By applying synthetic biology tools, such as the CRISPR/Cas technology, the expression of single or multiple genes identified in the "phenolic stressome" can be regulated in order to improve tolerance to phenolic compounds.

Finally, establishing the metabolomic profile of *S. cerevisiae* that are tolerant to a wide spectrum of individual phenolics may guide the development of biosensors to detect "signature metabolites" characteristic of tolerant and high-performing strains. Again, using synthetic biology, biosensors can be constructed with promoters (that are responsive to metabolites characteristic to tolerant and high-performing strains) and a reporting system (e.g., GFP), and inserted into a library of yeast mutants. Next, by applying microfluidics, a pool of heterogenous yeast mutants can be sorted to isolate phenolic tolerant strains that can be used in fermentation-based biomanufacturing to increase product yield and titers.

AUTHOR CONTRIBUTIONS

EF and KB conceived and designed the study. EF drafted the initial manuscript. KB edited and corrected the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We acknowledge funding from the Natural Science and Engineering Research Council (NSERC) of Canada and the Ontario Research Fund Research Excellence funded program titled Biochemicals from Cellulosic Biomass (BioCeB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.539902/full#supplementary-material

REFERENCES

Abe, F., and Hiraki, T. (2009). Mechanistic role of ergosterol in membrane rigidity and cycloheximide resistance in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta Biomembr.* 1788, 743–752. doi: 10.1016/j.bbamem.2008.12.002

Abo Bodjui, O., Gao, M., Wang, Y., Wu, C., Ma, H., and Wang, Q. (2019). Lignocellulosic biomass for bioethanol: an overview on pretreatment, hydrolysis and fermentation processes. *Rev. Environ. Health* 34, 57–68. doi: 10.1515/reveh-2018-0054

Adeboye, P. T., Bettiga, M., Aldaeus, F., Larsson, P. T., and Olsson, L. (2015). Catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid by Saccharomyces cerevisiae yields less toxic products. *Microb. Cell Fact.* 14, 149–149. doi: 10.1186/s12934-015-0338-x

Adeboye, P. T., Bettiga, M., and Olsson, L. (2014). The chemical nature of phenolic compounds determines their toxicity and induces distinct physiological

responses in Saccharomyces cerevisiae in lignocellulose hydrolysates. *AMB Express* 4, 46–46. doi: 10.1186/s13568-014-0046-7

Adeboye, P. T., Bettiga, M., and Olsson, L. (2017). ALD5, PAD1, ATF1 and ATF2 facilitate the catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid in *Saccharomyces cerevisiae*. *Sci. Rep.* 7:42635. doi: 10.1038/srep42635

Akhmetov, A., Laurent, J. M., Gollihar, J., Gardner, E. C., Garge, R. K., Ellington, A. D., et al. (2018). Single-step precision genome editing in yeast using CRISPR-Cas9. *Bio Protoc.* 8:e2765. doi: 10.21769/BioProtoc.2765

Ando, S., Arai, I., Kiyoto, K., and Hanai, S. (1986). Identification of aromatic monomers in steam-exploded poplar and their influences on ethanol fermentation by *Saccharomyces cerevisiae*. *J. Ferment. Technol.* 64, 567–570. doi: 10.1016/0385-6380(86)90084-1

Bastos, A. E. P., Marinho, H. S., Cordeiro, A. M., de Soure, A. M., and de Almeida, R. F. M. (2012). Biophysical properties of ergosterol-enriched lipid rafts in yeast and tools for their study: characterization of ergosterol/phosphatidylcholine membranes with three fluorescent membrane probes. *Chem. Phys. Lipids* 165, 577–588. doi: 10.1016/j.chemphyslip.2012.06.002

Becker, J., and Wittmann, C. (2019). A field of dreams: lignin valorization into chemicals, materials, fuels, and health-care products. *Biotechnol. Adv.* 37:107360. doi: 10.1016/j.biotechadv.2019.02.016

Biot-Pelletier, D., Pinel, D., Larue, K., and Martin, V. J. J. (2018). Determinants of selection in yeast evolved by genome shuffling. *Biotechnol. Biofuels* 11:282. doi: 10.1186/s13068-018-1283-9

Brochado, A. R., Matos, C., Møller, B. L., Hansen, J., Mortensen, U. H., and Patil, K. R. (2010). Improved vanillin production in baker's yeast through in silico design. *Microb. Cell Fact.* 9:84. doi: 10.1186/1475-2859-9-84

Carter, B., Squillace, P., Gilcrease, P. C., and Menkhaus, T. J. (2011). Detoxification of a lignocellulosic biomass slurry by soluble polyelectrolyte adsorption for improved fermentation efficiency. *Biotechnol. Bioeng.* 108, 2053–2060. doi: 10.1002/bit.23152

Chen, L., Zhang, Z., Hoshino, A., Zheng, H. D., Morley, M., Arany, Z., et al. (2019). NADPH production by the oxidative pentose-phosphate pathway supports folate metabolism. *Nat. Metab.* 1, 404–415. doi: 10.1038/s42255-019-0043-x

de Witt, R. N., Kroukamp, H., Van Zyl, W. H., Paulsen, I. T., and Volschenk, H. (2019). QTL analysis of natural *Saccharomyces cerevisiae* isolates reveals unique alleles involved in lignocellulosic inhibitor tolerance. *FEMS Yeast Res.* 19:foz047. doi: 10.1093/femsyr/foz047

de Witt, R. N., Kroukamp, H., and Volschenk, H. (2018). Proteome response of two natural strains of Saccharomyces cerevisiae with divergent lignocellulosic inhibitor stress tolerance. *FEMS Yeast Res.* 19:foy116. doi: 10.1093/femsyr/foy116

Di Gioia, D., Sciubba, L., Ruzzi, M., Setti, L., and Fava, F. (2009). Production of vanillin from wheat bran hydrolyzates via microbial bioconversion. *J. Chem. Technol. Biotechnol.* 84, 1441–1448. doi: 10.1002/jctb.2196

Ding, M.-Z., Wang, X., Liu, W., Cheng, J.-S., Yang, Y., and Yuan, Y.-J. (2012a). Proteomic research reveals the stress response and detoxification of yeast to combined inhibitors. *PLoS One* 7:e43474. doi: 10.1371/journal.pone.0043474

Ding, M.-Z., Wang, X., Yang, Y., and Yuan, Y.-J. (2012b). Comparative metabolic profiling of parental and inhibitors-tolerant yeasts during lignocellulosic ethanol fermentation. *Metabolomics* 8, 232–243. doi: 10.1007/s11306-011-0303-6

Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.* 12:64. doi: 10.1186/1475-2859-12-64

Eisenkolb, M., Zenzmaier, C., Leitner, E., and Schneiter, R. (2002). A specific structural requirement for ergosterol in long-chain fatty acid synthesis mutants important for maintaining raft domains in yeast. *Mol. Biol. Cell* 13, 4414–4428. doi: 10.1091/mbc.e02-02-0116

Endo, A., Nakamura, T., Ando, A., Tokuyasu, K., and Shima, J. (2008). Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 1:3. doi: 10.1186/1754-6834-1-3

Endo, A., Nakamura, T., and Shima, J. (2009). Involvement of ergosterol in tolerance to vanillin, a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.* 299, 95–99. doi: 10.1111/j.1574-6968.2009.01733.x

Evans, T. G. (2015). Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *J. Exp. Biol.* 218:1925. doi: 10.1242/jeb.114306

Fletcher, E., Feizi, A., Bisschops, M. M. M., Hallström, B. M., Khoomrung, S., Siewers, V., et al. (2017). Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. *Metab. Eng.* 39, 19–28. doi: 10.1016/j.ymben.2016.10.010

Fletcher, E., Gao, K., Mercurio, K., Ali, M., and Baetz, K. (2019). Yeast chemogenomic screen identifies distinct metabolic pathways required to tolerate exposure to phenolic fermentation inhibitors ferulic acid, 4-hydroxybenzoic acid and coniferyl aldehyde. *Metab. Eng.* 52, 98–109. doi: 10.1016/j.ymben.2018.11.010

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., et al. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741. doi: 10.1038/nature02046

Giaever, G., and Nislow, C. (2014). The yeast deletion collection: a decade of functional genomics. *Genetics* 197, 451–465. doi: 10.1534/genetics.114.161620

Giersch, R. M., and Finnigan, G. C. (2017). Yeast still a beast: diverse applications of CRISPR/Cas editing technology in *S. cerevisiae*. *Yale J. Biol. Med.* 90, 643–651.

Gómez, S., Navas-Yuste, S., Payne, A. M., Rivera, W., López-Estepa, M., Brangbour, C., et al. (2019). Peroxisomal catalases from the yeasts *Pichia pastoris* and *Kluyveromyces lactis* as models for oxidative damage in higher eukaryotes. *Free Radic. Biol. Med.* 141, 279–290. doi: 10.1016/j.freeradbiomed.2019.06.025

Hacısalihoğlu, B., Holyavkin, C., Topaloğlu, A., Kısakesen, H. I., and Çakar, Z. P. (2019). Genomic and transcriptomic analysis of a coniferyl aldehyde-resistant Saccharomyces cerevisiae strain obtained by evolutionary engineering. *FEMS Yeast Res.* 19:foz021. doi: 10.1093/femsyr/foz021

Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinform Biol. Insights* 9, 29–46. doi: 10.4137/BBI.S28991

Higgins, D. A., Young, M. K. M., Tremaine, M., Sardi, M., Fletcher, J. M., Agnew, M., et al. (2018). Natural variation in the multidrug efflux pump SGE1 underlies ionic liquid tolerance in yeast. *Genetics* 210, 219–234. doi: 10.1534/genetics.118.301161

Hou, L. (2009). Novel methods of genome shuffling in Saccharomyces cerevisiae. *Biotechnol. Lett.* 31, 671–677. doi: 10.1007/s10529-009-9916-5

Iwaki, A., Ohnuki, S., Suga, Y., Izawa, S., and Ohya, Y. (2013). Vanillin inhibits translation and induces messenger ribonucleoprotein (mRNP) granule formation in *Saccharomyces cerevisiae*: application and validation of high-content, image-based profiling. *PLoS One* 8:e61748. doi: 10.1371/journal.pone.0061748

Katzmann, D. J., Hallstrom, T. C., Voet, M., Wysock, W., Golin, J., Volckaert, G., et al. (1995). Expression of an ATP-binding cassette transporter-encoding gene (YOR1) is required for oligomycin resistance in Saccharomyces cerevisiae. *Mol. Cell. Biol.* 15, 6875–6883. doi: 10.1128/mcb.15.12.6875

Kodedová, M., and Sychrová, H. (2015). Changes in the sterol composition of the plasma membrane affect membrane potential, salt tolerance and the activity of multidrug resistance pumps in *Saccharomyces cerevisiae*. *PLoS One* 10:e0139306. doi: 10.1371/journal.pone.0139306

Koh, J. L. Y., Chong, Y. T., Friesen, H., Moses, A., Boone, C., Andrews, B. J., et al. (2015). CYCLoPs: a comprehensive database constructed from automated analysis of protein abundance and subcellular localization patterns in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 5, 1223–1232. doi: 10.1534/g3.115.017830

Larsson, O., and Sandberg, R. (2006). Lack of correct data format and comparability limits future integrative microarray research. *Nat. Biotechnol.* 24, 1322–1323. doi: 10.1038/nbt1106-1322

Li, C., Chen, C., Wu, X., Tsang, C.-W., Mou, J., Yan, J., et al. (2019). Recent advancement in lignin biorefinery: with special focus on enzymatic degradation and valorization. *Bioresour. Technol.* 291:121898. doi: 10.1016/j.biortech.2019.121898

Li, Z.-H., Meng, H., Ma, B., Tao, X., Liu, M., Wang, F.-Q., et al. (2020). Immediate, multiplexed and sequential genome engineering facilitated by CRISPR/Cas9 in *Saccharomyces cerevisiae*. *J. Indust. Microbiol. Biotechnol.* 47, 83–96. doi: 10.1007/s10295-019-02251-w

Luziatelli, F., Brunetti, L., Ficca, A. G., and Ruzzi, M. (2019). Maximizing the efficiency of vanillin production by biocatalyst enhancement and process optimization. *Front. Bioeng. Biotechnol.* 7:279. doi: 10.3389/fbioe.2019.00279

Lv, Y.-J., Wang, X., Ma, Q., Bai, X., Li, B.-Z., Zhang, W., et al. (2014). Proteomic analysis reveals complex metabolic regulation in *Saccharomyces cerevisiae* cells against multiple inhibitors stress. *Appl. Microbiol. Biotechnol.* 98, 2207–2221. doi: 10.1007/s00253-014-5519-8

Maeta, K., Izawa, S., Okazaki, S., Kuge, S., and Inoue, Y. (2004). Activity of the Yap1 transcription factor in &It;em> *Saccharomyces cerevisiae*&It;/em> is modulated by methylglyoxal, a metabolite derived from glycolysis. *Mol. Cell. Biol.* 24:8753. doi: 10.1128/MCB.24.19.8753-8764.2004

Magocha, T. A., Zabed, H., Yang, M., Yun, J., Zhang, H., and Qi, X. (2018). Improvement of industrially important microbial strains by genome shuffling: current status and future prospects. *Bioresour. Technol.* 257, 281–289. doi: 10.1016/j.biortech.2018.02.118

Mamnun, Y. M., Schüller, C., and Kuchler, K. (2004). Expression regulation of the yeast PDR5 ATP-binding cassette (ABC) transporter suggests a role in cellular detoxification during the exponential growth phase. *FEBS Lett.* 559, 111–117. doi: 10.1016/s0014-5793(04)00046-8

Marshall, D. D., and Powers, R. (2017). Beyond the paradigm: combining mass spectrometry and nuclear magnetic resonance for metabolomics. *Prog. Nucl. Magn. Reson. Spectrosc.* 100, 1–16. doi: 10.1016/j.pnmrs.2017.01.001

Nguyen, T. T. M., Iwaki, A., and Izawa, S. (2015). The ADH7 promoter of *Saccharomyces cerevisiae* is vanillin-inducible and enables mRNA translation under severe vanillin stress. *Front. Microbiol.* 6:1390. doi: 10.3389/fmicb.2015.01390

Nguyen, T. T. M., Iwaki, A., Ohya, Y., and Izawa, S. (2014). Vanillin causes the activation of Yap1 and mitochondrial fragmentation in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 117, 33–38. doi: 10.1016/j.jbiosc.2013.06.008

Nugroho, R. H., Yoshikawa, K., and Shimizu, H. (2015). Metabolomic analysis of acid stress response in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 120, 396–404. doi: 10.1016/j.jbiosc.2015.02.011

Ohnuki, S., Oka, S., Nogami, S., and Ohya, Y. (2010). High-content, image-based screening for drug targets in yeast. *PLoS One* 5:e10177. doi: 10.1371/journal.pone.0010177

Ong, S.-E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* 1, 119–126. doi: 10.1038/nmeth715

Overhage, J., Steinbüchel, A., and Priefert, H. (2002). Biotransformation of eugenol to ferulic acid by a recombinant strain of &It;em> Ralstonia eutropha&It;/em> H16. *Appl. Environ. Microbiol.* 68:4315. doi: 10.1128/aem.68.9.4315-4321.2002

Palmqvist, E., and Hahn-Hägerdal, B. (2000). Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresour. Technol.* 74, 25–33. doi: 10.1016/s0960-8524(99)00161-3

Park, E.-H., and Kim, M.-D. (2014). Genome-wide screening of *Saccharomyces cerevisiae* genes regulated by vanillin. *J. Microbiol. Biotechnol.* 25, 50–56. doi: 10.4014/jmb.1409.09064

Pereira, F. B., Teixeira, M. C., Mira, N. P., Sá-Correia, I., and Domingues, L. (2014). Genome-wide screening of *Saccharomyces cerevisiae* genes required to foster tolerance towards industrial wheat straw hydrolysates. *J. Indust. Microbiol. Biotechnol.* 41, 1753–1761. doi: 10.1007/s10295-014-1519-z

Pinel, D., Aoust, F., del Cardayre, S. B., Bajwa, P. K., Lee, H., and Martin, V. J. J. (2011). *Saccharomyces cerevisiae*&It;/span> genome shuffling through recursive population mating leads to improved tolerance to spent sulfite liquor. *Appl. Environ. Microbiol.* 77:4736. doi: 10.1128/AEM.02769-10

Piotrowski, J. S., Okada, H., Lu, F., Li, S. C., Hinchman, L., Ranjan, A., et al. (2015). Plant-derived antifungal agent poacic acid targets β-1,3-glucan. *Proc. Natl. Acad. Sci. U.S.A.* 112:E1490. doi: 10.1073/pnas.1410400112

Ponnusamy, V. K., Nguyen, D. D., Dharmaraja, J., Shobana, S., Banu, J. R., Saratale, R. G., et al. (2019). A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential. *Bioresour. Technol.* 271, 462–472. doi: 10.1016/j.biortech.2018.09.070

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., et al. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Qin, L., Li, W.-C., Liu, L., Zhu, J.-Q., Li, X., Li, B.-Z., et al. (2016). Inhibition of lignin-derived phenolic compounds to cellulase. *Biotechnol. Biofuels* 9:70. doi: 10.1186/s13068-016-0485-2

Radek, A., Tenhaef, N., Müller, M. F., Brüsseler, C., Wiechert, W., Marienhagen, J., et al. (2017). Miniaturized and automated adaptive laboratory evolution: evolving Corynebacterium glutamicum towards an improved d-xylose utilization. *Bioresour. Technol.* 245, 1377–1385. doi: 10.1016/j.biortech.2017.05.055

Raghavachari, N. (2013). "Microarray technology: basic methodology and application in clinical research for biomarker discovery in vascular diseases," in *Lipoproteins and Cardiovascular Disease: Methods and Protocols*, ed. L. A. Freeman (Totowa, NJ: Humana Press), 47–84. doi: 10.1007/978-1-60327-369-5_3

Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genomics Proteom. Bioinform.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Robinson, M. D., Grigull, J., Mohammad, N., and Hughes, T. R. (2002). FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinform.* 3:35. doi: 10.1186/1471-2105-3-35

Rodrigues-Pousada, C., Menezes, R. A., and Pimentel, C. (2010). The Yap family and its role in stress response. *Yeast* 27, 245–258. doi: 10.1002/yea.1752

Sandmann, S., de Graaf, A. O., Karimi, M., van der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H., et al. (2017). Evaluating variant calling tools for non-matched next-generation sequencing data. *Sci. Rep.* 7:43169. doi: 10.1038/srep43169

Sardi, M., Paithane, V., Place, M., Robinson, D. E., Hose, J., Wohlbach, D. J., et al. (2018). Genome-wide association across *Saccharomyces cerevisiae* strains reveals substantial variation in underlying gene requirements for toxin tolerance. *PLoS Genet.* 14:e1007217. doi: 10.1371/journal.pgen.1007217

Sardi, M., Rovinskiy, N., Zhang, Y., and Gasch, A. P. (2016). Leveraging genetic-background effects in &It;span class="named-content genus-species" id="named-content-1">Saccharomyces cerevisiae&It/span&gt to improve lignocellulosic hydrolysate tolerance. *Appl. Environ. Microbiol.* 82:5838. doi: 10.1128/aem.01603-16

Shen, Y., Li, H., Wang, X., Zhang, X., Hou, J., Wang, L., et al. (2014). High vanillin tolerance of an evolved Saccharomyces cerevisiae strain owing to its enhanced vanillin reduction and antioxidative capacity. *J. Indust. Microbiol. Biotechnol.* 41, 1637–1645. doi: 10.1007/s10295-014-1515-3

Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M., and Davis, R. W. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar–coding strategy. *Nat. Genet.* 14, 450–456. doi: 10.1038/ng1296-450

Skerker, J. M., Leon, D., Price, M. N., Mar, J. S., Tarjan, D. R., Wetmore, K. M., et al. (2013). Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. *Mol. Syst. Biol.* 9, 674–674. doi: 10.1038/msb.2013.30

Smith, A. M., Heisler, L. E., Mellor, J., Kaper, F., Thompson, M. J., Chee, M., et al. (2009). Quantitative phenotyping via deep barcode sequencing. *Genome Res.* 19, 1836–1842. doi: 10.1101/gr.093955.109

Sundström, L., Larsson, S., and Jönsson, L. J. (2010). Identification of *Saccharomyces cerevisiae* genes involved in the resistance to phenolic fermentation inhibitors. *Appl. Biochem. Biotechnol.* 161, 106–115. doi: 10.1007/s12010-009-8811-9

Thompson, O. A., Hawkins, G. M., Gorsich, S. W., and Doran-Peterson, J. (2016). Phenotypic characterization and comparative transcriptomics of evolved Saccharomyces cerevisiae strains with improved tolerance to lignocellulosic derived inhibitors. *Biotechnol. Biofuels* 9:200. doi: 10.1186/s13068-016-0614-y

Toledano, M. B., Delaunay-Moisan, A., Outten, C. E., and Igbaria, A. (2013). Functions and cellular compartmentation of the thioredoxin and glutathione pathways in yeast. *Antioxid Redox Signal.* 18, 1699–1711. doi: 10.1089/ars.2012.5033

Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., et al. (2018). Evaluation of oxford nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci. Rep.* 8:10931. doi: 10.1038/s41598-018-29334-5

Wang, H.-Y., Xiao, D.-F., Zhou, C., Wang, L.-L., Wu, L., Lu, Y.-T., et al. (2017). YLL056C from *Saccharomyces cerevisiae* encodes a novel protein with aldehyde reductase activity. *Appl. Microbiol. Biotechnol.* 101, 4507–4520. doi: 10.1007/s00253-017-8209-5

Wang, X., Liang, Z., Hou, J., Shen, Y., and Bao, X. (2017). The absence of the transcription factor Yrr1p, identified from comparative genome profiling, increased vanillin tolerance due to enhancements of ABC transporters expressing, rRNA processing and ribosome biogenesis in *Saccharomyces cerevisiae*. *Front. Microbiol.* 8:367. doi: 10.3389/fmicb.2017.00367

Wanichthanarak, K., Nookaew, I., and Petranovic, D. (2014). yStreX: yeast stress expression database. *Database (Oxford)* 2014:bau068. doi: 10.1093/database/bau068

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al. (1999). Functional characterization of the &It;em>S. cerevisiae&It;/em> genome by gene deletion and parallel analysis. *Science* 285:901. doi: 10.1126/science.285.5429.901

Wu, G., Xu, Z., and Jönsson, L. J. (2017). Profiling of Saccharomyces cerevisiae transcription factors for engineering the resistance of yeast to lignocellulose-derived inhibitors in biomass conversion. *Microb. Cell Fact.* 16:199. doi: 10.1186/s12934-017-0811-9

Xue, S., Jones, A. D., Sousa, L., Piotrowski, J., Jin, M., Sarks, C., et al. (2018). Water-soluble phenolic compounds produced from extractive ammonia pretreatment exerted binary inhibitory effects on yeast fermentation using synthetic hydrolysate. *PLoS One* 13:e0194012. doi: 10.1371/journal.pone.0194012

Zampieri, M., Sekar, K., Zamboni, N., and Sauer, U. (2017). Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* 36, 15–23. doi: 10.1016/j.cbpa.2016.12.006

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. III (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* 113, 2343–2394. doi: 10.1021/cr3003533

Zheng, D.-Q., Jin, X.-N., Zhang, K., Fang, Y.-H., and Wu, X.-C. (2017). Novel strategy to improve vanillin tolerance and ethanol fermentation performances of Saccharomycere cerevisiae strains. *Bioresour. Technol.* 231, 53–58. doi: 10.1016/j.biortech.2017.01.040

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fbioe.2020.595552

Multi-Omics Analysis of the Effect of cAMP on Actinorhodin Production in *Streptomyces coelicolor*

Katsuaki Nitta[1], Francesco Del Carratore[2], Rainer Breitling[2], Eriko Takano[2], Sastia P. Putri[1]* and Eiichiro Fukusaki[1]*

[1] Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan, [2] Department of Chemistry, Manchester Synthetic Biology Research Centre SYNBIOCHEM, Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom

Cyclic adenosine monophosphate (cAMP) has been known to play an important role in regulating morphological development and antibiotic production in *Streptomyces coelicolor*. However, the functional connection between cAMP levels and antibiotic production and the mechanism by which cAMP regulates antibiotic production remain unclear. In this study, metabolomics- and transcriptomics-based multi-omics analysis was applied to *S. coelicolor* strains that either produce the secondary metabolite actinorhodin (Act) or lack most secondary metabolite biosynthesis pathways including Act. Comparative multi-omics analysis of the two strains revealed that intracellular and extracellular cAMP abundance was strongly correlated with actinorhodin production. Notably, supplementation of cAMP improved cell growth and antibiotic production. Further multi-omics analysis of cAMP-supplemented *S. coelicolor* cultures showed an increase of guanine and the expression level of purine metabolism genes. Based on this phenomenon, supplementation with 7-methylguanine, a competitive inhibitor of reactions utilizing guanine, with or without additional cAMP supplementation, was performed. This experiment revealed that the reactions inhibited by 7-methylguanine are mediating the positive effect on growth and antibiotic production, which may occur downstream of cAMP supplementation.

Keywords: *Streptomyces coelicolor*, metabolomics, transcriptomics, cAMP, secondary metabolites, actinorhodin, 7-methylguanine

OPEN ACCESS

Edited by:
Young-Mo Kim,
Pacific Northwest National Laboratory (DOE), United States

Reviewed by:
Govind Chandra,
John Innes Centre, United Kingdom
Yinhua Lu,
Shanghai Normal University, China

****Correspondence:***
Sastia P. Putri
sastia_putri@bio.eng.osaka-u.ac.jp
Eriko Takano
eriko.takano@manchester.ac.uk

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *17 August 2020*
Accepted: *08 October 2020*
Published: *05 November 2020*

Citation:
Nitta K, Carratore FD, Breitling R, Takano E, Putri SP and Fukusaki E (2020) Multi-Omics Analysis of the Effect of cAMP on Actinorhodin Production in Streptomyces coelicolor. Front. Bioeng. Biotechnol. 8:595552. doi: 10.3389/fbioe.2020.595552

INTRODUCTION

Streptomyces species are soil-dwelling Gram-positive Actinobacteria and well-known important sources of bioactive secondary metabolites such as antibiotics (e.g., streptomycin; Davies et al., 1964), immunosuppressants (e.g., FK-506; Kino et al., 1987a,b), and anthelmintics (e.g., avermectin; Ikeda and Omura, 1997). Secondary metabolism is not necessary for bacterial growth and often occurs under nutrient limitation (Strauch et al., 1991), in specific medium condition (Lim et al., 2018), or in co-culture with other microbes (Onaka et al., 2011). The regulation of secondary metabolite production has therefore been a focus of research in natural product discovery and production (Liu et al., 2018). One way of improving secondary metabolite production is the addition of exogenous compounds such as cyclic adenosine monophosphate (cAMP)

supplementation in *Streptomyces coelicolor* (Susstrunk et al., 1998), rare earth elements (e.g., scandium) in *S. coelicolor* (Tanaka et al., 2010), S-adenosyl-L-methionine (SAM) in *Streptomyces lividans* (Kim et al., 2003), and hormones with a butanolide skeleton in *Streptomyces virginiae* (Kim et al., 1989). However, in most cases, it is not fully understood by which mechanism such approaches led to the increased production of secondary metabolites.

With the recent rapid progress of analytical equipment and methods that contribute to the accumulation of more analytical data, new opportunities arise for studying the regulation of secondary metabolism. Transcriptome (Hwang et al., 2019), proteome (Millan-Oropeza et al., 2017), metabolome (Senges et al., 2018), and lipidome analyses (Zhang et al., 2020) or various combinations of these (Wang et al., 2017; Gosse et al., 2019) in *Streptomyces* have been applied to studies of secondary metabolism regulation.

In this study, we exploit these new technological opportunities to develop a more detailed picture of the regulatory circuitry surrounding the production of antibiotics in the model *Streptomyces* species *S. coelicolor*. Liquid chromatography tandem mass spectrometry (LC-MS/MS)–based metabolome analysis with wide metabolite coverage and RNAseq-based transcriptome analysis were employed, for comparative analyses of strains with differing antibiotic production phenotypes to identify the possible factors affecting secondary metabolite production and to provide a more detailed understanding of the regulation of secondary metabolites. This multi-omics analysis revealed that antibiotic production is correlated with cAMP levels, and cAMP supplementation to *S. coelicolor* culture led to improvements of cell growth and antibiotic production. Furthermore, multi-omics analysis of cAMP-supplemented cultures was performed to explore how the effects of cAMP are mediated through changes in the *S. coelicolor* metabolome and transcriptome.

These results provide valuable mechanistic insights into the effects of cAMP and their regulation, which in the future might be exploited for manipulating the cell growth and antibiotic production of *S. coelicolor*.

RESULTS AND DISCUSSION

Phenotypes of *S. coelicolor* Strains for Omics-Based Comparative Analyses

In a previous study, the M1146 strain was generated from strain M145 thorough step-by-step deletion of the four major native antibiotics biosynthetic gene clusters in *S. coelicolor*: *act* (actinorhodin, Act), *red* (undecylprodigiosin), *cpk* (coelimycin P1), and *cda* (calcium-dependent antibiotic). The original intention was to achieve higher antibiotic production ability by deleting these carbon-consuming pathways and increasing available carbon source for other (heterologously expressed) antibiotics producing pathways and to simplify the detection of the secondary produced that will be produced (Gomez-Escribano and Bibb, 2011). Here, we make use of this "clean" M1146 strain as a reference for an omics-based comparison against M1146 expressing a heterologous Act biosynthetic gene cluster by an integrated cosmid (M1146 + ACT) to identify the molecular consequences of the difference in antibiotic production phenotype. Act was chosen as the target compound for this study, as its intense blue coloration allows easy detection and quantitation of metabolite production. Introduction of the Act biosynthetic genes into M1146 did not affect growth (**Figure 1A**), but it caused Act production as expected (**Figure 1B**). Glucose and phosphate consumption did not show any statistically significant difference, but the M1146 + ACT strain showed slightly higher glutamate consumption, presumably because glutamate was used as a nitrogen source for Act production in this nutrient condition (**Supplementary Figure 1A**). We also analyzed the effect of amplifying Act production in the M145 strain (parent of M1146). An additional Act biosynthetic genes cluster was introduced into M145, as was done in M1146. M145 showed reproducible antibiotic production as in a previous study (Nieselt et al., 2010), and while growth was not affected (**Figure 1C**), M145 with the additional Act biosynthesis gene cluster (M145 + ACT) showed earlier Act production and a 2.41-fold higher Act production than M145 (**Figure 1D**). Contrary to the analogous case of M1146/M1146 + ACT, there was no significant difference in glutamate consumption between M145 and M145 + ACT (**Supplementary Figure 1B**), probably because in this case both strains produce Act and utilize glutamate as a nitrogen source for this purpose. However, phosphate was consumed more slowly in M145 + ACT compared to M145. Interestingly, the production of undecylprodigiosin (RED) also increased in M145 + ACT, whereas coelimycin production was not observed (**Supplementary Figure 1C**). All the strains used in this study are summarized in **Supplementary Table 1**.

Metabolome Analysis of *S. coelicolor* Strains: Correlation of cAMP Levels With Actinorhodin Production

To understand the difference in metabolism with and without Act production, extracellular, and intracellular metabolome profiles of M1146 and M1146 + ACT were determined (**Figure 2**). Among the 99 metabolites analyzed (**Supplementary Table 2**), the major difference in both intracellular and extracellular metabolites between the two strains was that in cAMP levels (**Figure 2**). These were substantially higher in the M1146 + ACT compared to M1146 (**Figures 3A,B**). To support this observation, the intracellular and extracellular cAMP levels were also measured in M145 and M145 + ACT (**Supplementary Figure 2**) and showed higher cAMP levels in M145 + ACT compared to M145, as was observed in M1146 with and without ACT (**Figures 3C,D**). Comparing all four strains, cAMP production was strongly correlated with antibiotic production (e.g., production of Act and coelimycin P1; **Figures 3E,F**; and **Supplementary Figure 1C**).

Transcriptome Analysis of *S. coelicolor* Strains

To further support the results of the targeted metabolome analysis, fine-grained transcriptome analysis was conducted

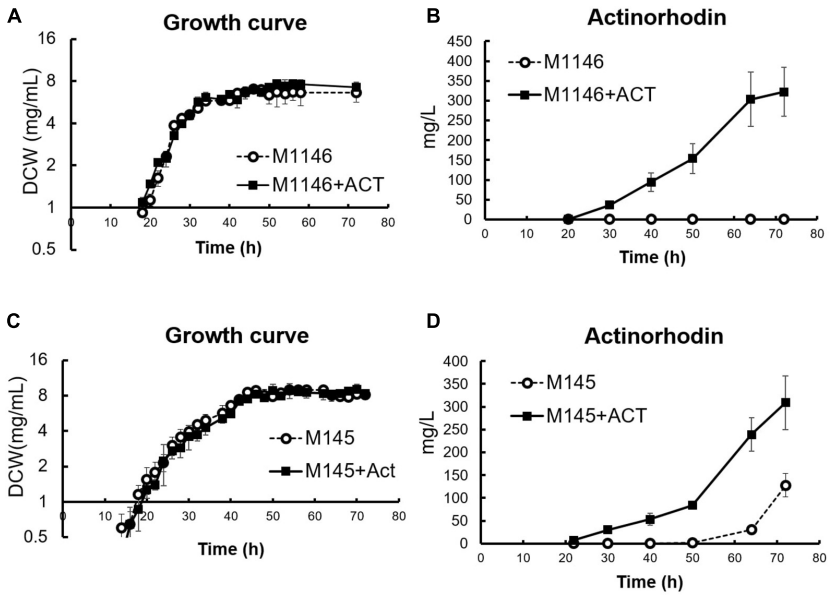


FIGURE 1 | Cell growth and actinorhodin production in four *S. coelicolor* strains. **(A)** Growth curve (DCW: mg/mL) in M1146 (dot line chart) and M1146 + ACT (black line chart). Error bar means standard deviation from three replicates. **(B)** Extracellular actinorhodin production in M1146 (dot line chart) and M1146 + ACT (black line chart). Error bar means standard deviation from three replicates. **(C)** Growth curve (DCW: mg/mL) in M145 (dot line chart) and M145 + ACT (black line chart). Error bar means standard deviation from three replicates. **(D)** Extracellular actinorhodin production in M145 (dot line chart) and M145 + ACT (black line chart). Error bar means standard deviation from three replicates.

on M1146 and M1146 + ACT at 10 time points during growth (**Figure 1A**).

Differentially expressed genes (DEGs) between M1146 and M1146 + ACT were identified in order to further evaluate the effects of Act production on gene expression levels. The false discovery rate (FDR) was controlled at 5%, and only genes with an absolute log2-fold change (FC) greater than 0.5 were included in the analysis. As expected, the genes of the Act biosynthetic gene cluster were only expressed in M1146 + ACT (at very high levels), whereas no expression was detected in M1146, which lacks this gene cluster (**Supplementary Table 3**). As this observation is trivial, these enzyme-coding genes were excluded from the subsequent analysis of differential expression. In total, 100 genes were identified as overexpressed in M1146 + ACT, whereas 15 genes were identified as less abundantly expressed in M1146 + ACT based on the DEG criteria described above (**Supplementary Table 4**). Based on the DEGs, Gene Set Enrichment Analysis (GSEA) was performed, using biological process, pathway, and keywords to define the gene sets (**Figure 4A** and **Supplementary Table 5**). The major overexpressed gene sets in M1146 + ACT were associated with biotin biosynthesis and oxidation/reduction processes (**Figure 4A**).

High expression of biotin biosynthetic genes is perhaps not surprising, as biotin is a necessary cofactor for malonyl-CoA synthesis, which is the essential building block for biosynthesis of actinorhodin (a polyketide; Katz and Donadio, 1993). This observation also matches the high levels of expression of the two acetyl-CoA carboxylase–encoding genes, SCO6271 (*accA1*) and SCO4921 (*accA2*) in M1146 + ACT (**Figure 4B**). When malonyl-CoA levels were analyzed by LC-MS/MS, malonyl-CoA decreased with time, and the malonyl-CoA level was lower in M1146 + ACT at 20 h (**Figure 4B**). Thus, malonyl-CoA was consumed earlier in the M1146 + ACT strain, which is presumably due to the malonyl-CoA being intensely utilized for Act biosynthesis. The difference in expression levels of genes encoding enzymes responsible for cAMP biosynthesis, adenylate cyclase (SCO4928), and cAMP degradation, phosphodiesterase (SCO6075), was not statistically significant (**Supplementary Figure 3**). The expression level of both genes is low and was only slightly increased in M1146 + ACT and showed a very similar trend, where a second peak of increased gene expression was seen at 30 h, which coincides with the cells entering the stationary phase. This in itself cannot explain the observed difference in cAMP production. The discrepancy in the expression of the genes encoding cAMP synthesis and degradation and cAMP levels is intriguing and will need further study to fully understand this phenomenon.

To reveal trends in the gene expression time courses, the transcriptome data were subjected to *k*-means clustering ($k = 20$; **Figure 4C**, **Supplementary Figure 4**, and **Supplementary Excel File 1**). All genes were clustered into 20 classes, and classes J and K genes were more highly expressed in M1146 + ACT at all sampling points. Class J consisted of 22 genes encoding enzymes involved in Act biosynthesis, whereas class K has 10 genes, most of which were reported as related to the *soxR* regulon (**Table 1**; Dela Cruz et al., 2010; Naseer et al., 2014). The *SoxR* regulon was reported to be upregulated by oxidative stress caused by high levels of Act production in previous studies (Shin et al., 2011;

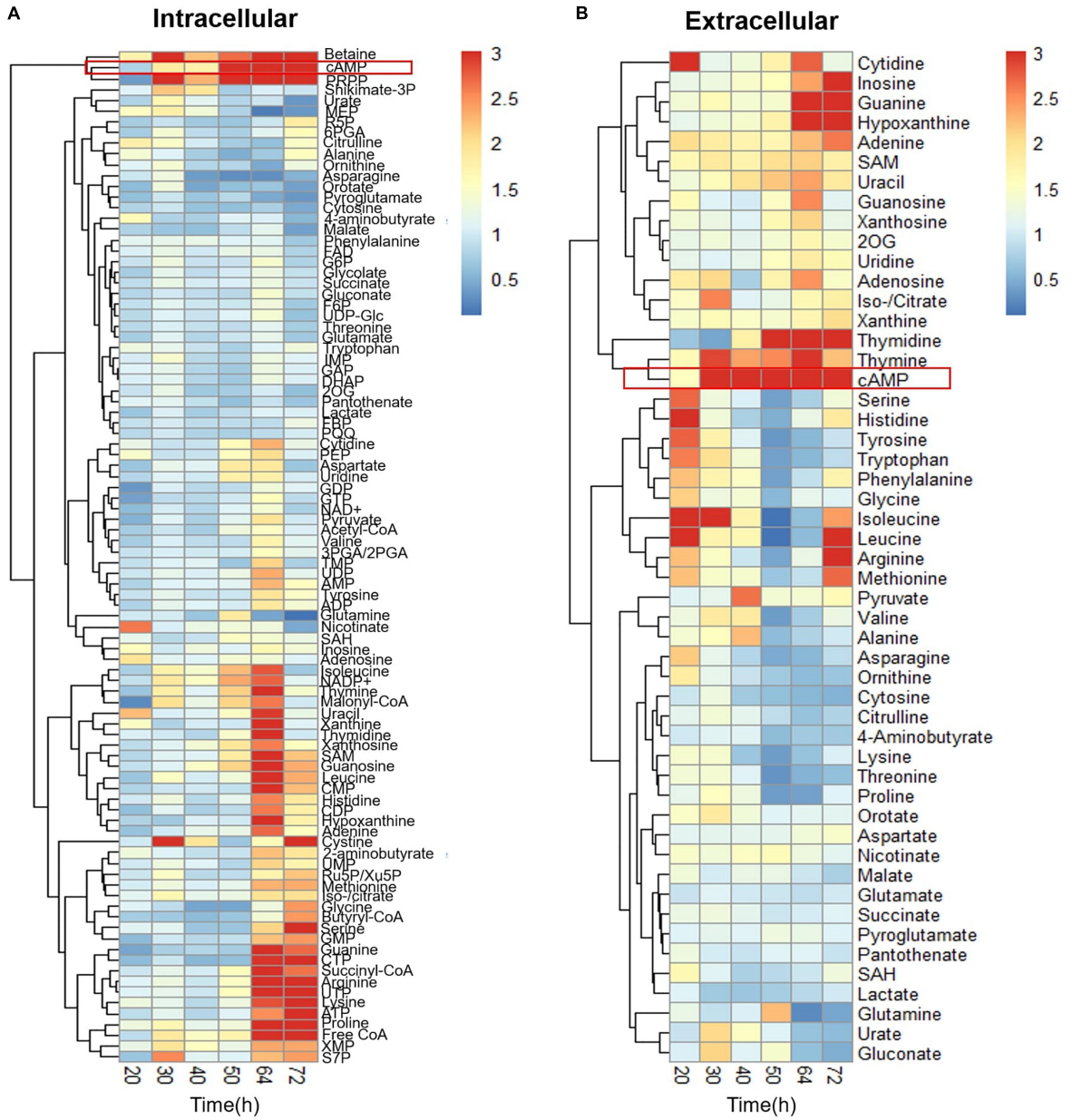


FIGURE 2 | Time-course metabolome analysis of M1146 and M1146 + ACT. **(A)** Intracellular metabolite levels shown by fold change value using heatmap analysis. Comparison between M1146 and M1146 + ACT. Average values from three replicates were used to calculate the fold change value. **(B)** Extracellular metabolite levels shown by fold change value using heatmap analysis. Comparison between M1146 and M1146 + ACT. Average values from three replicates were used to calculate the fold change value. Metabolites circled by red box represent cAMP.

Mak and Nodwell, 2017), which would be in agreement with our results.

cAMP Supplementation to *S. coelicolor* Increased Cell Growth and Secondary Metabolite Production

To explore whether the observed correlation of cAMP levels with antibiotic production was the result of a causal connection between the two phenomena, we performed a cAMP supplementation experiment (**Figure 5**). cAMP has been previously reported as an inducer of Act production (Susstrunk et al., 1998), and the deletion of the *cya* (SCO4928) encoding the adenylate cyclase abolished antibiotic production (Susstrunk et al., 1998; Kang et al., 1999), whereas the cAMP receptor protein CRP encoded by gene SCO3571 was shown to be important for morphological development, and ChIP-chip experiments showed secondary metabolism gene clusters including Act contained Crp-associated sites (Derouaux et al., 2004; Gao et al., 2012). Before conducting the cAMP supplementation,

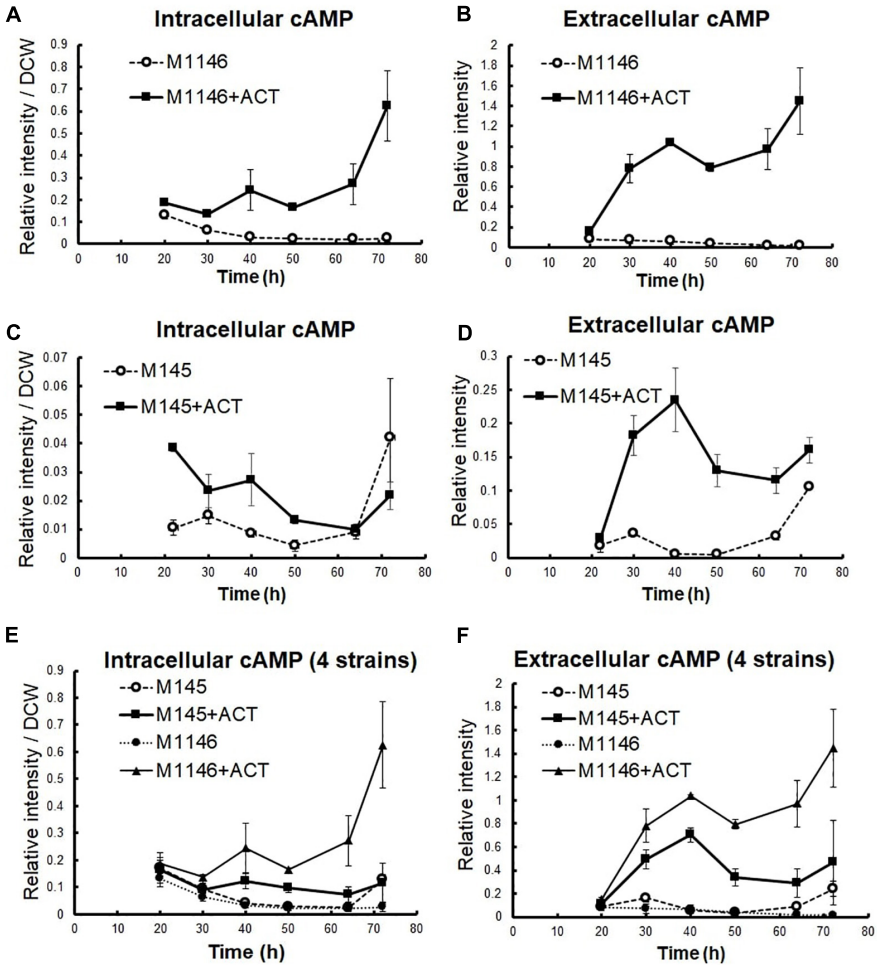


FIGURE 3 | cAMP level measurement in four *S. coelicolor* strains. Error bar means standard deviation from three replicates. **(A)** Intracellular cAMP-level comparison between M1146 (dot line chart) and M1146 + ACT (black line chart). **(B)** Extracellular cAMP-level comparison between M1146 (dot line chart) and M1146 + ACT (black line chart). **(C)** Intracellular cAMP-level comparison between M145 (dot line chart) and M145 + ACT (black line chart). **(D)** Extracellular cAMP-level comparison between M145 (dot line chart) and M145 + ACT (black line chart). **(E)** Intracellular cAMP-level comparison between four strains: M145 (dot line chart with blank circle), M145 + ACT (black line chart with square), M1146 (dot line chart with black circle), and M1146 + ACT (black line chart with triangle). **(F)** Extracellular cAMP-level comparison between all four strains, in an independent replication experiment: M145 (dot line chart with blank circle), M145 + ACT (black line chart with square), M1146 (dot line chart with black circle), and M1146 + ACT (black line chart with triangle).

cAMP stability in cell-free medium and cAMP uptake by M1146 at 1 h and 2 h after cAMP supplementation was confirmed (**Supplementary Figures 5A,B**). Interestingly, when M1146 and M1146 + ACT were supplemented with 10 μM cAMP at the mid–log phase (20 h), cell growth increased; the cell masses after supplementation were 1.20- and 1.22-fold higher in M1146 and M1146 + ACT, respectively (**Figure 5A**). Moreover, intracellular Act production increased in M1146 + ACT (**Figure 5B** and **Supplementary Figure 6A**) starting after 50 h, and at 72 h, the total Act production increase was 1.1-fold on average; this is a very minor increase, but statistically significant ($p < 0.05$; **Figure 5B**). Interestingly, extracellular supplemented cAMP rapidly decreased in M1146, but showed a slower decrease in M1146 + ACT (**Figures 5C,D**). After 50 h, a renewed increase of extracellular cAMP was observed with or without the addition of cAMP (**Figure 5C**), but the extracellular cAMP accumulation was higher in the cAMP-supplemented M1146 + ACT culture. The intracellular cAMP showed an increase in both strains with addition of external cAMP (**Figure 5D**).

To evaluate the effects of cAMP on gene expression levels, RNAseq-based transcriptome analysis of cAMP-supplemented cultures of M145 was performed, and DEGs were identified (**Supplementary Table 6**) based on DEG criteria. cAMP 3 μM was added to M145 cultures after 48 h of the growth and samples were taken at 50 h. Interestingly, cAMP addition increased expression of two genes from the Act biosynthetic gene cluster; SCO5086 (ketoacyl reductase) increased most dramatically by 3.05-fold on average, and SCO5085 (Act biosynthesis pathway-specific activator actII-ORF4) increased by 1.93-fold on average (**Figure 5E**). Act biosynthesis is known to be tightly controlled by *actII*-ORF4 (Gramajo et al., 1993), and Act production starts after the increased expression of this pathway-specific activator

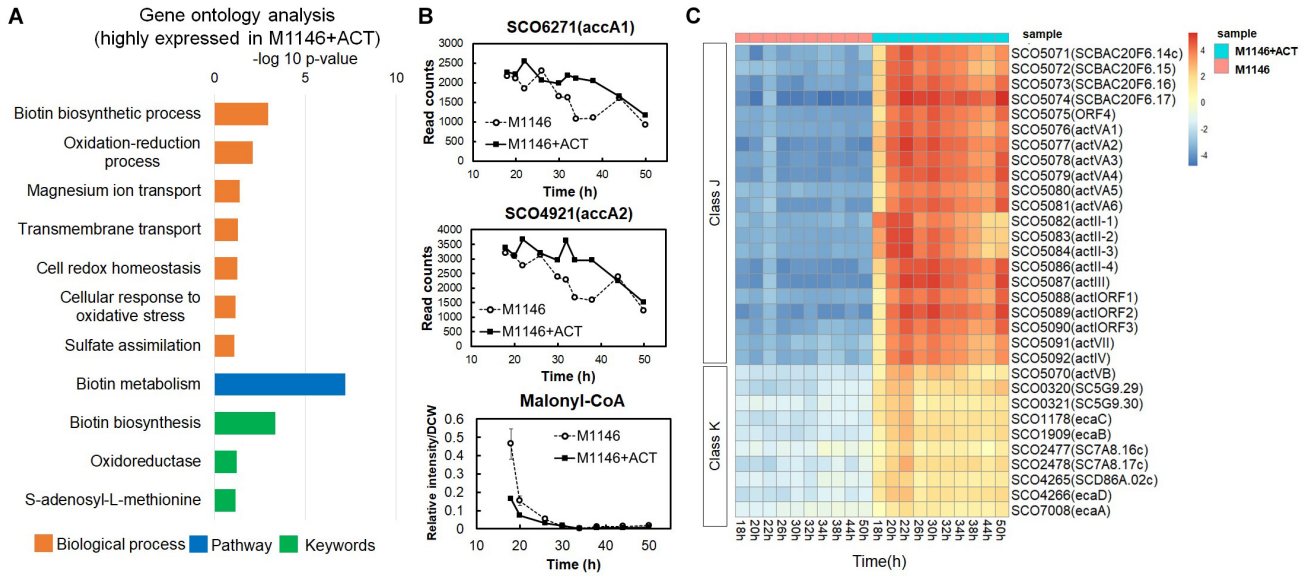


FIGURE 4 | Transcriptome analysis of M1146 and M1146 + ACT. **(A)** Gene set enrichment analysis by means of highly expressed genes from identified differentially expressed genes (DEGs) in time-course transcriptome analysis of M1146 and M1146 + ACT. Biological process, pathway, and keywords were used to define the gene sets. **(B)** Gene expression level comparison of two acetyl-CoA carboxylases and metabolite-level comparison of malonyl-CoA between M1146 (dot line chart) and M1146 + ACT (black line chart). The error bar in malonyl-CoA–levels line chart means standard deviation from three replicates. **(C)** K-means clustering of transcriptome analysis of M1146 and M1146 + ACT (k = 20). Gene expression trends in the focus classes J and K are shown here. K-means clustering results using all genes are shown in **Supplementary Figure 2**.

TABLE 1 | List of selected class K genes in heatmap analysis of time-course RNAseq of M1146 and M1146 + ACT.

Locus tag	Product	Note	Reference
SCO1178	NAD-dependent epimerase/dehydratase	SoxR regulon	Naseer et al., 2014
SCO1909	Monooxygenase	SoxR regulon	
SCO4266	Oxidoreductase	SoxR regulon	
SCO7008	ABC transporter ATP-binding protein	SoxR regulon	
SCO2478	Flavoprotein reductase	SoxR regulon	
SCO0320	Quinone oxidoreductase	Potential SoxR regulon	
SCO0321	Carboxylesterase	Potential SoxR regulon	

(Gramajo et al., 1993). It is also reported that increasing the transcription of *actII*-ORF4 results in the overproduction of Act (Gramajo et al., 1993). The increase of *actII*-ORF4 in our results is in line with these previous reports and provides a possible direct link between cAMP supplementation and Act production at the level of gene expression.

Interestingly, following the rapid disappearance of supplemented cAMP from the culture medium, a minor accumulation of extracellular cAMP was observed in M1146 at 72 h (**Figure 5C**). As we had previously observed that antibiotic production was correlated with extracellular cAMP accumulation, we hypothesized that this cAMP accumulation might be related to the production of other secondary metabolites in M1146—while this "clean" strain lacks the capacity to produce the four major antibiotics, 18 other potential biosynthetic gene clusters are known to be still present in the genome (Bentley et al., 2002). To understand if the effect of cAMP addition on the increase of Act production is accompanied by consequences for the production of other secondary metabolites, ethyl acetate extracts of cultures of M1146 and M1146 supplemented with cAMP were measured by LC-QTOF (Quadrupole Time-OF-Flight) MS. As shown in previous studies, antibiotic production heavily depends on the utilized medium (Cihak et al., 2017; Lim et al., 2018). Based on previous data on *S. coelicolor* secondary metabolites, candidate peaks were explored, and two secondary metabolites, namely, germicidin A (theoretical mass: 197.11777, observed mass: 197.118462, mass error: 4.0 ppm) and germicidin B (theoretical mass: 183.10212, observed mass: 183.102920, and mass error: 3.5 ppm), were detected in the extracts (**Figure 5F**). The annotation was confirmed by MS/MS analysis of each peak (**Supplementary Figures 6B,C**), and both MS/MS spectra matched the MS/MS spectra from a previous study (Cihak et al., 2017). Based on the peak areas, the production of the two germicidins significantly increased ($p < 0.01$) in the cAMP-supplemented culture (**Figure 5G**).

Metabolomics and Transcriptomics Analysis of cAMP-Supplemented *S. coelicolor* Cultures

In order to understand how cAMP improves secondary metabolite production and cell growth, comparative metabolome analysis of M1146 and M1146 + ACT with and without cAMP supplementation was performed (**Figure 6A** and **Supplementary Figure 7**). We measured a total of 98 metabolites. The analysis

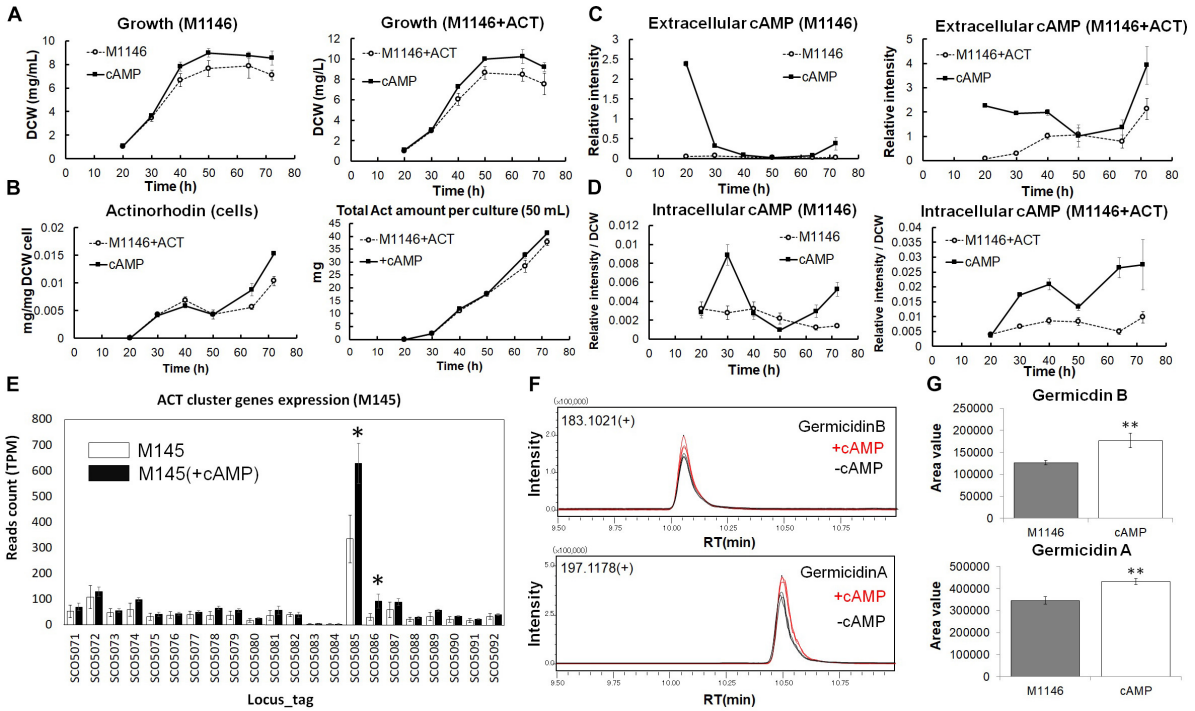


FIGURE 5 | Analysis of cAMP-supplemented to M1146 and M1146 + ACT. **(A)** Cell growth (mg/mL) comparison between the control sample (dot line chart) and cAMP-supplemented sample (black line chart) in M1146 and M1146 + ACT. cAMP was supplemented at the time of the first sampling (20 h). The left panel shows the growth comparison between M1146- and cAMP-supplemented M1146, whereas the right panel shows the growth comparison between M1146 + ACT- and cAMP-supplemented M1146 + ACT. Error bars indicate the standard deviation from three replicates. **(B)** Actinorhodin production (mg/L) comparison between the control sample (dot line chart) and cAMP-supplemented sample (black line chart) in M1146 + ACT during the growth. The left panel shows intracellular actinorhodin production comparison between M1146 + ACT- and cAMP-supplemented M1146 + ACT, whereas the right panel shows total actinorhodin (inside and outside the cell, per 50-mL culture) comparison between M1146 + ACT- and cAMP-supplemented M1146 + ACT. Error bars indicate the standard deviation from three replicates. **(C)** Extracellular cAMP-level comparison between the control sample (dot line chart) and cAMP-supplemented sample (black line chart) during the growth. The left panel shows extracellular cAMP-level comparison between M1146- and cAMP-supplemented M1146, whereas the right panel shows extracellular cAMP-level comparison between M1146 + ACT- and cAMP-supplemented M1146 + ACT. Error bars indicate the standard deviation from three replicates. **(D)** Intracellular cAMP-level comparison between the control sample (dot line chart) and cAMP-supplemented sample (black line chart) during the growth. The left panel shows intracellular cAMP-level comparison between M1146- and cAMP-supplemented M1146, whereas the right panel shows intracellular cAMP-level comparison between M1146 + ACT- and cAMP-supplemented M1146 + ACT. Error bars indicate the standard deviation from three replicates. **(E)** Gene expression level comparison between non–cAMP-supplemented M145 (white bar graph) and cAMP-supplemented M145 (black bar graph). Read count normalized by TPM (transcripts per million) was used, and error bar means standard deviation from three replicates. Gene names are listed in **Supplementary Table 7**. Asterisk means significant difference ($p < 0.01$). **(F)** Chromatograms of germicidin B and germicidin A from ethyl acetate extracts from M1146 culture with (red line) and without (black line) the addition of cAMP. **(G)** Area value comparison between non–cAMP-supplemented and cAMP-supplemented M1146. Error bar means standard deviation from three replicates. Double asterisk means significant difference ($p < 0.01$).

here focuses on metabolites, which consistently increased in M1146 and M1146 + ACT when cAMP was supplemented.

Interestingly, levels of purine bases, such as hypoxanthine guanine and xanthine, substantially increased at 64 h and 72 h when cAMP was supplemented (**Figures 6A,B**). Therefore, we hypothesized that increased guanine, xanthine, and hypoxanthine levels and the purine base–utilizing reactions might be a key factor mediating increased cell growth and antibiotic production improvement following cAMP supplementation. This would be in line with earlier observations of a link between purine metabolism and antibiotic production, through their involvement in the synthesis of guanosine tetraphosphates and pentaphosphates (p)ppGpp and the second messenger cyclic-di-GMP (c-di-GMP; Sivapragasam and Grove, 2019).

To support the metabolome analysis result and to understand the phenomenon caused by cAMP supplementation at the level of gene expression, transcriptome analysis of cAMP was performed. To distinguish between the direct effects of cAMP supplementation and indirect effects of Act production, we performed transcriptome analysis on M1146 with addition of cAMP. Based on the same differential expression criteria as described above, DEGs were identified between time points 0, 1, and 2 h (**Supplementary Figure 8** and **Supplementary Table 8**) to identify genes that showed different expression in response to cAMP supplementation. Here, gene expression related to purine metabolism was mapped to the metabolic pathway (**Figure 7**). The results showed increased expression levels of genes, encoding enzymes involved in the synthesis of 5-aminoimidazole-4-carboxamide ribonucleotide (a shared

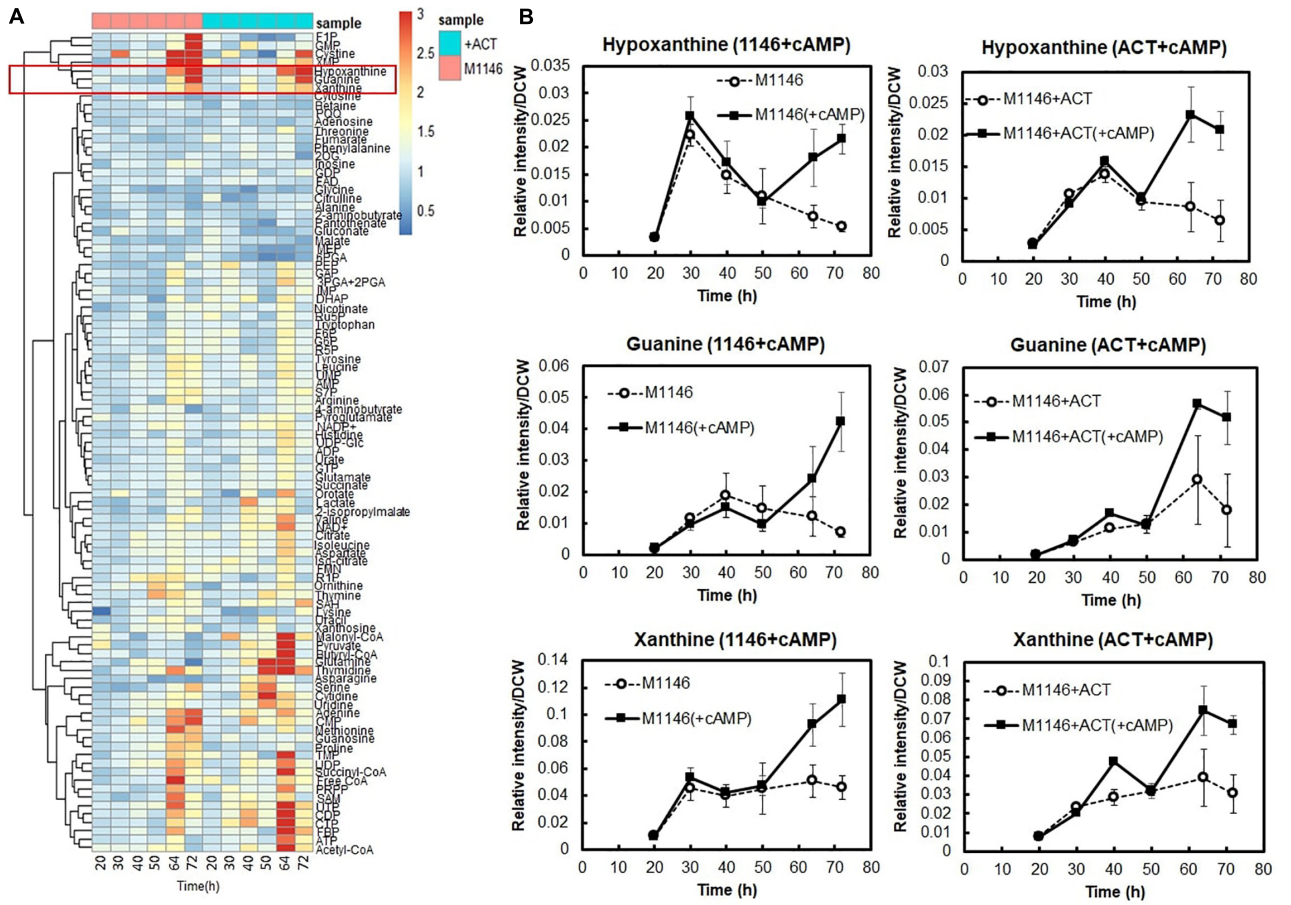


FIGURE 6 | Metabolome analysis of cAMP-supplemented M1146 and M1146 + ACT. **(A)** Intracellular metabolome analysis of cAMP-supplemented M1146 and M1146 + ACT during growth. The metabolite amount difference is shown by fold change of area value (cAMP supplemented/non–cAMP supplemented) in heatmap clustering. Average values from three replicates were used to calculate the fold change values. This heatmap consists of data of cAMP-supplemented of M1146 (pink, left six panels) and M1146 + ACT (blue, right six panels). The metabolites are listed in **Supplementary Table 2**. **(B)** Intracellular selected metabolite-level (hypoxanthine, guanine, and xanthine) comparison between non–cAMP-supplemented and cAMP-supplemented culture. The left figure shows the comparison between M1146 (dot line, circle) and cAMP-supplemented M1146 (black line, square), whereas the right figure shows the comparison between M1146 + ACT (dot line, circle), and cAMP-supplemented M1146 + ACT (black line, square). Error bar means standard deviation from three replicates.

precursor of purines) in response to cAMP supplementation: SCO1254 (2SCG1.29), SCO4068 (*purD*), SCO4087 (*purM*), and SCO4813 (*purN*). This transcriptome trend is consistent with purine bases (guanine, xanthine, hypoxanthine, and adenine) increasing at the metabolome level. In addition, expression levels of genes encoding salvage pathway enzymes [SCO1514, *apt*, and SCO3405, *hprT* (FDR = 0.077 and log2FC = 0.64)] increased. These phenomena at transcriptome level are consistent with the metabolome-based observation that guanine and GMP levels increase as cAMP is supplemented (**Figure 7**). Therefore, among the candidate key metabolites (guanine, xanthine, and hypoxanthine) identified to be increasing in the metabolome analysis of cAMP-supplemented culture, it was suggested that guanine or a guanine-related reaction may be mediating the effect of cAMP on cell growth and antibiotic production.

To test this hypothesis, we supplemented M1146 + ACT at the mid–log phase with 20 μM 7-methylguanine, which is an analog of guanine that inhibits guanine utilizing reactions by competitive inhibition (Goodenough-Lashua and Garcia, 2003; Fernandez et al., 2010). In addition, 20 μM 7-methylguanosine, which is a possible competitive inhibitor of guanosine utilizing reactions (Pathak et al., 2005), was also supplemented independently at the mid–log phase of M1146 + ACT, to test the idea that guanosine instead of guanine might be the active metabolite in this regulatory system (**Figures 8A,B**). Interestingly, both 7-methylguanine, and 7-methylguanine supplementation impaired Act production by 2.22- and 1.85-fold, respectively. In addition, both 7-methylguanine and 7-methylguanosine supplementation impaired guanine levels by 2.81- and 2.31-fold, respectively (**Figure 8D**). In order to identify whether this impairment occurs downstream or upstream of the positive effect of cAMP supplementation, cAMP was also supplemented to the two inhibitor-treated cultures. cAMP supplementation recovered about 50% of Act production only in the 7-methylguanosine–supplemented culture (**Figures 8A,B**). Cell growth was monitored at 72 h

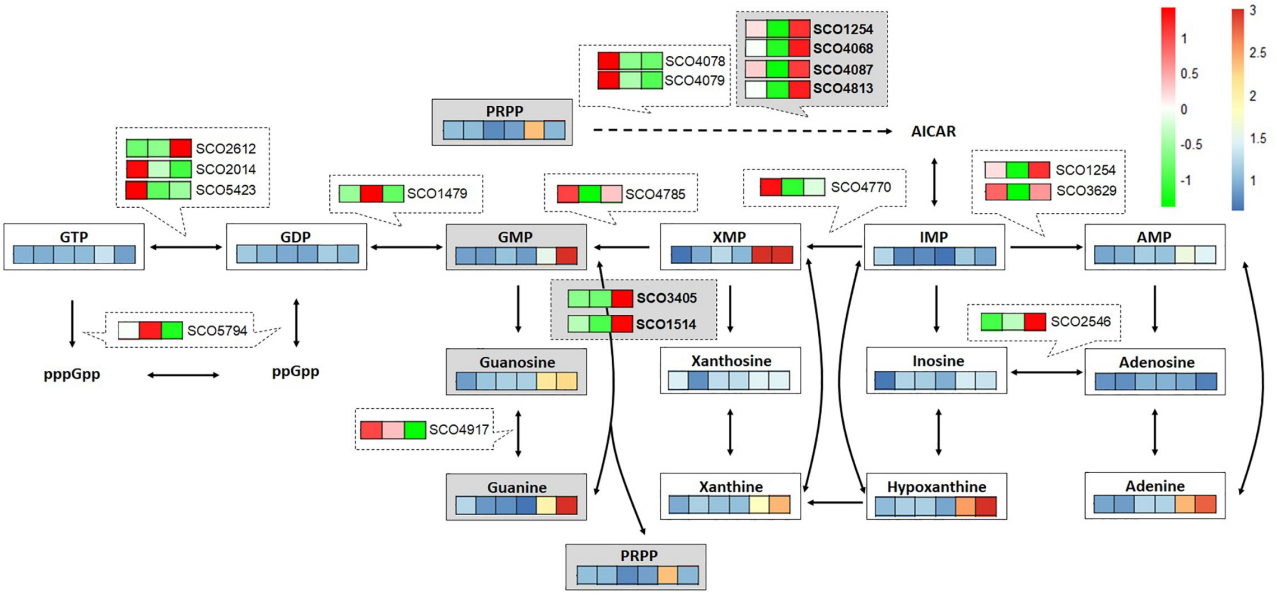


FIGURE 7 | Multi-omics analysis of cAMP-supplemented M1146 mapped to purine metabolism. Time course (20, 30, 40, 50, 64, and 72 h: from left to right) of metabolite-level fold change between non–cAMP-supplemented M1146- and cAMP-supplemented M1146 is shown on a color scale from blue to red. Average values from three replicates were used to calculate the fold change. Time-course gene expression (0, 1, and 2 h: from left to right) changes in response to cAMP supplementation are shown on a color scale from green to red. The average value of read counts was normalized by autoscaling to see the difference well.

after cAMP was supplemented to the 7-methylguanine– or 7-methylguanosine–treated M1146 + ACT. 7-methylguanine–treated cultures after cAMP supplementation still showed a decreased amount of cells ($p < 0.01$), whereas cAMP supplementation of 7-methylguanosine–treated cultures restored the cell amount to untreated levels ($p < 0.05$; **Figure 8C**). Guanine levels in the cells in inhibitor- and cAMP-supplemented cultures were also monitored and compared at 72 h; here, the addition of cAMP did not restore the guanine level in 7-methylguanine–supplemented culture, whereas addition of cAMP restored the guanine level in 7-methylguanosine–supplemented cultures (by 1.66-fold, statistically not significant; **Figure 8D**). This is in partial agreement with our original hypothesis that an increase in guanine is necessary for cell growth and antibiotic production improvement by cAMP supplementation. In summary, the addition of cAMP to 7-methylguanine–supplemented culture did not restore Act production and cell growth, whereas the addition of cAMP to 7-methylguanosine–supplemented culture restored the impaired Act production and cell growth. Therefore, the negative effect of 7-methylguanine supplementation occurs upstream of the effects of cAMP supplementation, and we suggest that the reaction(s) inhibited by 7-methylguanine are mediating the positive effect on growth and antibiotic production in *S. coelicolor*.

CONCLUSION

In this study, a metabolomics- and transcriptomics-based approach was applied to elucidate the effect of Act production in *S. coelicolor*. The analysis showed that Act production was highly correlated with an increase in extracellular cAMP levels, and cAMP supplementation was found to increase antibiotic production and cell growth. Further multi-omics analysis of cAMP-supplemented cultures showed that guanine levels increased in response to cAMP supplementation, and inhibition of guanine utilizing reactions by the analog 7-methylguanine confirmed that some reaction or reactions inhibited by 7-methylguanine mediate the positive effect of cAMP supplementation on growth and antibiotic production.

MATERIALS AND METHODS

Bacterial Strain, Medium, Growth Condition

Bacterial strains were grown in 50 mL liquid minimum nutrition medium with same composition as previous study (Nieselt et al., 2010) in well-siliconized 250-mL flasks containing stainless-steel springs. For inoculation, 1×10^9 colony-forming units of spores were inoculated to the 50 mL medium. Incubation speed and temperature were set to 220 rpm and 30°C, respectively. To measure growth, cells were collected in 2 mL Eppendorf tube and centrifuged to discard supernatant and washed well by dH_2O and lyophilized by freeze drying for measurement of dry cell weight (DCW).

Antibiotic Production Quantification

Procedure to quantify extracellular actinorhodin was based on previous study (Nieselt et al., 2010). Cells for intracellular actinorhodin quantification were sampled by fast filtration with

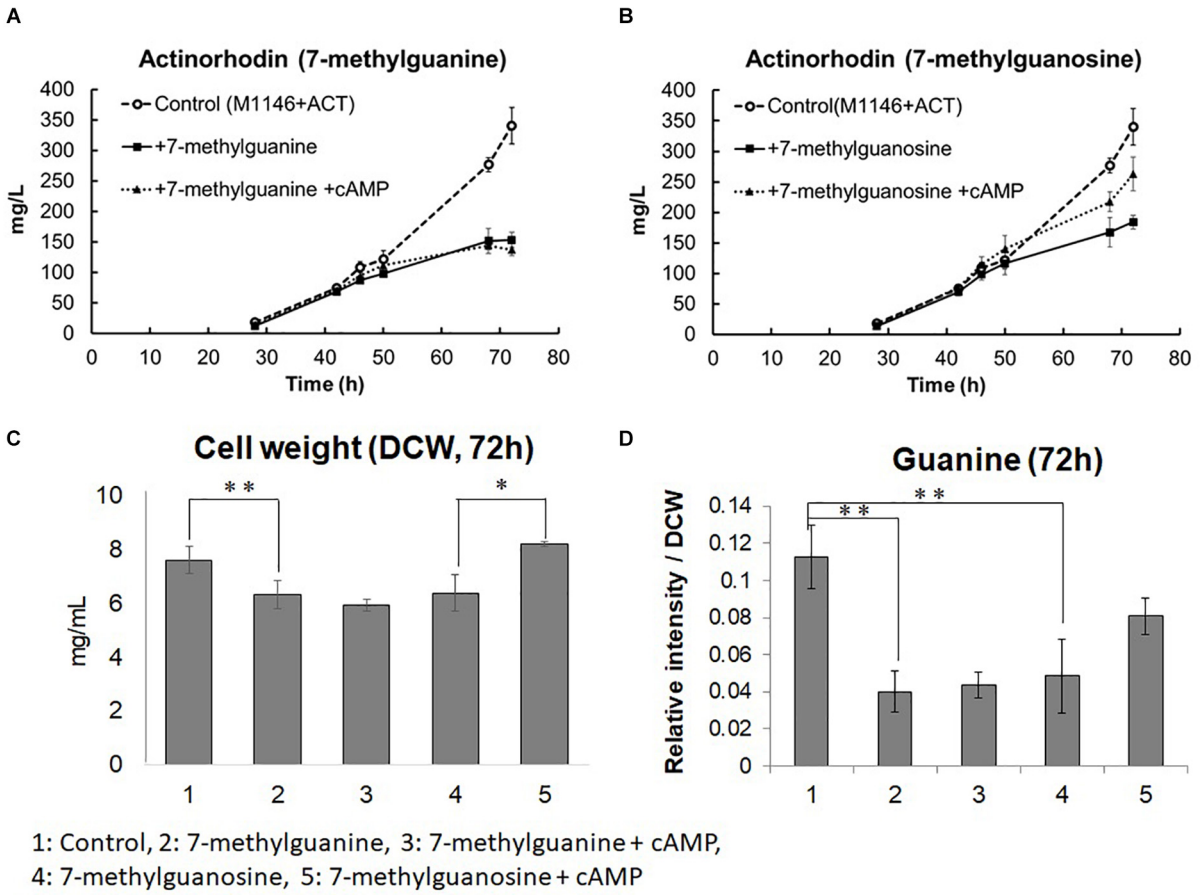


FIGURE 8 | Actinorhodin production inhibition by 7-methylguanine and 7-methylguanosine in M1146 + ACT and restoration by addition of cAMP. **(A)** Actinorhodin production in control (M1146 + ACT; semi dotted line, circle), 7-methylguanine (20 μM)-supplemented (black line, square), and 7-methylguanine (20 μM)– and cAMP (10 μM)–supplemented (dotted line, triangle) cultures. Error bar means standard deviation from three replicates. Circle means control sample, square means 7-methylguanine–supplemented, and triangle means 7-methylguanine– and cAMP-supplemented culture. **(B)** Actinorhodin production in control (M1146 + ACT; semi dotted line, circle), 7-methylguanosine (20 μM)-supplemented (black line, square), and 7-methylguanosine (20 μM)– and cAMP (10 μM)–supplemented (dotted line, triangle) cultures. Error bar means standard deviation from three replicates. **(C)** Cell weight (DCW: mg/mL) comparison at 72 h. Error bar means standard deviation from three replicates. Asterisk means significant difference (**$p < 0.01$, *$p < 0.05$). **(D)** Guanine amount comparison at 72 h. Error bar means standard deviation from three replicates. Asterisk means significant difference (**$p < 0.01$).

5.0-μm pore size nylon membrane filter and subjected to vacuum fast filtration, and the cells were quenched in 15-mL Falcon tube by liquid nitrogen and kept at −80°C until extraction. The cells were extracted by 4 mL of mix-solvent solution [methanol, chloroform, and water 5:2:2 (vol/vol/vol)] with three cycles of freeze and thaw cycle (freezing at −80°C and thawing at −30°C). The extracts were centrifuged at 10,000 rpm for 10 min, and 2 mL supernatant was transferred to new 15-mL Falcon tube, and 1 mL ultrapure water was added to the supernatant and vortexed well. The mixture was separated to two layers by centrifugation at 10,000 rpm for 10 min, and the upper layer was used for intracellular actinorhodin quantification. The colored extracts were diluted with the same amount of 1 M NaOH solution, and OD608 was measured by photometer. Here, a culture with an already known actinorhodin concentration was extracted exactly the same way as was done for intracellular actinorhodin extraction, and the OD608 was measured. The value was used to quantify the unknown amount of intracellular actinorhodin.

Culture Sampling for Metabolome Analysis

Based on the growth curve of each strain, more than 10 mg DCW cultures were sampled by nylon membrane filter with 5.0-μm pore size and 47-mm diameter and subjected to vacuum fast filtration. Cells on the filter were washed using twice amount of 3.5% (vol/vol) $NaHCO_3$ solution to wash the medium component away. The filter with cells was immediately quenched in 15-mL Falcon tubes by liquid nitrogen and kept at −80°C until extraction. One-milliliter culture was sampled for extracellular metabolome analysis and medium components analysis and Act quantification. The culture was centrifuged at 10,000 rpm for 10 min, and supernatant was kept at −80°C until extracellular

metabolites extraction and medium components consumption analysis. For time-course metabolome analysis and medium components analysis, 20, 30, 40, 50, 64, and 72 h were selected for sampling points based on the growth curve. In the case of comparing the metabolome analysis data with transcriptome analysis data, 18, 20, 22, 26, 30, 32, 34, 38, 44, and 50 h were selected for extracellular metabolome analysis, whereas less time points, 18, 22, 26, 30, 34, 38, 44, and 50 h, were selected for intracellular metabolome analysis due to lack of sampling volume.

Medium Components (Glucose, Glutamate, and Phosphate) Consumption Quantification

Medium components, glucose, glutamate, and phosphate, were measured by commercially available quantification kits. F-kit-D-glucose (J. K. International), L-glutamate assay kit Yamasa NEO (Yamasa), and PiBlue phosphate assay kit (BioAssay Systems) were used for quantification of glucose, glutamate and phosphate in medium, respectively, by following manufacturing guides.

Extraction of Cells and Medium for Metabolome Analysis

For intracellular metabolome analysis, 4 mL of mix solvent [methanol, chloroform, and water ratio 5:2:2 (vol/vol/vol)] with 50 μg/L (+) 10-camphorsulfonic acid as an internal standard was added to frozen cells in 15-mL Falcon tubes and extracted with three cycles of freeze and thawing cycles (freezing at –80°C, thawing at –30°C, vortexing for 10 s, and sonication for 10 s). The extracted cells were centrifuged at 10,000 rpm, 4°C for 10 min, and the supernatant was used for the following extraction procedures. To the supernatant (2 mL), 1 mL ultrapure water was added and vortexed for 5 s and separated to two layers by centrifugation (10,000 rpm, 4°C) for 10 min. The upper polar phase was filtered by 0.2 μm PTFE hydrophilic membrane filter, and the 2 mL extract solution was applied to centrifugal concentration by spin dryer for 2 h, and the samples were lyophilized overnight. The lyophilized sample was kept at −80°C until analysis. The lyophilized sample was reconstituted with 400 μL ultrapure water when analyzed. The extracted cells were washed by ultrapure water and subjected to freeze drying to calculate DCW used for intracellular metabolome analysis and normalize metabolite abundance by the cell amount.

For extracellular metabolome analysis, 50 μL culture supernatant was added to 1.8 mL mix solvent with the same composition for intracellular metabolites extraction. The mixture was vortexed for 5 s and kept at -30°C, and the extracts were centrifuged at 10,000 rpm and 4°C for 10 min. The supernatant (1.3 mL) was mixed with 0.65 mL ultrapure water and vortexed for 5 s and separated to two layers by centrifugation at 10,000 rpm and 4°C for 10 min, and supernatant was filtered by 0.2 μm PTFE hydrophilic membrane filter. Filtered sample (1 mL) was applied to centrifugal concentration by spin dryer for 1 h, and the sample was lyophilized overnight, and lyophilized sample was kept at -80°C until analysis. The lyophilized sample was reconstituted with 200 μL ultrapure water when analyzed. Intracellular and extracellular reconstituted sample solution (40 μL) was transferred to LC vial for following LC-MS/MS analysis.

LC-MS/MS Analysis for Intracellular and Extracellular Metabolome Analysis

In this study, two kinds of LC-MS/MS platform were employed for achieving a wider range of metabolite coverage. One is ion-pair LC-MS/MS with negative ionization mode, which is previously described (Nitta et al., 2017). Acquired data from this analysis were analyzed by Lab solution (Shimadzu).

For sugar phosphate isomer separation, different gradient (see below) was employed for better chromatographic separation, whereas other parameters were set to be the same. The gradient is as follows: Percentage of mobile phase B was held at 0% for 1 min and raised to 50% in 30 min and raised to 100% in a minute. After holding at 100% for 1 min, the percentage was decreased to 0% in a minute and held at 0% for 6 min for column equilibration for the next analysis. All metabolites' abundance was normalized by area value of internal standard [(+) 10-camphorsulfonic acid] and the DCW used for metabolite extraction.

The other platform is LC-MS/MS in positive ionization mode with ESI (electrospray ionization) mode. Nexera X2 UHPLC coupled to LCMS 8050 (Shimadzu) with Discovery HS F5-3 (3 μm, 150 × 2.1 mm) column was used. Formic acid 0.1% (vol/vol) in ultrapure water was used for mobile phase A, whereas acetonitrile was used for mobile phase B.

Column oven temperature was set to 40°C, and injection volume was set to 3 μL. The chromatographic separation was conducted by gradient mode as follows. The total flow rate was set to 0.2 mL/min, and percentage of mobile phase B was held at 0% for 5 min and raised with a gradient for 4%/min until 40% and held for 10 min and then raised with a gradient 60%/min until 100%. After holding at 100% for 2.5 min, the percentage was decreased to 0% in a minute and held at 0% for 6 min for column equilibration for the next analysis. Mass spectrometer parameters were set to as follows: DL (desolvent line) temperature was set to 250°C; heating block temperature was set to 400°C; nebulizer gas flow was set to 3 L/min; drying gas flow rate was set to 10 L/min; heating gas flow rate was set to 10 L/min; and interface temperature was set to 400°C. All analyses were performed by MRM mode. Acquired data from this analysis were analyzed by Lab solution (Shimadzu). All metabolites' abundance was normalized by area value of internal standard [(+) 10-camphorsulfonic acid]. Metabolome analysis data have been deposited to the Metabolights public repository under accession numbers MTBLS1984 and MTBLS2025.

Culture Sampling for RNAseq Analysis

Before sampling, all solutions used for RNA extraction was autoclaved two times to avoid RNase contamination. Based on the growth curve, more than 2.0 mg DCW cell cultures were sampled and mixed with twice the amount of RNA protection reagent (Qiagen), immediately. The cells were vortexed for 10 s, kept

for 5 min at room temperature, and centrifuged at 10,000 rpm, 20°C for 10 min. Supernatant was eliminated as much as possible, and the cells were quenched by liquid nitrogen immediately and kept at −80°C until RNA isolation. For time-course RNAseq of M1146 and M1146 + ACT, the samples were collected at 18, 20, 22, 26, 30, 32, 34, 38, 44, and 50 h. For RNAseq of cAMP-supplemented M145, 3 μM cAMP was supplemented to M145 at 48 h, and the samples were collected at 50 h. For RNA-Seq of cAMP-supplemented M1146, 10 μM cAMP was supplemented at 20 h, and the samples were collected at 21 and 22 h.

Cell Lysis, RNA Extraction, Purification, and Quality Check

Cells were resuspended by double-autoclaved empty tip, and 0.17 mL of 15 mg/mL lysozyme was added to the cells and incubated at 30°C for 10 min. The solutions including cells were transferred to tube containing lysing matrix E (MPbio medicals) and 0.6 mL RLT buffer (Qiagen) supplemented with β-mercaptoethanol (100:1, vol/vol) and vortexed for 5 s. Three pulses were applied to the tube by Fast Prep (6.5 m/s, 30 s), whereas the tubes were kept on ice between pulses for 30 s. The tubes were centrifuged at 10,000 rpm and 4°C for 1 min, and the lysate was recovered. After centrifuging heavy Phase Lock Gel (PLG) tube to pack the resin for 1 min, 0.65 mL recovered lysate was transferred to the heavy Phase Lock Gel (PLG) tube; 0.65 mL isoamyl alcohol and acid phenol mixture were added to the lysate in heavy PLG tube and mixed by inversion for 1 min. The heavy PLG tubes containing extracts were centrifuged for 5 min, and superior aqueous phase was recovered. RNA purification was performed with Direct-zol RNA MiniPrep Plus (Zymo research). After 0.6 mL ethanol was added to recovered 0.6 mL of aqueous phase and mixing by pipette, Zymo-spin IIICG column was assembled to collection tube, and 0.6 mL mixture solution was transferred to the assembled column-collection tube and centrifuged at 10,000 rpm for 30 s. Flow-through from the collection tube was discarded, and the remaining 0.6 mL of mixture of aqueous extract and ethanol was transferred to the assembled column-collection tube, consequently centrifuged at 10,000 rpm for 30 s. The flow-through in the collection tube was discarded again. RNA wash buffer completed with ethanol (0.4 mL) was added to the column and centrifuged at 10,000 rpm for 30 s, and the flow-through was discarded. To the column matrix, 80 μL of DNase (6 U/μL) was added and incubated at 30°C for 15 min. After incubation, RNA was washed for two cycles (adding 0.4 mL of Direct-zol RNA prewash to the column, centrifuging at 10,000 rpm for 30 s and discarding the flow through). To the column, 0.7 mL RNA wash buffer was added and centrifuged for 2 min to ensure complete removal of wash buffer. Consequently, the column was transferred to a new RNase-free tube; 50 μL of DNase/RNase-free water was added and incubated at room temperature for 1 min. The column with sample containing water was centrifuged at 10,000 rpm for 1 min, and the tube with the RNA sample was frozen by liquid nitrogen until the next step. RNA concentration was quantified by NanoDrop (ThermoFisher Scientific), and RNA quality based on RNA integrity number was evaluated by Bio-analyzer (Agilent Technologies).

Ribosomal RNA Deletion, cDNA Library Construction, and Sequencing

All rRNA deletion, strand-specific cDNA library construction, and the sequencing experiment were outsourced to Vertis Co. (Germany) and Genewiz Japan (Japan) and performed by their in-house–developed methods. Time-course RNAseq analysis of M1146 and M1146 + ACT and RNAseq of cAMP-supplemented M145 was performed by Vertis (Germany), and cDNA pools were sequenced on an Illumina NextSeq 5000 system using 75-bp length with 1 × 10-M reading depth. Time-course RNAseq analysis of cAMP-supplemented M1146 was performed by Genewiz (Japan), and cDNA pools were sequenced by Illumina HiSeq X-ten system using 150-bp length with 2 × 10-M reading depth.

RNASeq Data Analysis

Sequencing data were obtained from Vertis Co. and Genewiz Co. in FASTQ format. Reference genome data of *S. coelicolor* were downloaded from NCBI, and generation of reference genome library and read mapping to the reference genome library were performed by STAR version 2.7 (Dobin et al., 2013) on a Linux PC. Obtained SAM files were converted to BAM file by Samtools (Li et al., 2009). Defining gene model, reads counting, normalization, and differential gene expression analysis procedures were based in a previous study (Love et al., 2015). All data were transferred to R version 3.6.1, and BAM file list was generated by R package Rsamtools; reference gene model was generated by R package GenomicFeatures (Lawrence et al., 2013); sequencing reads were counted R package GenomicAlignments (Lawrence et al., 2013), and finally RLE normalization and differential gene expression analysis were performed by R package DESeq2 (Love et al., 2014). For defining differential gene expression, a threshold of FDR less than 5% and log2 FC greater than 0.5 was chosen. RNAseq analysis data have been deposited to the NCBI Gene Expression Omnibus public repository under accession numbers GSE155796, GSE158810, and GSE158811.

GSEA

Gene Set Enrichment Analysis was performed using the DAVID Functional Annotation Bioinformatics Microarray Analysis version 6.8 (Huang da et al., 2009a,b). The GSEA was performed based on biological process, pathway, and keywords.

Multivariate Analysis (Heatmap Analysis and K-Means Clustering)

Heatmap analysis of metabolome data was conducted using the R package pheatmap under R version 3.6.1. K-means clustering of time-course transcriptome data was performed using iDEP90 (Ge et al., 2018).

cAMP Addition and Analysis of the Culture Extracts by LC-QTOF MS

cAMP (10 μM) was supplemented to culture of M1146 at the mid–log phase, and culture was sampled at 72 h. Medium (20 mL) was vortexed well with 20 mL ethyl acetate. After centrifugation (10,000 rpm, 4°C) for 10 min, the 20 mL supernatant was subjected to centrifugal concentration and lyophilized by freeze drying. The lyophilized sample was reconstituted with 400 μL 50% acetonitrile, and the 40 μL solution was transferred to LC vial and subjected to LC-QTOF MS analysis. LC-QTOF MS 9030 (Shimadzu) was used for secondary metabolite detection and Nexera XR UHPLC (Shimadzu), and the Inert Sustain AQ-C18 (3 μm, 150 × 2.1 mm; GLscience) column was used for chromatographic separation. Formic acid 0.1% in ultrapure water (vol/vol) was used for mobile phase A. whereas acetonitrile with 0.1% formic acid (vol/vol) was used for mobile phase B. The chromatographic separation was performed by gradient mode. Gradient condition is as follows. The ratio of mobile phase B was kept at 0% for 3 min, raised until 100% by 8 min, and kept at 100% for 2 min. The B ratio was reduced to 0% by 3 min and kept at 0% for the next analysis. The flow rate was set to 0.2 mL/min. Column oven temperature was set to 40°C. The injection volume was set to 3 μL. For mass spectrometer parameter, the following settings were used. Interface voltage was set to 4 kV, needle voltage was set to 4.5 kV, flow rate of nebulizer gas was set to 2.0 L/min, flow rate of heating gas was set to 10 L/min, flow rate of drying gas was set to 10 L/min, interface temperature was set to 300°C, DL temperature was set to 250°C, and heat block temperature was set to 250°C. Metabolites were ionized in positive ionization mode by DUIS (dual ion source) mode. For secondary metabolite peak discovery, TOF *m/z* range was set to 100 to 700. For specific germicidin analysis, TOF *m/z* range was set to 100–250. For MS/MS analysis, collision energy was set to -15 eV.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The metabolome analysis data and transcriptome analysis data were uploaded to public data repositories "Metabolights database" and "NCBI GEO database". The accession numbers can be found in the article.

AUTHOR CONTRIBUTIONS

EF, SP, ET, and RB conceived the idea of this study. The experiments were designed by all authors. All experiments were performed by KN. The data interpretation was done by KN, ET, SP, RB, and FC. The manuscript was written by all authors. All authors read and approved the final manuscript.

FUNDING

This research was supported by JSPS KAKENHI grant number 19J10382 (KN); JSPS Grants-in-Aid for Scientific Research grant number 20308851 (SP), the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 720793 TOPCAPI—thoroughly Optimized Production Chassis for Advanced Pharmaceutical Ingredients (FC, RB, and ET); University of Manchester (RB); and BBSRC Japan-Manchester Partnering Award (BB/N021975/1; RB and ET).

ACKNOWLEDGMENTS

We thank Associate Professor Masae Kuboniwa and Dr. Akito Sakanaka in Graduate School of Dentistry of Osaka University for the use of the cell homogenizer and NanoDrop for RNAseq experiment. We would also like to thank Dr. Suhui Ye Huang and Dr. Oksana Bilyk at the University of Manchester for assistance in *Streptomyces* genetics and RNA isolation. We would also like to thank Dr. Junko Iida and Mr. Yohei Arao at Shimadzu corporation for assistance to use LC-Q TOF MS. This study will be included in dissertation submitted by KN to Osaka University in partial fulfillment of the requirement for his doctor degree.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.595552/full#supplementary-material

REFERENCES

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147. doi: 10.1038/417141a

Cihak, M., Kamenik, Z., Smidova, K., Bergman, N., Benada, O., Kofronova, O., et al. (2017). Secondary metabolites produced during the germination of *Streptomyces coelicolor*. *Front. Microbiol.* 8:2495. doi: 10.3389/fmicb.2017.02495

Davies, J., Gilbert, W., and Gorini, L. (1964). Streptomycin, suppression, and the code. *Proc. Natl. Acad. Sci. U.S.A.* 51, 883–890. doi: 10.1073/pnas.51.5.883

Dela Cruz, R., Gao, Y., Penumetcha, S., Sheplock, R., Weng, K., and Chander, M. (2010). Expression of the *Streptomyces coelicolor* SoxR regulon is intimately linked with actinorhodin production. *J. Bacteriol.* 192, 6428–6438. doi: 10.1128/JB.00916-10

Derouaux, A., Halici, S., Nothaft, H., Neutelings, T., Moutzourelis, G., Dusart, J., et al. (2004). Deletion of a cyclic AMP receptor protein homologue diminishes germination and affects morphological development of *Streptomyces coelicolor*. *J. Bacteriol.* 186, 1893–1897. doi: 10.1128/jb.186.6.1893-1897.2004

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Fernandez, J. R., Sweet, E. S., Welsh, W. J., and Firestein, B. L. (2010). Identification of small molecule compounds with higher binding affinity to guanine deaminase (cypin) than guanine. *Bioorg. Med. Chem.* 18, 6748–6755. doi: 10.1016/j.bmc.2010.07.054

Gao, C., Hindra, Mulder, D., Yin, C., and Elliot, M. A. (2012). Crp is a global regulator of antibiotic production in *Streptomyces*. *mBio* 3:e00407-12. doi: 10.1128/mBio.00407-12

Ge, S. X., Son, E. W., and Yao, R. (2018). iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* 19:534. doi: 10.1186/s12859-018-2486-6

Gomez-Escribano, J. P., and Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* 4, 207–215. doi: 10.1111/j.1751-7915.2010.00219.x

Goodenough-Lashua, D. M., and Garcia, G. A. (2003). tRNA–guanine transglycosylase from *E. coli*: a ping-pong kinetic mechanism is consistent with nucleophilic catalysis. *Bioorgan. Chem.* 31, 331–344. doi: 10.1016/s0045-2068(03)00069-5

Gosse, J. T., Ghosh, S., Sproule, A., Overy, D., Cheeptham, N., and Boddy, C. N. (2019). Whole genome sequencing and metabolomic study of cave *Streptomyces* isolates ICC1 and ICC4. *Front. Microbiol.* 10:1020. doi: 10.3389/fmicb.2019.01020

Gramajo, H. C., Takano, E., and Bibb, M. J. (1993). Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Mol. Microbiol.* 7, 837–845. doi: 10.1111/j.1365-2958.1993.tb01174.x

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Hwang, S., Lee, N., Jeong, Y., Lee, Y., Kim, W., Cho, S., et al. (2019). Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Nucleic Acids Res.* 47, 6114–6129. doi: 10.1093/nar/gkz471

Ikeda, H., and Omura, S. (1997). Avermectin biosynthesis. *Chem. Rev.* 97, 2591–2610. doi: 10.1021/cr960023p

Kang, D. K., Li, X. M., Ochi, K., and Horinouchi, S. (1999). Possible involvement of cAMP in aerial mycelium formation and secondary metabolism in *Streptomyces griseus*. *Microbiology* 145(Pt 5), 1161–1172. doi: 10.1099/13500872-145-5-1161

Katz, L., and Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu. Rev. Microbiol.* 47, 875–912. doi: 10.1146/annurev.mi.47.100193.004303

Kim, D. J., Huh, J. H., Yang, Y. Y., Kang, C. M., Lee, I. H., Hyun, C. G., et al. (2003). Accumulation of S-adenosyl-L-methionine enhances production of actinorhodin but inhibits sporulation in *Streptomyces lividans* TK23. *J. Bacteriol.* 185, 592–600. doi: 10.1128/jb.185.2.592-600.2003

Kim, H. S., Nihira, T., Tada, H., Yanagimoto, M., and Yamada, Y. (1989). Identification of binding protein of virginiae butanolide C, an autoregulator in virginiamycin production, from *Streptomyces virginiae*. *J. Antibiot.* 42, 769–778. doi: 10.7164/antibiotics.42.769

Kino, T., Hatanaka, H., Hashimoto, M., Nishiyama, M., Goto, T., Okuhara, M., et al. (1987a). FK-506, a novel immunosuppressant isolated from a *Streptomyces*. I. Fermentation, isolation, and physico-chemical and biological characteristics. *J. Antibiot.* 40, 1249–1255. doi: 10.7164/antibiotics.40.1249

Kino, T., Hatanaka, H., Miyata, S., Inamura, N., Nishiyama, M., Yajima, T., et al. (1987b). FK-506, a novel immunosuppressant isolated from a *Streptomyces*. II. Immunosuppressive effect of FK-506 in vitro. *J. Antibiot.* 40, 1256–1265. doi: 10.7164/antibiotics.40.1256

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lim, Y., Jung, E. S., Lee, J. H., Kim, E. J., Hong, S. J., Lee, Y. H., et al. (2018). Non-targeted metabolomics unravels a media-dependent prodiginines production pathway in *Streptomyces coelicolor* A3(2). *PLoS One* 13:e0207541. doi: 10.1371/journal.pone.0207541

Liu, R., Deng, Z., and Liu, T. (2018). *Streptomyces* species: ideal chassis for natural product discovery and overproduction. *Metab. Eng.* 50, 74–84. doi: 10.1016/j.ymben.2018.05.015

Love, M. I., Anders, S., Kim, V., and Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research* 4:1070. doi: 10.12688/f1000research.7035.1

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Mak, S., and Nodwell, J. R. (2017). Actinorhodin is a redox-active antibiotic with a complex mode of action against Gram-positive cells. *Mol. Microbiol.* 106, 597–613. doi: 10.1111/mmi.13837

Millan-Oropeza, A., Henry, C., Blein-Nicolas, M., Aubert-Frambourg, A., Moussa, F., Bleton, J., et al. (2017). Quantitative proteomics analysis confirmed oxidative metabolism predominates in *Streptomyces coelicolor* versus glycolytic metabolism in *Streptomyces lividans*. *J. Proteome Res.* 16, 2597–2613. doi: 10.1021/acs.jproteome.7b00163

Naseer, N., Shapiro, J. A., and Chander, M. (2014). RNA-Seq analysis reveals a six-gene SoxR regulon in *Streptomyces coelicolor*. *PLoS One* 9:e106181. doi: 10.1371/journal.pone.0106181

Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, O. M., et al. (2010). The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics* 11:10. doi: 10.1186/1471-2164-11-10

Nitta, K., Lavina, W. A., Pontrelli, S., Liao, J. C., Putri, S. P., and Fukusaki, E. (2017). Orthogonal partial least squares/projections to latent structures regression-based metabolomics approach for identification of gene targets for improvement of 1-butanol production in *Escherichia coli*. *J. Biosci. Bioeng.* 124, 498–505. doi: 10.1016/j.jbiosc.2017.05.015

Onaka, H., Mori, Y., Igarashi, Y., and Furumai, T. (2011). Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* 77, 400–406. doi: 10.1128/AEM.01337-10

Pathak, C., Jaiswal, Y. K., and Vinayak, M. (2005). Hypomodification of transfer RNA in cancer with respect to queuosine. *RNA Biol.* 2, 143–148. doi: 10.4161/rna.2.4.2417

Senges, C. H. R., Al-Dilaimi, A., Marchbank, D. H., Wibberg, D., Winkler, A., Haltli, B., et al. (2018). The secreted metabolome of *Streptomyces chartreusis* and implications for bacterial chemistry. *Proc. Natl. Acad. Sci. U.S.A.* 115, 2490–2495. doi: 10.1073/pnas.1715713115

Shin, J. H., Singh, A. K., Cheon, D. J., and Roe, J. H. (2011). Activation of the SoxR regulon in *Streptomyces coelicolor* by the extracellular form of the pigmented antibiotic actinorhodin. *J. Bacteriol.* 193, 75–81. doi: 10.1128/JB.00965-10

Sivapragasam, S., and Grove, A. (2019). The link between purine metabolism and production of antibiotics in *Streptomyces*. *Antibiotics* 8:76. doi: 10.3390/antibiotics8020076

Strauch, E., Takano, E., Baylis, H. A., and Bibb, M. J. (1991). The stringent response in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 5, 289–298. doi: 10.1111/j.1365-2958.1991.tb02109.x

Susstrunk, U., Pidoux, J., Taubert, S., Ullmann, A., and Thompson, C. J. (1998). Pleiotropic effects of cAMP on germination, antibiotic biosynthesis and morphological development in *Streptomyces coelicolor*. *Mol. Microbiol.* 30, 33–46. doi: 10.1046/j.1365-2958.1998.01033.x

Tanaka, Y., Hosaka, T., and Ochi, K. (2010). Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in *Streptomyces coelicolor* A3(2). *J. Antibiot.* 63, 477–481. doi: 10.1038/ja.2010.53

Wang, J., Liu, H., Huang, D., Jin, L., Wang, C., and Wen, J. (2017). Comparative proteomic and metabolomic analysis of *Streptomyces tsukubaensis* reveals the metabolic mechanism of FK506 overproduction by feeding soybean oil. *Appl. Microbiol. Biotechnol.* 101, 2447–2465. doi: 10.1007/s00253-017-8136-5

Zhang, J., Liang, Q., Xu, Z., Cui, M., Zhang, Q., Abreu, S., et al. (2020). The inhibition of antibiotic production in *Streptomyces coelicolor* over-expressing the TetR regulator SCO3201 is correlated with changes in the lipidome of the strain. *Front. Microbiol.* 11:1399. doi: 10.3389/fmicb.2020.01399

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 02 December 2020
doi: 10.3389/fbioe.2020.603488

High-Throughput Large-Scale Targeted Proteomics Assays for Quantifying Pathway Proteins in *Pseudomonas putida* KT2440

*Yuqian Gao[1,2], Thomas L. Fillmore[2], Nathalie Munoz[1,2], Gayle J. Bentley[1,3], Christopher W. Johnson[1,3], Joonhoon Kim[1,2], Jamie A. Meadows[1,4], Jeremy D. Zucker[1,2], Meagan C. Burnet[1,2], Anna K. Lipton[1,2], Aivett Bilbao[1,2], Daniel J. Orton[2], Young-Mo Kim[1,2], Ronald J. Moore[2], Errol W. Robinson[1,2], Scott E. Baker[1,2], Bobbie-Jo M. Webb-Robertson[1,2], Adam M. Guss[1,5], John M. Gladden[1,4], Gregg T. Beckham[1,3], Jon K. Magnuson[1,2] and Kristin E. Burnum-Johnson[1,2]**

[1] Department of Energy, Agile BioFoundry, Emeryville, CA, United States, [2] Pacific Northwest National Laboratory, Richland, WA, United States, [3] National Renewable Energy Laboratory, Golden, CO, United States, [4] Sandia National Laboratories, Livermore, CA, United States, [5] Oak Ridge National Laboratory, Oak Ridge, TN, United States

OPEN ACCESS

Edited by:
Pablo Ivan Nikel,
Novo Nordisk Foundation Center for Biosustainability (DTU Biosustain), Denmark

Reviewed by:
Lars M. Blank,
RWTH Aachen University, Germany
Dae-Hee Lee,
Korea Research Institute of Bioscience and Biotechnology (KRIBB), South Korea

***Correspondence:**
Kristin E. Burnum-Johnson
Kristin.Burnum-Johnson@pnnl.gov

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 07 September 2020
Accepted: 10 November 2020
Published: 02 December 2020

Citation:
Gao Y, Fillmore TL, Munoz N, Bentley GJ, Johnson CW, Kim J, Meadows JA, Zucker JD, Burnet MC, Lipton AK, Bilbao A, Orton DJ, Kim Y-M, Moore RJ, Robinson EW, Baker SE, Webb-Robertson B-JM, Guss AM, Gladden JM, Beckham GT, Magnuson JK and Burnum-Johnson KE (2020) High-Throughput Large-Scale Targeted Proteomics Assays for Quantifying Pathway Proteins in Pseudomonas putida KT2440. Front. Bioeng. Biotechnol. 8:603488. doi: 10.3389/fbioe.2020.603488

Targeted proteomics is a mass spectrometry-based protein quantification technique with high sensitivity, accuracy, and reproducibility. As a key component in the multi-omics toolbox of systems biology, targeted liquid chromatography-selected reaction monitoring (LC-SRM) measurements are critical for enzyme and pathway identification and design in metabolic engineering. To fulfill the increasing need for analyzing large sample sets with faster turnaround time in systems biology, high-throughput LC-SRM is greatly needed. Even though nanoflow LC-SRM has better sensitivity, it lacks the speed offered by microflow LC-SRM. Recent advancements in mass spectrometry instrumentation significantly enhance the scan speed and sensitivity of LC-SRM, thereby creating opportunities for applying the high speed of microflow LC-SRM without losing peptide multiplexing power or sacrificing sensitivity. Here, we studied the performance of microflow LC-SRM relative to nanoflow LC-SRM by monitoring 339 peptides representing 132 enzymes in *Pseudomonas putida* KT2440 grown on various carbon sources. The results from the two LC-SRM platforms are highly correlated. In addition, the response curve study of 248 peptides demonstrates that microflow LC-SRM has comparable sensitivity for the majority of detected peptides and better mass spectrometry signal and chromatography stability than nanoflow LC-SRM.

Keywords: targeted proteomics, *Pseudomonas putida* KT2440, mass spectrometry, selected reaction monitoring, central carbon metabolism

INTRODUCTION

Liquid chromatography (LC) selected reaction monitoring (SRM, or multiple reaction monitoring – MRM) targeted proteomics is a popular mass spectrometry (MS)-based protein quantification technique (Picotti and Aebersold, 2012; Ebhardt et al., 2015). Highly sensitive and accurate protein quantification is afforded by LC-SRM analysis of enzymatic digests of proteins in the presence of isotope-labeled internal peptide standards. Most targeted proteomics assays are limited to a few dozen proteins per run and the samples

are often acquired using nanoflow LC-SRM in order to achieve high sensitivity (Picotti et al., 2009; Huttenhain et al., 2012). As mass spectrometry techniques advance, the sensitivity and scan speed of mass spectrometers have greatly improved, yielding analyte detection with high signal to noise in short dwell times. LC-SRM workflows can now monitor hundreds of peptides in a single analysis (Lee et al., 2020), which provides researchers opportunities for deeper exploration into biological systems. In the current era of high-throughput biology, there is also an increasing need to systematically capture detailed information about biological systems with high-throughput experiments. Therefore, the long overhead time (i.e., sample loading, column washing, and equilibrating) of nanoflow LC-SRM can no longer meet the demands of high-throughput studies. Nanoflow LC-SRM also lacks robustness due to the difficulty in keeping stable electrospray over a long period of time. Compared to nanoflow LC-SRM, microflow LC-SRM provides higher throughput and better reproducibility, advantages that overshadow its slightly less sensitivity (Bian et al., 2020).

Pseudomonas putida KT2440 (*P. putida*) is a metabolically versatile, Gram-negative soil bacterium with excellent environmental tolerance since it can grow on a wide variety of carbon sources and thrive in diverse environmental conditions (e.g., aquatic and soil). It has considerable potential for a wide range of biotechnological applications (Linger et al., 2014; Loeschcke and Thies, 2015; Nikel et al., 2016; Johnson et al., 2019). It is critical to understand the intrinsic metabolism of *P. putida* before redesigning it to function as an efficient cell factory for desired bioproduct production through synthetic biology-guided engineering.

Here, we performed a systematic comparison of key characteristics of microflow LC-SRM and the conventional nanoflow LC-SRM platforms through a response curve study of 248 *P. putida* peptides in pooled *P. putida* digests, including throughput, sensitivity, reproducibility, and stability. We also applied both platforms to quantify the expression levels of 132 enzymes (i.e., 339 peptides) in *P. putida*, including enzymes from carbohydrate metabolism, amino acid metabolism, and other pathways. The bacteria were grown in 8 different conditions (*p*-coumarate in MOPS medium, glucose in MOPS medium, glucose in M9 medium, gluconate in M9 medium, fructose in M9 medium, glucose plus gluconate in M9 medium, fructose plus glucose in M9 medium, and fructose plus glucose plus gluconate in M9 medium). Together, we demonstrated that microflow LC-SRM is a robust, high-throughput targeted proteomic approach with little or no loss of sensitivity relative to nanoflow LC-SRM, and it works well in quantifying metabolic pathway enzymes and providing deep insights into the metabolism of *P. putida*.

MATERIALS AND METHODS

P. putida KT2440 Cell Cultivation

Pseudomonas putida KT2440 cells were grown in either MOPS minimal media (LaBauve and Wargo, 2012) or modified minimal M9 medium (comprising 6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 18 μM $FeSO_4$), and 30 mM total of the respective carbon source(s). In MOPS minimal media, cell cultures were cultivated in two individual carbon sources, glucose and *p*-coumarate. In modified minimal M9 medium, cell cultures were cultivated in three individual carbon sources, as well as permutations of each carbon source combination. Specifically, cell cultures were grown individually on glucose, gluconate, and fructose, and on mixed carbon sources: glucose plus gluconate, fructose plus glucose, fructose plus glucose plus gluconate. Cell cultures were inoculated to a starting optical density measured at 600 nm (OD_{600}, measured by a Beckman DU640 spectrophotometer) of 0.1 in 50 mL of medium, according to previously reported methods (Bentley et al., 2020). The cultures were then incubated at 30°C in 250 mL baffled flasks, shaking at 225 rpm until an OD600 of 0.7 was reached, reflecting mid-log phase of growth. The cells were pelleted by centrifugation at 4,500 rpm for 5 min, the supernatant was decanted, and the pellets were washed with ice cold phosphate-buffered saline (PBS) and flash frozen in liquid nitrogen and stored at −80°C until analysis.

Protein Extraction and Tryptic Digestion

Proteins were extracted from the cell pellets using a metabolite, protein, lipid extraction (MPLEx) method (Nakayasu et al., 2016; Burnum-Johnson et al., 2017; Kim and Heyman, 2018). Briefly, in solvent resistant tubes (Sorenson), the cell pellets were resuspended in H_2O and a solvent mixture of four volumes of cold 2:1 chloroform:methanol mix was added. Samples were vigorously vortexed for 30 s, placed on ice for 5 min, vortexed again for 30 s, and centrifuged at 15,000 × *g* for 5 min at 4°C. After centrifugation, the denatured protein interphase was washed in 1 mL of cold 100% methanol, vortexed, and centrifuged again at 15,000 × *g* for 5 min at 4°C to pellet the protein. The methanol was removed, and samples were dried in a fume hood.

The protein pellet was resuspended in 100 mM NH_4HCO_3 containing 8 M urea and protein concentration was measured by a bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, Waltham, MA, United States). Disulfide bonds were reduced by adding dithiothreitol (DTT) to a final concentration of 5 mM and incubating at 60°C for 30 min with constant shaking at 850 rpm. Samples were alkylated with a final concentration of 40 mM iodoacetamide for 1 h at 37°C at 850 rpm. The reaction was then diluted 10-fold with 100 mM NH_4HCO_3 followed by the addition of $CaCl_2$ to 1 mM final concentration. Sequencing-grade trypsin (Promega, Madison, WI, United States) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio and incubated for 3 h at 37°C. Digested samples were desalted with 1 mL Discovery C18 SPE columns (Supelco, Bellefonte, PA, United States), using the following protocol: 3 mL of methanol was added for conditioning the column followed by 2 mL of 0.1% TFA in H_2O. The samples were then loaded onto each column followed by 4 mL of 95:5: H_2O:ACN, 0.1% TFA. Samples were eluted with 1 mL 80:20 ACN:H_2O, 0.1% TFA. The samples were concentrated down to ∼100 μL using a Speed Vac and a final BCA was performed to determine the peptide concentration.

Targeted Proteomics Assay Development

Targeted peptides were selected for 132 proteins in major pathways of *P. putida* KT2440, including carbohydrate metabolism, amino acid metabolism, biosynthesis of terpenoids and polyketides, energy metabolism, xenobiotics biodegradation, lipid metabolism, and nucleotide metabolism pathways (**Supplementary Table 1**) derived from RefSeq assembly accession GCF_000007565.2 using a BioCyc pathway/genome database (Caspi et al., 2013; Paley et al., 2017; Karp et al., 2019). Peptide selection was based on the spectral count data from our global proteomics and then evaluated by Prego (Searle et al., 2015) and CONSeQuence (Eyers et al., 2011) scores. All peptides were further blasted against *P. putida* KT2440 proteome using Protein Coverage Summarizer[1] for their uniqueness to target proteins. The crude synthetic heavy isotope-labeled (e.g., $^{13}C/^{15}N$ on C-terminal lysine and arginine unless otherwise noted) peptides were purchased from New England Peptide (Gardner, MA, United States; FlashPure^{TM} Custom Peptide Array Tier 3). All the cysteines of the synthetic heavy peptides were modified by carbamidomethylation (CAM). Upon receiving the crude synthetic heavy peptides, they were mixed together and diluted with 0.1% formic acid in 15% acetonitrile in water to get a nominal concentration of 1 μM for each individual peptide. The heavy peptide stock solution was aliquoted and stored at −80°C until further use.

To evaluate the peptide quality and select the best responsive transitions for each peptide, 500 fmol/μL of heavy peptide mixture was subjected to high-resolution mass spectrometry analysis (e.g., LTQ Velos Orbitrap MS) since the peptide fragmentation patterns from HCD MS/MS on Orbitrap MS is similar to those from CID MS/MS on triple quadrupole MS (Wu et al., 2014). Firstly, the six most intensive fragment ions for each peptide were selected based on their corresponding MS/MS spectra. The collision energies for individual transitions were obtained by using empirical equations from the Skyline software (MacLean et al., 2010). Secondly, we employed LC-SRM to further evaluate all heavy peptides for the LC performance (e.g., the stability of peptide retention time), MS response (e.g., reliable heavy peptides identification), transition interferences, and endogenous peptide detectability by spiking them into water and the samples. In the end, 2–3 transitions per peptide and 1–3 peptides per protein were selected for the final panel of targeted proteomics assay. There were 339 peptides representing 132 proteins monitored in the final assay (**Supplementary Tables 1, 2**).

Heavy Peptide Spike-In and Sample Loading

There are two sets of samples. One is individual *P. putida* samples taken directly from peptide digests, and the other is the pooled *P. putida* samples, where the individual samples were pooled together to make peptide digests with large volume and used exclusively for response curve studies.

[1]https://omics.pnl.gov/software/protein-coverage-summarizer

For individual *P. putida* samples of microflow LC-SRM analysis, crude heavy peptide mixture stock solution was spiked in the 0.50 μg/μL peptide samples at a nominal concentration of 25 fmol/μL for each peptide. For individual *P. putida* sample of nanoflow LC-SRM analysis, crude heavy peptide mixture stock solution was spiked in the 0.125 μg/μL peptide samples at a nominal concentration of 6.25 fmol/μL for each peptide.

In the response curve study, the response curves of 248 peptides representing 111 proteins were evaluated by spiking heavy isotope labeled peptides in pooled *P. putida* samples at concentrations of 0 (blank), 0.002, 0.008, 0.04, 0.24, 1.2, 6, 30, 120, and 600 fmol/μg. Each of the above samples was subject to both microflow LC-SRM and nanoflow LC-SRM with loading of 25 μg for microflow LC-SRM and 0.25 μg for nanoflow LC-SRM. The response curve samples were injected from lowest to highest with triplicated technical replicates performed on each sample and platform combination.

Microflow and Nanoflow LC-SRM Analysis

Microflow LC-SRM analysis utilized a nanoACQUITY H-Class UHPLC® system (Waters Corporation, Milford, MA, United States), while nanoflow LC-SRM analysis utilized a M-Class UHPLC® system (Waters Corporation, Milford, MA, United States). Both are coupled online to a TSQ Altis^{TM} triple quadrupole mass spectrometer (Thermo Fisher Scientific). The microflow LC-SRM's UHPLC® system was equipped with an ACQUITY UHPLC BEH 1.7 μm C18 column (1,000 μm i.d. × 15 cm), while the nanoflow LC-SRM's UHPLC® system was equipped with an ACQUITY UHPLC BEH 1.7 μm C18 column (100 μm i.d. × 10 cm). In both systems, the mobile phases were (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. The sample loading for microflow LC-SRM is 50 μL of sample (i.e., 25 μg of peptides), while that for nanoflow LC-SRM is 2 μL of sample (i.e., 0.25 μg of peptides). The gradient profile for the microflow LC contained a duty cycle of 32.0 min and a gradient length of 18.9 min (detailed as following, 0.0:90:7, 1.1:90:7, 12.0:90:28, 18.0:90:60, 20.0:90:95, 22.0:90:95, 23.0:90:1, 24.0:90:50, 25.0:90:1, 26.0:90:7, 32.0:90:7, in terms of min:flow-rate-μL/min:%B), while the gradient profile for the nanoflow LC contained a duty cycle of 110.0 min and a gradient length of 78.0 min (detailed as following, 0.0:0.4:1, 1.0:0.6:1, 6.0:0.6:1, 7.0:0.4:1, 9.0:0.4:6, 40.0:0.4:13, 70.0:0.4:22, 80.0:0.4:40, 85.0:0.4:95, 90.0:0.4:95, 91.0:0.5:95, 92.0:0.5:95, 93.0:0.5:50, 94.0:0.5:95, 95.0:0.6:1, 98.0:0.4:1, 110.0:0.4:1, in terms of min:flow-rate-μL/min:%B). Both LC columns were operated with a temperature of 45°C. The TSQ Altis^{TM} triple quadrupole mass spectrometer was operated with ion spray voltages of 4000 ± 100 V and a capillary inlet temperature of 325°C in microflow SRM mode, while it was operated with ion spray voltages of 2100 ± 100 V and a capillary inlet temperature of 350°C in nanoflow SRM mode. In both microflow LC-SRM and nanoflow LC-SRM, tube lens voltages were obtained from automatic tuning and calibration without further optimization. Both Q1 and Q3 were set at unit resolution of 0.7 FWHM and Q2 gas pressure was optimized at 1.5 mTorr. The transitions were

scanned with a minimal dwell time of 0.879 msec for microflow SRM and 0.806 msec for nanoflow SRM, respectively.

Data Analysis

All the LC-SRM data were imported into the Skyline software and the peak boundaries were manually inspected to ensure correct peak assignment and peak boundaries. The normalized dot product of the light transition peak areas with the heavy transition peak areas (i.e., DotProduct as denoted in Skyline) was calculated by the Skyline software, and it can be used to check whether the transition peak areas in the two label types are in the same ratio to each other determining their spectral similarity.

For individual samples, the total peak area ratios of endogenous light peptides and their corresponding heavy isotope-labeled internal standards were calculated by the Skyline software. The detectability of a spectra was defined by having DotProduct above 0.86 and maximum intensity of light above 1,300. Peptide-peptide correlation within single proteins were evaluated and there were 16 peptides whose abundance profile across 30 samples were significantly different from other peptides in the same proteins. The final 323 peptides were proceeded with final protein abundance rollup by removing those 16 peptides. Specifically, data were log2 transformed, compared to assure no biases (Webb-Robertson et al., 2011), and normalized by global median centering based on rank-invariant peptides (Callister et al., 2006), where rank invariance was determined by a *p*-value threshold of 0.2. Protein quantification was performed using R-rollup (Polpitiya et al., 2008; Matzke et al., 2013), which scaled the peptides associated with each protein by a reference peptide (the peptide with the least missing data) and then set the median of the scaled peptides as the protein abundance. Pairwise-univariate statistical comparisons were carried out between each of the other carbon sources and glucose in M9 medium using an analysis of variance (ANOVA) with a Dunnet multiple test correction, or between *p*-coumarate and glucose in MOPS medium and between MOPS medium and M9 medium with glucose as carbon source using a standard two-sample *t*-test (Webb-Robertson et al., 2010, 2017).

For response curve study, the response curves of peptides were generated using the heavy-over-light peak area ratios and the heavy peptides concentrations. Similar to the analysis of individual samples, the DotProduct for all the replicates at the LODs and LOQs level need to be above 0.86 while the maximum intensity of heavy above 1,300. The limit of detection (LOD) was determined from the blanks using the average plus three times the standard deviation of the blank signals, while the limit of quantification (LOQ) using the average plus 10 times the standard deviation of the blank signals. Additionally, LOQs also have coefficient variations of less than 20%. The final LOD and LOQ were listed in **Supplementary Table 3**. Peak capacity was calculated using the formula $p = 1 + t_g/w$, where t_g is the length of the length of the LC gradient and w is the peak width in terms of the full width at half maximum (FWHM). FWHMs were exported from Skyline. The gradient lengths are 78.0 min and 18.9 min for nanoflow and microflow LC-SRM, respectively, while the average FWHMs of all the 248 response curve peptides are 0.22 min and 0.12 min for nanoflow and microflow LC-SRM, respectively. The calculated peak capacities are 356 and 159 for nanoflow and microflow LC-SRM, respectively.

RESULTS

The goal of this study was to develop a high-speed platform that can expedite targeted proteomics analysis, thereby increasing the sample analysis throughput for studying biological systems without significantly reducing the sensitivity. Implementation of this high-speed LC-SRM platform can analyze enzymatic digests of *P. putida* protein extracts in the presence of hundreds of isotope-labeled internal peptide standards, enabling rapid and accurate protein quantification and deep exploration of metabolic pathways (**Figure 1**).

Microflow and Nanoflow LC-SRM Platform Comparison in Response Curve Study

The microflow LC-SRM platform utilized a 1 mm i.d. column packed with 1.7 μm C18 particles, with a total run time of 32 min. By comparison, the nanoflow LC-SRM system employed a 100 μm i.d. column packed with the same C18 particles, with a total run time of 110 min, and thus the microflow LC-SRM platform increases the sample analysis throughput by more than 3-fold. The microflow LC-SRM platform can potentially result in analyzing 10,000 more samples than the nanoflow LC-SRM system each year (**Figure 2**). Many software tools (MacLean et al., 2010; Choi et al., 2014; Gibbons et al., 2019) can be used to facilitate the efficient analysis of the large-scale SRM data, and the data analysis time will be well below the instrument run time. The utilization of microflow LC-SRM is often considered to have lower sensitivity but increased robustness relative to the nanoflow LC-SRM system (Gatlin et al., 1998; Bian et al., 2020). In order to evaluate the effectiveness and efficacy of microflow LC-SRM, we performed a thorough comparison between microflow LC-SRM and nanoflow LC-SRM.

The study monitored the responses of 248 heavy isotope labeled synthetic peptides (**Supplementary Table 3**) spiked-in at various concentrations on both microflow LC-SRM and nanoflow LC-SRM platforms. The endogenous peptides in the pooled *P. putida* samples were used here as references in LC-SRM analysis. The effective gradient length of nanoflow LC-SRM is about five times that of microflow LC-SRM (**Supplementary Figure 1**). On average, the chromatographic peak width (FWHM) in microflow LC-SRM is about 2 times that of nanoflow LC-SRM (**Figure 3A**). The resulting peak capacity of the nanoflow LC-SRM is substantially higher than that of the microflow LC-SRM (>2 times). This indicates that the extended LC gradient greatly improves peak capacity, even when the column length of microflow LC-SRM (150 mm) is longer than that of nanoflow LC-SRM (100 mm). The lower peak capacity of microflow LC-SRM resulted in lower separation power compared to that of nanoflow LC-SRM, as seen in their total ion chromatography (**Supplementary Figure 1**). Even though microflow LC-SRM will benefit from the narrower peak width by increasing analyte concentration, the high flow significantly

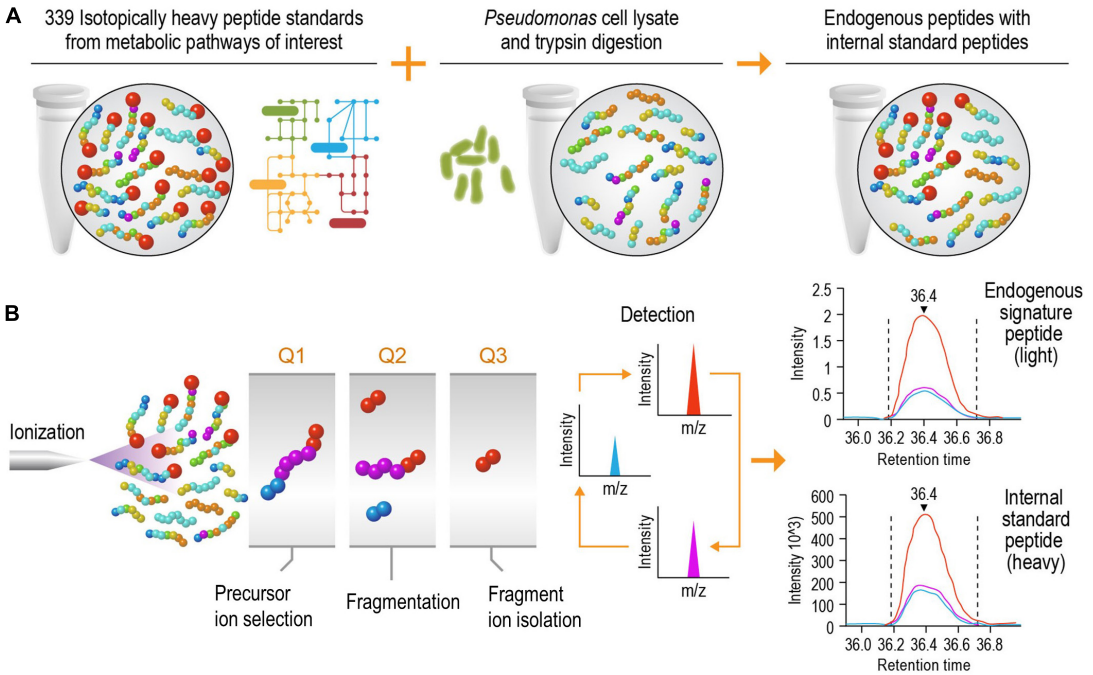


FIGURE 1 | Schematic diagram of the liquid chromatography selected reaction monitoring (LC-SRM) workflow. **(A)** Sample establishment. The heavy isotope labeled internal standard peptides from metabolic pathways of interest are spiked into the tryptic peptides digested from cell lysates, and this results in a mixture of endogenous peptides and internal standard peptides. The mixture will go through liquid chromatography separation and be analyzed by mass spectrometers. **(B)** Selected reaction monitoring using a triple quadrupole mass spectrometer. The eluents of liquid chromatography (LC) separation, carrying both endogenous and internal standard peptides at the same time, are ionized using electrospray ionization. The precursor ions are filtered in Q1, and they are then dissociated via collision into multiple fragment ions in Q2. In Q3, each fragment ion is monitored individually (normally 1 to 5 fragment ions per precursor ion) over the LC elution time. The endogenous peptides share the same retention time and fragmentation profile as their internal standard counterparts. The authors thank PNNL Graphic Designer Nathan Johnson for preparing the figure.

dilutes analyte concentration (90 μL/min versus 0.4 μL/min). To offset this dilution, 100 times more sample mass was loaded on to the microflow LC-SRM column. Sample loading might be a concern for some biological studies with limited sample volumes. By loading 100 times more sample onto the column, similar concentrations of analytes at the time of elution were achieved between nanoflow LC-SRM and microflow LC-SRM. The peak areas of peptides in nanoflow LC-SRM were on average 4 times higher than those in microflow LC-SRM (**Supplementary Figures 2A,C**), mainly due to less interference and better ionization efficiency of nanoflow LC-SRM. However, the stability of the ESI signal in microflow LC-SRM is much higher than that in nanoflow LC-SRM, as demonstrated by the coefficient of variation (CV) of peptide peak areas of three replicated samples (**Supplementary Figures 2B,D**) as well as the CV of peptide peak area ratios of three replicated samples (**Supplementary Figure 2E**). Moreover, the peptide retention time is also more stable in microflow LC-SRM compared to that in nanoflow LC-SRM. As shown in **Figure 3B**, the average standard deviation of peptide retention time is 0.21 min for nanoflow LC-SRM while 0.01 min for microflow LC-SRM. The standard deviation of peptide retention time is an important factor in determining the time window (i.e., start and end times of acquisition) of each peptide for a large-scale multiplexed LC-SRM assay. The smaller the standard deviation of peptide retention time, the narrower the time window. Microflow LC-SRM will be able to use a much narrower time window to fit more peptides in the assay, which will improve the peptide-multiplexing power in a single LC-SRM run.

In order to evaluate the sensitivity of microflow LC-SRM and nanoflow LC-SRM, the LODs and LOQs of 248 peptides were measured. The LODs and LOQs of all the 248 peptides are listed in **Supplementary Table 3**. In general, the distributions of overall LODs and LOQs in microflow LC-SRM are very similar to those in nanoflow LC-SRM (**Figures 3C,D**). Comparing the LODs and LOQs of microflow LC-SRM to those of nanoflow LC-SRM at the individual peptide level (**Figures 3E,F**), the LODs and LOQs of the majority of peptides are the same, while the number of peptides whose LODs are higher in microflow LC-SRM are more than the number of peptides whose LODs are lower in microflow LC-SRM. This indicates that microflow LC-SRM provides equal or slightly lower sensitivity compared to nanoflow LC-SRM. Microflow LC-SRM is likely to lose some sensitivity compared to nanoflow LC-SRM due to its smaller peak area and lower separation power, but the increased stability of microflow LC-SRM overcomes this potential limitation.

In summary, compared to nanoflow LC-SRM, microflow LC-SRM has comparable or slightly lower sensitivity and similar

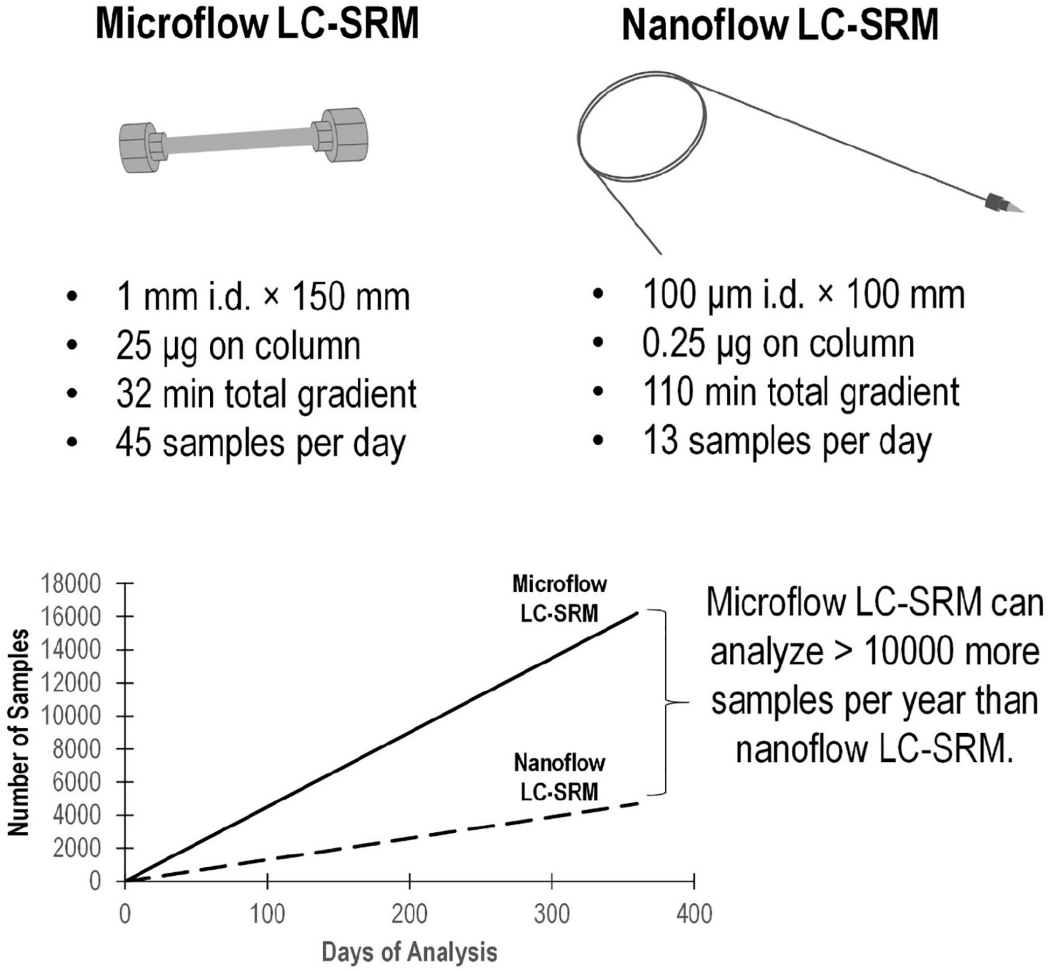


FIGURE 2 | Comparison of the microflow LC-SRM and nanoflow LC-SRM platforms. The top figure includes the size of the column, sample loading, total gradient length, and number of samples per day. The bottom figure shows the number of samples run versus the days spent on the analysis in an ideal situation regardless of the instrument down time. In 1 year, the microflow LC-SRM platform can analyze >10000 more samples than the nanoflow LC-SRM platform.

multiplexing power, but much better sample throughput and stability. The main criteria for applying microflow LC-SRM is whether there is enough biological material (i.e., 25–50 µg of peptide digests) to load onto the larger microflow analytical column.

LC-SRM Analysis of Metabolic Pathway Enzymes in 30 *P. putida* Samples

Both microflow LC-SRM and nanoflow LC-SRM were used to analyze a total of 132 enzymes, including 92 in carbohydrate metabolism, 26 in amino acid metabolism, 4 in nucleotide metabolism, 3 in energy metabolism, 4 in biosynthesis of terpenoids and polyketides, 2 in lipid metabolism, and 1 in xenobiotics biodegradation, as listed in **Supplementary Table 1**. All the 339 peptides corresponding to 132 proteins monitored were detected by both microflow LC-SRM and nanoflow LC-SRM. The detectability (i.e., number of samples where peptides are detected) in the majority of the peptides (i.e., 304 peptides) are the same between microflow LC-SRM and nanoflow LC-SRM, while nanoflow LC-SRM has slightly better detectability in 34 peptides and worse detectability in 1 peptide than microflow LC-SRM (see **Supplementary Table 4**). Overall, the peptide abundance measured by microflow LC-SRM and nanoflow LC-SRM are highly correlated for the same sample (**Supplementary Figure 3**).

In this study, the samples fed with glucose were grown in either M9 minimal salts medium or MOPS minimal medium, while samples fed with *p*-coumarate were grown in MOPS medium and samples fed with gluconate, fructose and mixed carbon sources (glucose + gluconate, fructose + glucose, fructose + glucose + gluconate) were grown in M9 medium. Statistical comparisons between conditions at protein level demonstrated great similarity between microflow LC-SRM and nanoflow LC-SRM in their findings of significantly differentiated proteins

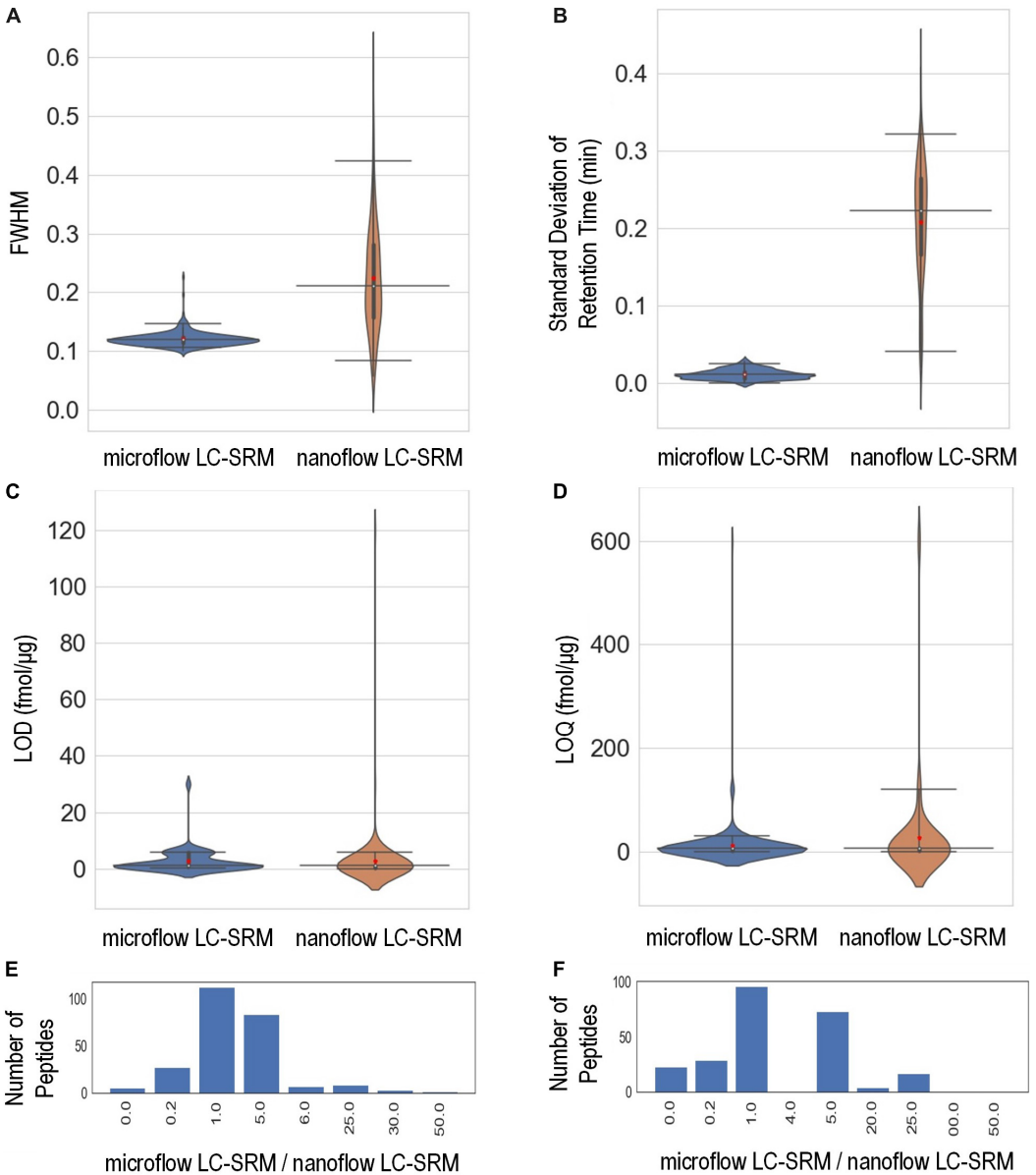


FIGURE 3 | Performance characteristics of the microflow LC-SRM system versus nanoflow LC-SRM system from the response curve study of 248 *P. putida* peptides. The microflow LC-SRM platform is shown in blue and the nanoflow LC-SRM platform in brown for **(A–D)**. **(A)** Violin plot comparing the average full width at half maximum (FWHM) of the three replicated samples at 600 fmol/µg heavy isotope labeled peptide spike-in concentration. **(B)** Violin plot comparing the standard deviation of the retention times of the three replicated samples at 600 fmol/µg heavy isotope labeled peptide spike-in concentration. **(C)** Violin plot comparing the limit of detection (LOD). **(D)** Violin plot comparing the limit of quantification (LOQ). **(E)** Histogram of the LOD differences between microflow LC-SRM platform and nanoflow LC-SRM platform of individual peptides, in terms of ratios of LOD in microflow LC-SRM over LOD in nanoflow LC-SRM. **(F)** Histogram of the LOQ difference between the microflow LC-SRM platform and nanoflow LC-SRM platform of individual peptides, in terms of ratios of LOQ in microflow LC-SRM over LOQ in nanoflow LC-SRM. The three horizontal lines across the violin plots are 2.5, 50, and 97.5% quartiles, respectively, while the red dots in the violin plots are the mean value.

(i.e., p-value < 0.05 and fold change > 2), as shown in **Supplementary Figures 4A–G**.

The stability of peptide retention time across 30 samples in nanoflow LC-SRM is worse than that in microflow LC-SRM (**Supplementary Figure 5**), and these differences were larger when analyzing 30 different samples compared to only analyzing 3 samples of the same matrix composition in the response curve study (**Figure 3B**). The less stable peptide retention time using nanoflow LC-SRM will make the time window scheduling challenging, especially for the analysis of complex sample extracts. Microflow LC-SRM is better suited to facilitate high-throughput time-scheduled SRM transition

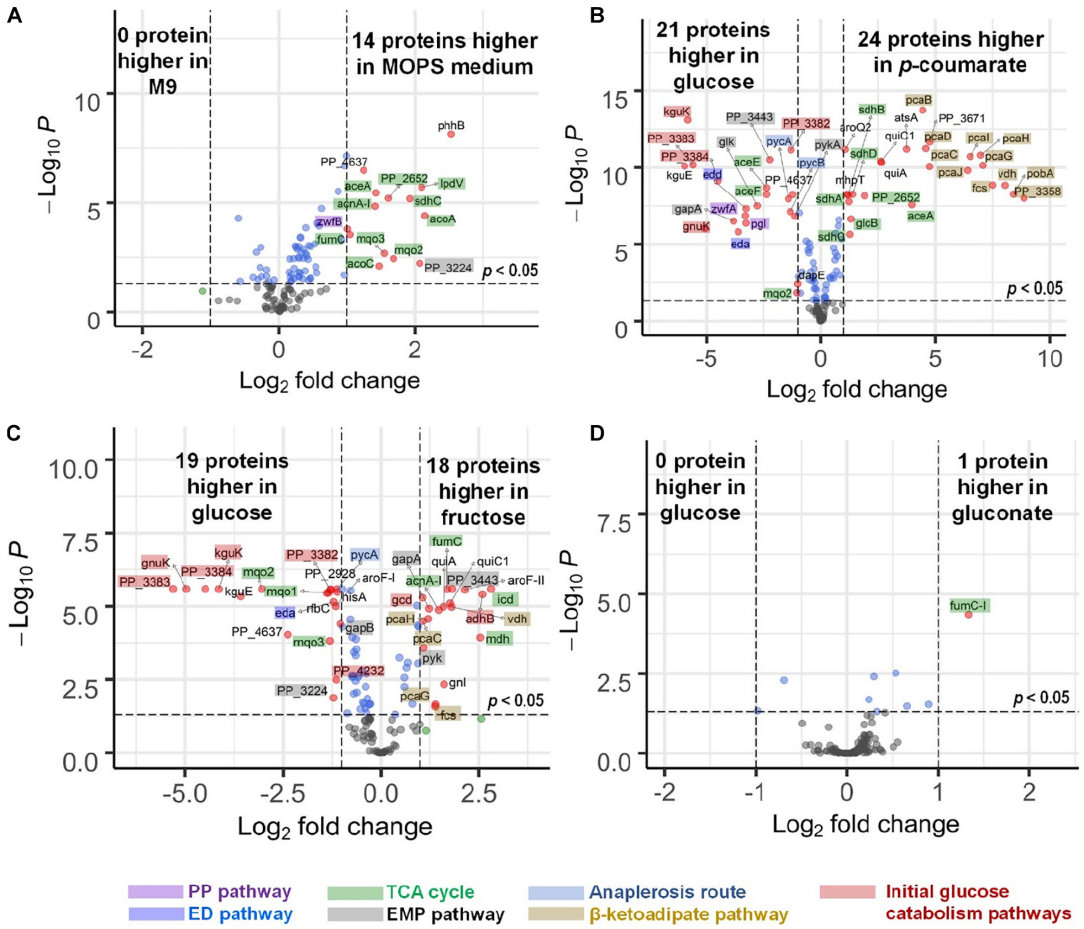


FIGURE 4 | Volcano plots displaying differential expressed genes in four comparisons presented by microflow LC-SRM results of protein expression level of pathway genes in *P. putida*. **(A)** Comparison of strains grown in MOPS medium to those grown in M9 medium, both with glucose as carbon source. **(B)** Comparison of strains grown in *p*-coumarate against those grown in glucose, both in MOPS medium. **(C)** Comparison of strains grown in fructose against those grown in glucose, both in M9 medium. **(D)** Comparison of strains grown in gluconate against those grown in glucose, both in M9 medium. The vertical axis (*y*-axis) corresponds to the significance in terms of -log10 *P* (*p*-value), and the horizontal axis (*x*-axis) displays the log2 fold change value. The red dots represent significantly differentially expressed genes (*p*-value < 0.05, | fold change| > 2) that are either increased (right) or decreased (left); the blue dots represent the genes whose fold change is less than two folds in either direction but with enough significance (*p*-value < 0.05); the green dots represent genes whose fold change is more than two folds in either direction without enough significance (*p*-value > 0.05); the black dots represent genes whose fold change is less than two folds in either direction without enough significance (*p*-value > 0.05). All the significantly differentially expressed genes are label with their gene names. The total of variables plotted contain results of 132 genes. The shades on the gene labels indicate their pathway categories, including the Embden-Meyerhof-Parnas (EMP) pathway of glycolysis, pentose phosphate (PP) pathway, Entner-Doudoroff (ED) pathway, anaplerosis routes, tricarboxylic acid cycle (TCA), the initial glucose catabolism pathways, and β-ketoadipate pathway.

acquisition of large target numbers (hundreds of peptides) across large sample sets.

Enzyme Expression Levels of *P. putida* KT2440 Strains Grown in Different Carbon Sources

Central carbon metabolism (**Supplementary Figure 6**) consists of a series of enzymatic activities to convert carbon sources into valuable metabolic precursors (Noor et al., 2010), and in *P. putida* includes the Embden-Meyerhof-Parnas (EMP) pathway of glycolysis, pentose phosphate (PP) pathway, Entner-Doudoroff (ED) pathway, anaplerosis routes, and tricarboxylic acid cycle (TCA) (Nikel et al., 2015). *P. putida* can grow on a wide variety of carbon sources, from multiple carbohydrates (e.g., glucose, gluconate, fructose) to aromatic carbon (e.g., *p*-coumarate). Glucose and gluconate are transported into the cell either directly or through the conversion process of glucose to gluconate to 2-ketogluconate in the periplasmic space (**Supplementary Figure 6**) (Rojo, 2010). Once inside, glucose, gluconate, and 2-ketogluconate go through the initial glucose catabolism pathways and converge onto the central carbon metabolism (Chavarria et al., 2012). In contrast, fructose is transported by phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) and converted to fructose-1,6-bisphosphate by 1-phosphofructokinase encoded by genes in the fruBKA operon (Chavarria et al., 2016). *P. putida* lacks a classical EMP pathway due to the absence of 6-phosphofructokinase, and utilizes

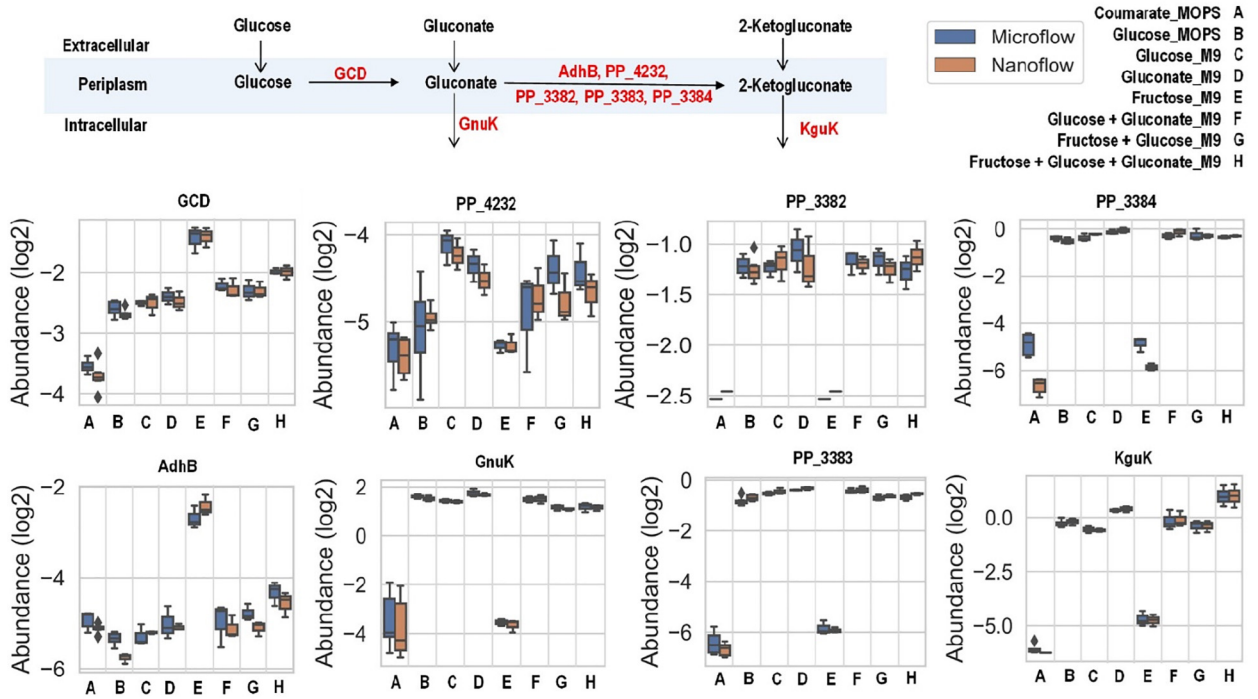


FIGURE 5 | Boxplot of the relative abundance of enzymes in the initial glucose catabolism pathways of up-taking glucose and gluconate with *P. putida* grown on eight different growth conditions: A, *p*-coumarate in MOPS medium; B, glucose in MOPS medium; C, glucose in M9 medium; D, gluconate in M9 medium; E, fructose in M9 medium; F, glucose and gluconate in M9 medium; G, fructose and glucose in M9 medium; H, fructose, glucose and gluconate in M9 medium. The enzymes are quinoprotein glucose dehydrogenase (GCD), alcohol dehydrogenase cytochrome c subunit (AdhB), cytochrome c family protein (PP_4232), gluconate 2-dehydrogenase cytochrome c subunit (PP_3382), gluconate 2-dehydrogenase flavoprotein subunit (PP_3383), gluconate 2-dehydrogenase gamma subunit (PP_3384), gluconokinase (GnuK), and putative 2-ketogluconokinase (KguK). Each box represents the distribution of expression levels of the corresponding enzymes in at least three independent biological replicated samples, including minimum (bottom bar), maximum (top bar), median (line inside the box), first quartile (bottom edge of the box), third quartile (top edge of the box) and diamond (outliers).

these hexose sugars through a cycle formed by enzymes in the ED, EMP, and PP pathways (Nikel et al., 2015). On the other hand, *p*-coumarate is metabolized via the β-ketoadipate pathway before joining the central carbon metabolism via acetyl-CoA and succinate.

Among the 132 enzymes monitored in this study, 83 of them comprise central carbon metabolism, the β-ketoadipate pathway, and the initial glucose catabolism pathways, as shown in **Supplementary Figure 6**. Overall, microflow LC-SRM and nanoflow LC-SRM resulted in similar quantitative patterns across 84 proteins and 30 biological samples and the hierarchy of clustering of genes obtained after performing unsupervised clustering was the same for both platforms (**Supplementary Figure 7**).

There are slight differences of enzyme expression levels between glucose-fed samples grown in MOPS medium versus those in M9 medium, and the major variant enzymes between the two conditions are the ones in the TCA, EMP, and PP pathways (see **Figure 4A**). Comparing strains grown on different carbon sources in the same medium (for example, *p*-coumarate versus glucose in MOPS medium, fructose versus glucose in M9 medium), the enzymes in pathways associated with the intracellular entering route of the carbon sources into *P. putida* have the most significantly (i.e., p-value < 0.05 and fold change > 2) altered expression levels. When *p*-coumarate is the sole carbon source, the majority of the enzymes in the β-ketoadipate pathway are increased, while the majority of the enzymes in the initial glucose catabolism pathways and all the enzymes in the ED pathway are significantly decreased (**Figure 4B**). However, few enzymes in the EMP pathway, PP pathway, and TCA cycle are significantly altered, while the enzymes in the TCA cycle exhibit some differences. Namely, some are decreased (AceE, AceF, and Mqo2) and others are increased (SdhA, SdhB, SdhC, SdhD, AceA, and PP_2652). When fructose is the sole carbon source, the downstream enzymes (i.e., GnuK, PP_4232, PP_3382, PP_3383, PP_3384, and KguK) in the initial glucose catabolism pathways of the glucose-gluconate uptake system are expressed at a very low level, but in contrast, the levels of early pathway enzymes (i.e., GCD and AdhB) in the periplasmic space are expressed at a significantly higher level compared to the presence of other carbon sources (**Figures 4C**, 5). In addition, several enzymes in the β-ketoadipate pathway are significantly increased and half of the enzymes in the ED pathway are significantly decreased (**Figure 4C**). However, only a few enzymes in the EMP pathway and TCA cycle are significantly altered and the rest show minor changes in relative abundance,

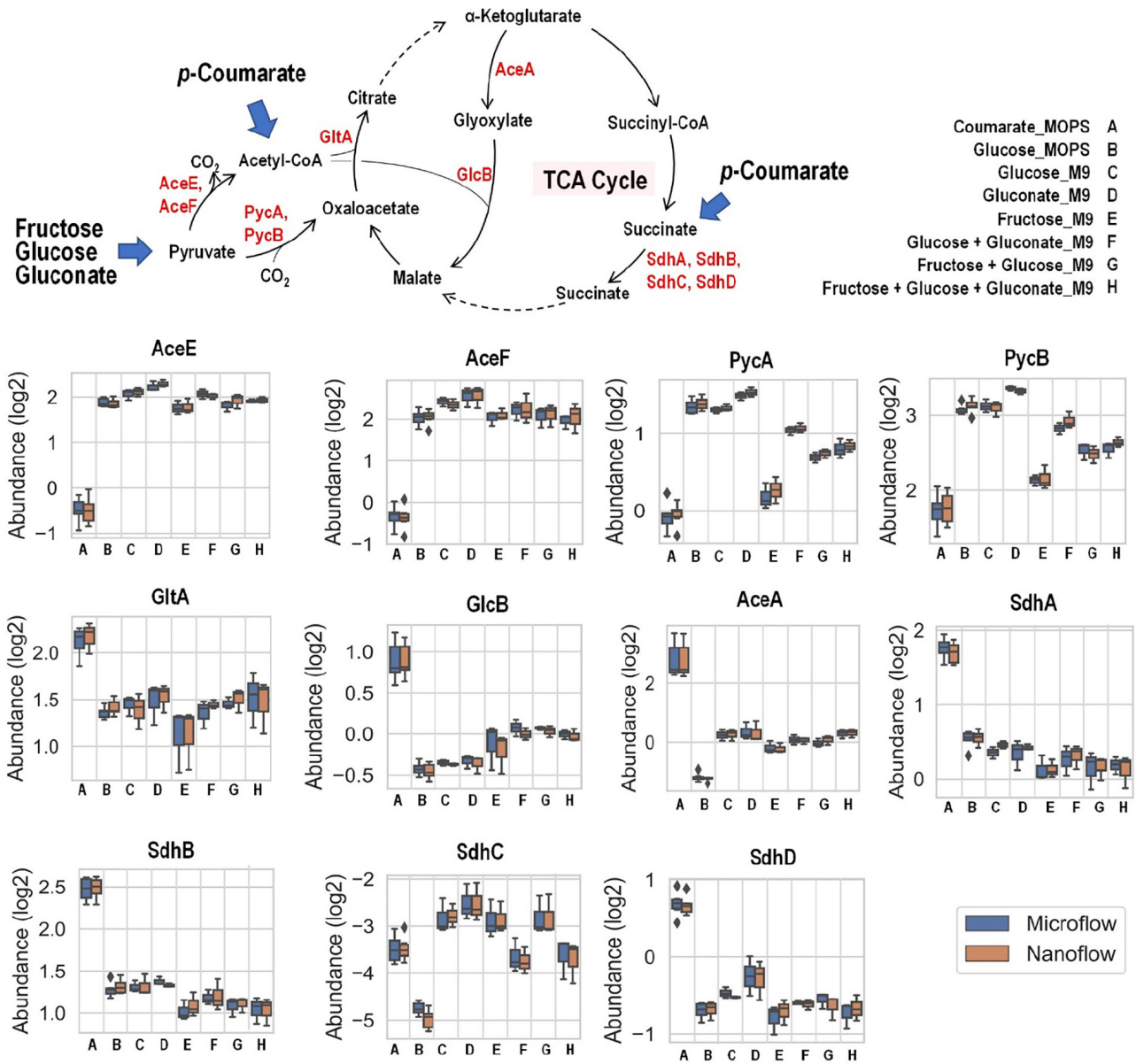


FIGURE 6 | Boxplot of the relative abundance of enzymes in *P. putida* grown on eight different growth conditions: A, *p*-coumarate in MOPS medium; B, glucose in MOPS medium; C, glucose in M9 medium; D, gluconate in M9 medium; E, fructose in M9 medium; F, glucose and gluconate in M9 medium; G, fructose and glucose in M9 medium; H, fructose, glucose and gluconate in M9 medium. These enzymes either facilitate the entrance of key organic carbon products (pyruvate and acetyl-CoA) into the tricarboxylic acid cycle (TCA cycle) or fuel TCA cycle via succinate as substrates. While pyruvate is generated from aliphatic carbon sources (i.e., fructose, glucose and gluconate), acetyl-CoA and succinate are resulted from aromatic carbon source (i.e., *p*-coumarate). The enzymes are pyruvate carboxylase subunit A (PycA), pyruvate carboxylase subunit A (PycB), citrate synthase (GltA), malate synthase G (GlcB), isocitrate lyase (AceA), succinate dehydrogenase flavoprotein subunit (SdhA), succinate dehydrogenase iron-sulfur subunit (SdhB), succinate dehydrogenase membrane b-556 subunit (SdhC), and succinate dehydrogenase hydrophobic membrane anchor subunit (SdhD). Each box represents the distribution of expression levels of the corresponding enzymes in at least three independent biological replicated samples, including minimum (bottom bar), maximum (top bar), median (line inside the box), first quartile (bottom edge of the box), third quartile (top edge of the box) and diamond (outliers).

and similar to what was observed in the comparison between *p*-coumarate and glucose, the enzymes in the TCA can be either decreased or increased depending on carbon sources.

Most of the enzymes are not altered significantly when comparing fructose mixed with either glucose or glucose plus gluconate against glucose (**Supplementary Figures 4D,E**), and gluconate either alone (**Figure 4D**) or mixed with glucose against glucose (**Supplementary Figure 4G**). The expression levels of only a few enzymes are changed significantly. This is likely due to either the convergence of the metabolic pathways

utilized by gluconate and glucose and/or the co-presence of glucose in the system.

The uptake of glucose and gluconate into the cell are incorporated through the initial glucose catabolism pathways. Glucose can be converted to gluconate in the periplasmic space by quinoprotein glucose dehydrogenase (encoded by the *gcd* gene). Once in cytoplasm, glucose will first be phosphorylated by glucokinase (encoded by the *glk* gene) and then converted to 6-phosphogluconate by glucose-6-phosphate 1-dehydrogenase (encoded by the *zwfA*, *zwfB,* and *zwf* genes) followed with 6-phosphogluconolactonase (encoded by the *pgl* gene), while gluconate is phosphorylated directly to 6-phosphogluconate by gluconokinase (encoded by the *gnuK* gene). Interestingly, even grown solely in either glucose or gluconate (both with M9 medium), there are no variations of expression levels for the enzymes converting these carbon sources to 6-phosphogluconate, except slight increase of KguK in samples grown in gluconate (**Figure 5**).

Fructose, glucose, and gluconate metabolism eventually converge to pyruvate and then into the TCA cycle, either directly or through acetyl-CoA as intermediate, while carbon from *p*-coumarate enters the TCA cycle through acetyl-CoA and succinate (**Figure 6**). When *p*-coumarate is the sole carbon source, the levels of enzymes at the entrance point of acetyl-CoA into TCA cycle (GltA and GlcB) and those utilizing succinate into TCA cycle (SdhA, SdhB, and SdhD) are relatively increased. In contrast, the levels of pyruvate carboxylase subunit A and B (PycA and PycB) and pyruvate dehydrogenase E1 and E2 component (AceE and AceF) are decreased when *p*-coumarate is used relative to glucose. This agrees with the fact metabolism of *p*-coumarate generates succinate and acetyl-CoA via β-ketoadipate without pyruvate. In *P. putida* KT2440, benzoate is also known to be degraded to succinate and acetyl-CoA via β-ketoadipate, and its catabolism has been well studied using kinetic modeling (Sudarsan et al., 2016), transcriptomics (Sudarsan et al., 2014), global proteomics and fluxomics (Kukurugya et al., 2019). The transcriptomics study found that these genes involved in pyruvate metabolism and TCA cycle were not differentially expressed at steady state between benzoate and glucose, but the downregulation of succinate dehydrogenase was observed transiently when carbon source was shifted from benzoate to glucose (Sudarsan et al., 2014). More similar observations were made in the proteomics study comparing cells grown on glucose and benzoate to glucose only, including the upregulation of citrate synthase and succinate dehydrogenase as well as the downregulation of pyruvate dehydrogenase (Kukurugya et al., 2019). Interestingly, different observation was made for the expression of genes involved in glyoxylate cycle. In the global proteomics study isocitrate lyase (AceA) was significantly downregulated comparing cells grown on glucose and benzoate to glucose only and malate synthase G (GlcB) was not detected in either cases (Kukurugya et al., 2019), whereas in our targeted proteomics study AceA and GlcB were both increased in *p*-coumarate versus glucose (**Figure 6**).

CONCLUSION

In this study, we systematically compared the performance of two LC-SRM platforms, microflow LC-SRM and nanoflow LC-SRM, through monitoring hundreds of targeted peptides in response curve samples as well as individual samples grown in different environmental conditions. The results of this evaluation clearly demonstrated the promise of microflow LC-SRM as a robust protein quantification system biology tool with high sensitivity, high peptide-multiplexing capability, and high sample throughput. Compared to nanoflow LC-SRM, microflow LC-SRM improves the speed by 3-fold, while providing comparable sensitivity over hundreds of peptides. The results of 132 enzymes in *P. putida* reveals reliable and highly correlated quantification by microflow LC-SRM and nanoflow LC-SRM. In addition, the quantification of enzymes in the central carbon metabolism, the initial glucose catabolism pathways, and β-ketoadipate pathway reveals the changes of these enzyme expression levels of *P. putida* in response to various carbon sources and media composition. The increased throughput and measurement reliability of the presented microflow LC-SRM platform makes it an exceptional test tool for synthetic biology-guided engineering by reducing the cycle time of Design-Build-Test-Learn cycles for enhanced microbial bioproduct production.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and link can be found below: Panorama Public, https://panoramaweb.org/ABF_P_Putida_KT2440_HighFlow_SRM.url.

AUTHOR CONTRIBUTIONS

KB-J and YG planned and designed the targeted proteomic studies. GTB, JG, AG, GJB, CJ, JAM, JKM, and KB-J planned and designed the *P. putida* studies. NM and MB prepared the samples for proteomic analysis. YG and TF performed the targeted proteomic experiments and data analysis. B-JW-R performed the statistical analysis. YG and JK performed the data analyses. YG wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript.

FUNDING

A portion of this research was performed at Pacific Northwest National Laboratory (PNNL) using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. PNNL is a multiprogram national laboratory operated by Battelle for the Department of Energy (DOE) under Contract DE-AC05-76RLO 1830. This work was authored in part by Alliance for Sustainable

Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under contract DE-AC05-00OR22725. Funding was provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Bioenergy Technologies Office (BETO) for the Agile BioFoundry (under Award No. DE-NL0030038). The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

ACKNOWLEDGMENTS

We thank Jay Fitzgerald at DOE and members of the Agile BioFoundry for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.603488/full#supplementary-material

REFERENCES

Bentley, G. J., Narayanan, N., Jha, R. K., Salvachúa, D., Elmore, J. R., Peabody, G. L., et al. (2020). Engineering glucose metabolism for enhanced muconic acid production in *Pseudomonas* putida KT2440. *Metab. Eng.* 59:12.

Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y. C., Meng, C., et al. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat. Commun.* 11:157.

Burnum-Johnson, K. E., Kyle, J. E., Eisfeld, A. J., Casey, C. P., Stratton, K. G., Gonzalez, J. F., et al. (2017). MPLEx: a method for simultaneous pathogen inactivation and extraction of samples for multi-omics profiling. *Analyst* 142, 442–448. doi: 10.1039/c6an02486f

Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W. J., Webb-Robertson, B. J., et al. (2006). Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Prot. Res.* 5, 277–286. doi: 10.1021/pr050300l

Caspi, R., Dreher, K., and Karp, P. D. (2013). The challenge of constructing, classifying, and representing metabolic pathways. *FEMS Microbiol. Lett.* 345, 85–93. doi: 10.1111/1574-6968.12194

Chavarria, M., Goni-Moreno, A., De Lorenzo, V., and Nikel, P. I. (2016). A Metabolic widget adjusts the phosphoenolpyruvate-dependent fructose influx in *Pseudomonas* putida. *mSystems* 1, e00154-16.

Chavarria, M., Kleijn, R. J., Sauer, U., Pfluger-Grau, K., and De Lorenzo, V. (2012). Regulatory tasks of the phosphoenolpyruvate-phosphotransferase system of *Pseudomonas* putida in central carbon metabolism. *mBio* 3, e00028-12.

Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., Maclean, B., et al. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524–2526. doi: 10.1093/bioinformatics/btu305

Ebhardt, H. A., Root, A., Sander, C., and Aebersold, R. (2015). Applications of targeted proteomics in systems biology and translational medicine. *Proteomics* 15, 3193–3208. doi: 10.1002/pmic.201500004

Eyers, C. E., Lawless, C., Wedge, D. C., Lau, K. W., Gaskell, S. J., and Hubbard, S. J. (2011). CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol. Cell Prot.* 10:M110003384.

Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., and Yates, J. R. III (1998). Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* 263, 93–101. doi: 10.1006/abio.1998.2809

Gibbons, B. C., Fillmore, T. L., Gao, Y., Moore, R. J., Liu, T., Nakayasu, E. S., et al. (2019). Rapidly assessing the quality of targeted proteomics experiments through monitoring stable-isotope labeled standards. *J. Prot. Res.* 18, 694–699. doi: 10.1021/acs.jproteome.8b00688

Huttenhain, R., Soste, M., Selevsek, N., Rost, H., Sethi, A., Carapito, C., et al. (2012). Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* 4, 142ra194.

Johnson, C. W., Salvachúa, D., Rorrer, N. A., Black, B. A., Vardon, D. R., John, P. C., et al. (2019). Innovative chemicals and materials from bacterial aromatic catabolic pathways. *Joule* 3:15.

Karp, P. D., Midford, P. E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., et al. (2019). Pathway tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.* bbz104.

Kim, Y. M., and Heyman, H. M. (2018). Mass spectrometry-based metabolomics. *Methods Mol. Biol.* 1775, 107–118.

Kukurugya, M. A., Mendonca, C. M., Solhtalab, M., Wilkes, R. A., Thannhauser, T. W., and Aristilde, L. (2019). Multi-omics analysis unravels a segregated metabolic flux network that tunes co-utilization of sugar and aromatic carbons in *Pseudomonas* putida. *J. Biol. Chem.* 294, 8464–8479. doi: 10.1074/jbc.ra119.007885

LaBauve, A. E., and Wargo, M. J. (2012). Growth and laboratory maintenance of *Pseudomonas aeruginosa*. *Curr. Protoc. Microbiol.* 6:1.

Lee, J. Y., Shi, T., Petyuk, V. A., Schepmoes, A. A., Fillmore, T. L., Wang, Y. T., et al. (2020). Detection of head and neck cancer based on longitudinal changes in serum protein abundance. *Cancer Epidemiol. Biomark. Prev.* 29, 1665–1672. doi: 10.1158/1055-9965.epi-20-0192

Linger, J. G., Vardon, D. R., Guarnieri, M. T., Karp, E. M., Hunsinger, G. B., Franden, M. A., et al. (2014). Lignin valorization through integrated biological funneling and chemical catalysis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12013–12018. doi: 10.1073/pnas.1410657111

Loeschcke, A., and Thies, S. (2015). *Pseudomonas* putida-a versatile host for the production of natural products. *Appl. Microbiol. Biotechnol.* 99, 6197–6214. doi: 10.1007/s00253-015-6745-4

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. doi: 10.1093/bioinformatics/btq054

Matzke, M. M., Brown, J. N., Gritsenko, M. A., Metz, T. O., Pounds, J. G., Rodland, K. D., et al. (2013). A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics* 13, 493–503. doi: 10.1002/pmic.201200269

Nakayasu, E. S., Nicora, C. D., Sims, A. C., Burnum-Johnson, K. E., Kim, Y. M., Kyle, J. E., et al. (2016). MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *mSystems* 1, e00043-16.

Nikel, P. I., Chavarria, M., Danchin, A., and De Lorenzo, V. (2016). From dirt to industrial applications: *Pseudomonas* putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Curr. Opin. Chem. Biol.* 34, 20–29. doi: 10.1016/j.cbpa.2016.05.011

Nikel, P. I., Chavarria, M., Fuhrer, T., Sauer, U., and De Lorenzo, V. (2015). *Pseudomonas* putida KT2440 strain metabolizes glucose through a cycle formed by enzymes of the entner-doudoroff, embden-meyerhof-parnas, and pentose phosphate pathways. *J. Biol. Chem.* 290, 25920–25932. doi: 10.1074/jbc.m115.687749

Noor, E., Eden, E., Milo, R., and Alon, U. (2010). Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* 39, 809–820. doi: 10.1016/j.molcel.2010.08.031

Paley, S., Parker, K., Spaulding, A., Tomb, J. F., O'maille, P., and Karp, P. D. (2017). The Omics Dashboard for interactive exploration of gene-expression data. *Nucl. Acids Res.* 45, 12113–12124. doi: 10.1093/nar/gkx910

Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* 9, 555–566. doi: 10.1038/nmeth.2015

Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of S. *cerevisiae by targeted proteomics*. *Cell* 138, 795–806. doi: 10.1016/j.cell.2009.05.051

Polpitiya, A. D., Qian, W. J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G., et al. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556–1558. doi: 10.1093/bioinformatics/btn217

Rojo, F. (2010). Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol. Rev.* 34, 658–684. doi: 10.1111/j.1574-6976.2010.00218.x

Searle, B. C., Egertson, J. D., Bollinger, J. G., Stergachis, A. B., and Maccoss, M. J. (2015). Using Data Independent Acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Mol. Cell Prot.* 14, 2331–2340. doi: 10.1074/mcp.m115.051300

Sudarsan, S., Blank, L. M., Dietrich, A., Vielhauer, O., Takors, R., Schmid, A., et al. (2016). Dynamics of benzoate metabolism in *Pseudomonas* putida KT2440. *Metab. Eng. Commun.* 3, 97–110. doi: 10.1016/j.meteno.2016.03.005

Sudarsan, S., Dethlefsen, S., Blank, L. M., Siemann-Herzberg, M., and Schmid, A. (2014). The functional structure of central carbon metabolism in *Pseudomonas* putida KT2440. *Appl. Environ. Microbiol.* 80, 5292–5303. doi: 10.1128/aem.01643-14

Webb-Robertson, B. J., Matzke, M. M., Jacobs, J. M., Pounds, J. G., and Waters, K. M. (2011). A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics* 11, 4736–4741. doi: 10.1002/pmic.201100078

Webb-Robertson, B. J., Mccue, L. A., Waters, K. M., Matzke, M. M., Jacobs, J. M., Metz, T. O., et al. (2010). Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. *J. Prot. Res.* 9, 5748–5756. doi: 10.1021/pr1005247

Webb-Robertson, B. M., Bramer, L. M., Jensen, J. L., Kobold, M. A., Stratton, K. G., White, A. M., et al. (2017). P-MartCancer-interactive online software to enable analysis of shotgun cancer proteomic datasets. *Cancer Res.* 77, e47–e50.

Wu, C., Shi, T., Brown, J. N., He, J., Gao, Y., Fillmore, T. L., et al. (2014). Expediting SRM assay development for large-scale targeted proteomics experiments. *J. Prot. Res.* 13, 4479–4487. doi: 10.1021/pr500500d

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer LB declared a past co-authorship with one of the authors GTB to the handling editor.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fbioe.2020.612832

Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in *Rhodosporidium toruloides*

Joonhoon Kim[1,2,3], Samuel T. Coradetti[1,4], Young-Mo Kim[1,3], Yuqian Gao[1,3], Junko Yaegashi[2,3], Jeremy D. Zucker[1,3], Nathalie Munoz[1,3], Erika M. Zink[3], Kristin E. Burnum-Johnson[1,3], Scott E. Baker[1,2,3], Blake A. Simmons[1,2,5], Jeffrey M. Skerker[6], John M. Gladden[1,2,4] and Jon K. Magnuson[1,2,3]**

[1] Department of Energy, Agile BioFoundry, Emeryville, CA, United States, [2] Department of Energy, Joint BioEnergy Institute, Emeryville, CA, United States, [3] Pacific Northwest National Laboratory, Richland, WA, United States, [4] Sandia National Laboratories, Livermore, CA, United States, [5] Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [6] Department of Bioengineering, University of California, Berkeley, Berkeley, CA, United States

OPEN ACCESS

Edited by:
Dong-Yup Lee,
Sungkyunkwan University,
South Korea

Reviewed by:
Christopher Rao,
University of Illinois
at Urbana–Champaign, United States
Julio Augusto Freyre-Gonzalez,
National Autonomous University
of Mexico, Mexico

***Correspondence:**
John M. Gladden
jmgladden@lbl.gov
Jon K. Magnuson
jon.magnuson@pnnl.gov

Specialty section:
This article was submitted to
Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 30 September 2020
Accepted: 04 December 2020
Published: 08 January 2021

Citation:
Kim J, Coradetti ST, Kim Y-M,
Gao Y, Yaegashi J, Zucker JD,
Munoz N, Zink EM,
Burnum-Johnson KE, Baker SE,
Simmons BA, Skerker JM,
Gladden JM and Magnuson JK
(2021) Multi-Omics Driven Metabolic
Network Reconstruction and Analysis
of Lignocellulosic Carbon Utilization
in Rhodosporidium toruloides.
Front. Bioeng. Biotechnol. 8:612832.
doi: 10.3389/fbioe.2020.612832

An oleaginous yeast *Rhodosporidium toruloides* is a promising host for converting lignocellulosic biomass to bioproducts and biofuels. In this work, we performed multi-omics analysis of lignocellulosic carbon utilization in *R. toruloides* and reconstructed the genome-scale metabolic network of *R. toruloides*. High-quality metabolic network models for model organisms and orthologous protein mapping were used to build a draft metabolic network reconstruction. The reconstruction was manually curated to build a metabolic model using functional annotation and multi-omics data including transcriptomics, proteomics, metabolomics, and RB-TDNA sequencing. The multi-omics data and metabolic model were used to investigate *R. toruloides* metabolism including lipid accumulation and lignocellulosic carbon utilization. The developed metabolic model was validated against high-throughput growth phenotyping and gene fitness data, and further refined to resolve the inconsistencies between prediction and data. We believe that this is the most complete and accurate metabolic network model available for *R. toruloides* to date.

Keywords: *Rhodosporidium toruloides*, multi-omics, metabolic networks, genome-scale models, lignocellulosic biomass

INTRODUCTION

An oleaginous yeast *Rhodosporidium toruloides* is a non-model basidiomycete fungus known for its ability to produce carotenoids and accumulate lipids. The high flux in lipid and carotenoid biosynthetic pathways makes *R. toruloides* a promising host organism for producing biofuels and value-added bioproducts from carbon sources derived from lignocellulosic biomass (Wiebe et al., 2012; Fei et al., 2016; Yaegashi et al., 2017; Park et al., 2018; Zhuang et al., 2019). It is also known for the tolerance of inhibitory compounds in lignocellulosic biomass hydrolyzate

as well as the ability to consume aromatic compounds related to lignin degradation products (Yaegashi et al., 2017; Sundstrom et al., 2018). For example, *R. toruloides* can utilize hexoses, pentoses, and aromatic compounds that are found in lignocellulosic biomass hydrolyzate such as glucose, xylose, and *p*-coumaric acid, and produce bisabolene or amorphadiene (Yaegashi et al., 2017). Genome sequence and annotation are available for several *R. toruloides* strains and efficient transformation methods have been developed (Zhu et al., 2012; Zhang et al., 2016; Liu et al., 2017; Coradetti et al., 2018). More advanced genetic tools and parts to engineer *R. toruloides* have been recently developed including RB-TDNAseq, CRISPR/Cas9 editing, RNA interference, and promoter libraries (Coradetti et al., 2018; Liu et al., 2019; Nora et al., 2019; Otoupal et al., 2019).

Previous studies of *R. toruloides* metabolism primarily focused on the lipid production and carotenoid production and several multi-omics studies have been performed to date (Zhu et al., 2012; Lee et al., 2014; Bommareddy et al., 2017; Coradetti et al., 2018). However, it is still not fully clear how different carbon sources present in lignocellulosic biomass hydrolyzate are utilized. There are multiple reports indicating that *R. toruloides*'s metabolism of glucose, xylose, or glycerol is different from *Saccharomyces cerevisiae*'s. For example, *R. toruloides* is an oleaginous yeast and generates cytosolic acetyl-CoA from citrate using ATP-citrate lyase whereas *S. cerevisiae* does not have ATP-citrate lyase and uses the pyruvate dehydrogenase bypass (Rodriguez et al., 2016). When grown on D-xylose, *R. toruloides* transiently accumulates D-arabinitol (or D-arabitol) while *S. cerevisiae* accumulates xylitol (Jagtap and Rao, 2018). The regulation of genes involved in glycerol metabolism including glycerol kinase and glycerol 3–phosphate dehydrogenase was also found to be different between *R. toruloides* and *S. cerevisiae* (Bommareddy et al., 2017). Fatty acids are degraded by peroxisomal β-oxidation in *S. cerevisiae*, but both mitochondrial and peroxisomal β-oxidation pathways are shown to be present and necessary for efficient fatty acid degradation in *R. toruloides*. The growth of *S. cerevisiae* is known to be inhibited by some phenolic compounds that are found in lignocellulosic biomass hydrolyzate including *p*-coumaric acid, ferulic acid, and coniferyl aldehyde, and *S. cerevisiae* can convert them to less inhibitory products, but it is unable to grow on them as sole carbon sources (Adeboye et al., 2015). On the other hand, *R. toruloides* grows on *p*-coumaric acid, ferulic acid, vanillic acid, *p*-hydroxybenzoic acid, and benzoic acid (Yaegashi et al., 2017), but the catabolism of aromatic compounds related to lignin is not well studied in fungi yet. Therefore, there is a need for a metabolic network model to systematically investigate the metabolism of non-model oleaginous basidiomycete yeast, *R. toruloides*. In this work, we reconstruct the genome-scale metabolic network of *R. toruloides* using high-quality published models and perform manual curation using functional annotation and multi-omics data in a fully reproducible manner. Every step of the reconstruction and curation was written in electronic notebooks starting from the reconstruction of a draft metabolic network to the evaluation of the resulting metabolic model. The developed metabolic model and multi-omics data were used to study the utilization of carbon sources that are present in lignocellulosic biomass hydrolyzate in *R. toruloides*.

MATERIALS AND METHODS

Metabolic Network Reconstruction

The *R. toruloides* IFO0880 genome sequence, gene models, and gene annotation from a previous study (Coradetti et al., 2018) was used for the metabolic network reconstruction. The same study identified *R. toruloides* proteins that have orthologous proteins in several different eukaryotic organisms using OrthoMCL (Li et al., 2003). To this list we further added orthologous proteins in *Escherichia coli* K-12 MG1655 and *Pseudomonas putida* KT2440, identified with a separate OrthoMCL analysis including proteins from *R. toruloides* NP11, *Saccharomyces cerevisiae*, *Lipomyces starkeyi*, and *Yarrowia lipolytica*. The list of orthologous proteins was used to gather metabolic reactions from BiGG Models (King et al., 2016), a repository containing high-quality manually curated genome-scale metabolic models. Among the models available in BiGG Models, genome-scale metabolic models of *S. cerevisiae* (Mo et al., 2009), *Chlamydomonas reinhardtii* (Chang et al., 2011), human (Duarte et al., 2007), mouse (Sigurdsson et al., 2010), *E. coli* (Monk et al., 2017), and *P. putida* (Nogales et al., 2008) were used for reconstruction since these models covered most of the orthologous proteins found in *R. toruloides*. In addition, a genome-scale metabolic model of another oleaginous yeast *Y. lipolytica* CLIB122 (Wei et al., 2017) was included. For each protein in *R. toruloides*, metabolic reactions from other metabolic models were added to the *R. toruloides* metabolic network if orthologous proteins were associated with the reactions, and their gene association was updated with the *R. toruloides* protein identifiers. The function and localization of proteins was determined by Joint Genome Institute's annotation on MycoCosm (Grigoriev et al., 2014), WoLF PSORT (Horton et al., 2007) prediction, and the presence of peroxisomal targeting sequences PTS1 and PTS2 predicted by FIMO (Grant et al., 2011) using MEME (Bailey and Elkan, 1994) motifs from known peroxisomal protein sequences.

Metabolic Network Modeling

COBRApy (Ebrahim et al., 2013) was used for curation, evaluation, and modeling of the reconstructed metabolic network. Metabolic models were imported from either the BiGG Models directly or a JSON file constructed from supplemental files from publications. Lieven et al. (2014) was also used to track the development and evaluate the quality of the metabolic model. Escher (King et al., 2015) was used to build metabolic pathway maps and visualize omics data. BOFdat (Lachance et al., 2019) was used to update the biomass composition from experimental data.

Phenotype Microarrays

Phenotype microarray plates and standard components for yeast phenotypic analysis were obtained from Biolog Inc. (Hayward, CA). Wild type *R. toruloides* IFO0880 was precultured to log phase in LB broth at 30°C, 200 RPM in 10 mL culture tubes.

Cells were centrifuged 5 min at 3000 RCF, 22°C, washed twice in sterile water, then resuspended OD 600 of 0.005 in Biolog inoculation fluid IFY-0 with 1 μM nicotinic acid (Sigma, N4126), 1 μM myo-inositol (Sigma, I5125), 1 μM thiamine HCl (Sigma, T1270), 1 μM *p*-aminobenzoic acid (Sigma, A9878), and 1 μM calcium pantothenate (Sigma, 21210) plus Biolog dye mix E (a proprietary, tetrazolium-based dye) and 1 μM menadione sodium bisulfite (Sigma, M5750). For nitrogen, phosphorous, and sulfur sources 100 mM glucose was added to the inoculation fluid. Hundred microliters of the cell suspension was added to each well in plates PM1, PM2, PM3, and PM4. Plates were sealed with clear sealing film (Axygen, CTP-103) and incubated for 120 h at 30°C in the dark. Respiration in each condition was detected by measuring reduction of the dye by comparing absorbance at 590 nm to absorbance at 750 nm.

Fitness Analysis With RB-TDNAseq

Fitness analysis was performed as described in a previous study (Coradetti et al., 2018). Briefly, the three aliquots of the random insertion mutant pool were thawed on ice and recovered in 100 mL YPD (BD Difco, 242820) for two generations (OD 0.2 to OD 0.8). A 10 mL of each starter culture was pelleted and frozen as an initial "time 0" sample. The remaining cells were pelleted 5 min at 4000 RCF, 22°C, washed twice with sterile water and inoculated at OD 0.1 in 50 mL SD media plus 76 mM KH_2PO_4 (Sigma, P9791), 24 mM K_2HPO_4 (Sigma, P3786), and 100 nM $FeSO_4$ (Sigma, 215422) with 1% w/v carbon source. Cultures were grown to OD 600 = 5–10 (20–50 h depending on carbon source) at 30°C, 200 RPM in baffled flasks (25630-250, DWK Life Sciences). Ten milliliters mL samples were pelleted and frozen for DNA extraction. DNA extraction, barcode amplification, and sequencing was performed as described in Coradetti et al. (2018), except that we used primers including dual indexes to prevent "index swapping" on the HiSeq 4000 instrument (Costello et al., 2018) and different lengths of random bases for improved phasing (Price et al., 2019). Fitness analysis was performed with the RBseq software package version 1.0.6, an updated implementation of the algorithms (available at https://github.com/stcoradetti/RBseq). Raw barcode sequencing data are available at the NCBI Sequence Read Archive (PRJNA595384). Fitness scores are available at the fungal fitness browser.[1]

Lipid Content by Fatty Acid Methyl-Ester Analysis

Twenty milligrams of lyophilized cell mass was suspended in 750 μl 3N methanolic HCl (Sigma, 40104-U) and 50 μl chloroform (Sigma, CX1050-1). A total of 100 μl of 10 mg/mL tridecanoic acid methyl ester (Sigma, T0627) in methanol (Sigma, 34860) was added as an internal standard and all the resuspended cell mass was transferred to bead-bug screw-top tubes with glass beads. Tubes were shaken vigorously in a Retsch Tissue Lyser at 30 cycles/second for 5 min to break up cell aggregates and disrupt cell walls, then incubated in an 80°C water bath for 2 h with occasional vortexing. fatty acid methyl-esters (FAMEs) were extracted with 500 μl n-hexane (Sigma, 650552), and diluted 1:10 in hexane. Methyl esters of palmitic, palmitoleic, heptadecanoic, stearic, oleic, linoleic, alpha linoleic, arachidic, behenic, and lignoceric acid were separated in a DB-wax column (Agilent, 123-7012) on a Thermo Scientific Focus gas chromatograph (AS 3000 II) with a flame ionization detector. Standard curves of ratios of peak areas from standards of those FAMEs (Sigma) to tridecanoic acid methyl ester (internal standard) were established. FAME concentrations were determined by comparing ratios of peak areas of FAMEs to the internal standard in the samples.

Media and Growth Conditions for Multi-Omics Experiment

Wild type *R. toruloides* IFO0880 was grown in synthetic defined (SD) medium supplemented with different carbon sources (1% glucose, 1% glucose + 1% D-xylose, 1% D-xylose, 1% L-arabinose, or 1% *p*-coumarate). The SD medium was made with Difco yeast nitrogen base without amino acids (Becton, Dickinson & Co., Sparks, MD) and complete supplemental mixture (Sunrise Science Products, San Diego, CA). Cells were pre-cultured in LB media, and pelleted and washed once with sterile ddH2O before inoculation. Initial pH was adjusted to 7.4 with NaOH, and cells were inoculated to 30 mL of medium with a starting optical density at 600 nm of 0.1. Cultures were grown at 30°C and shaken at 200 rpm. Samples were taken in triplicates at 24 and 48 h in SD glucose, at 24, 66, and 90 h in SD glucose + D-xylose, at 66 and 90 h in SD D-xylose, at 40 and 66 h in SD L-arabinose, and at 40 and 90 h in SD *p*-coumarate. RNA extraction was performed on Promega's Maxwell RSC machine using Plant RNA extraction kit (Promega Corporation, Madison, WI). For proteomics and metabolomics, cells were pelleted and washed twice with 100 mM NH_4HCO_3 at pH 7.8, and spent media was filtered through a 0.45 μ m filter.

RNA Sequencing and Analysis

RNA samples were sequenced and processed at Joint Genome Institute (SRP143805, SRP143806, SRP143807, SRP143808, SRP143809, SRP143810, SRP143811, SRP143812, SRP143813, SRP143814, SRP143815, SRP143816, SRP143817, SRP143818, SRP143819, SRP143820, SRP143821, SRP143822, SRP143823, SRP143824, SRP143825, SRP143826, SRP143827, SRP143828, SRP143829, SRP143830, SRP143831, SRP143832, SRP143833, SRP143834, SRP143835, SRP143836, and SRP143838). Raw read counts were used to perform differential gene expression analysis using DESeq2 (Love et al., 2014).

Metabolite Extraction

Metabolites were extracted from the cell pellets using MPLEx method (Nakayasu et al., 2016; Kim and Heyman, 2018). Briefly, the cell pellets were extracted with a solvent mixture of four volumes of a chloroform and methanol mix (2:1) with a volume of nanopure water. Strong vortexing and ice-cold temperature were also used in the protocol to aid the disruption of the cells. After centrifugation, the aqueous layer

[1] http://fungalfit.genomics.lbl.gov

and half of the organic layer (containing polar and non-polar metabolites, respectively) were combined for GC-MS analysis. The remaining volume of the organic layer was kept for lipidomics analysis. Collected liquid fractions were transferred to new clean vials and subsequently dried in a speed-vacuum concentrator. The denatured protein disk, located between the aqueous and organic layers during the MPLEx protocol, were separately stored for proteomics analysis. All the samples were dried completely and stored in the −80°C freezer until the instrumental analysis.

Metabolomics Analysis

The stored metabolite extracts were completely dried under speed-vacuum to remove moisture and chemically derivatized as previously reported (Kim et al., 2015). Briefly, the extracted metabolites were derivatized by methoxyamination and trimethylsilyation (TMS), then the samples were analyzed by GC-MS. GC-MS raw data files were processed using the Metabolite Detector software, version 2.5 beta (Hiller et al., 2009). Retention indices (RI) of detected metabolites were calculated based on the analysis of a FAMEs mixture, followed by their chromatographic alignment across all analyses after deconvolution. Metabolites were initially identified by matching experimental spectra to a PNNL augmented version of Agilent GC-MS metabolomics Library, containing spectra and validated RI for over 850 metabolites. Then, the unknown peaks were additionally matched with the NIST17/Wiley11 GC-MS library. All metabolite identifications and quantification ions were validated and confirmed to reduce deconvolution errors during automated data-processing and to eliminate false identifications. All metabolomics raw data files are available at OSF data depository https://osf.io/tnqwx/.

Lipidomics

The lipid samples were analyzed using liquid chromatography tandem mass spectrometry (LC-MS/MS) as outlined before (Kyle et al., 2017). Briefly, lipid fractions were re-dried *in vacuo* to remove moisture and reconstituted in 50 μl methanol, 10 μl of which was injected onto a reversed phase Waters CSH column (3.0 mm × 150 mm × 1.7 μm particle size) connected to a Waters Acquity UPLC H class system interfaced with a Velos-ETD Orbitrap mass spectrometer. Lipid molecular species were separated over a 34 min gradient [mobile phase A: acetonitrile/water (40:60) containing 10 mM ammonium acetate; mobile phase B: acetonitrile/isopropanol (10:90) containing 10 mM ammonium acetate] at a flow rate of 250 μl/min. Samples were analyzed in both positive and negative ionization using higher-energy collision dissociation and collision-induced dissociation to obtain high coverage of the lipidome. Confident lipid identifications were made using in-house developed identification software LIQUID (Kyle et al., 2017) where the tandem mass spectrum was examined for diagnostic ion fragments along with associated hydrocarbon chain fragment information. To facilitate quantification of lipids, a reference database for lipids identified from the MS/MS data was created and features from each analysis were then aligned to the reference database based on their identification, m/z and retention time using MZmine 2 (Pluskal et al., 2010). Aligned features were manually curated and peak apex intensity values were generated for subsequent statistical analysis.

Proteomics

The protein disks were dissolved in 100 mM NH_4HCO_3 containing 8 M urea and the protein concentration was measured by BCA assay. Disulfide bonds were reduced by adding dithiothreitol to a final concentration of 5 mM and incubating at 60°C for 30 min. Samples were alkylated with a final concentration of 40 mM iodoacetamide for 1 h at 37°C. The reaction was then diluted 10-fold with 100 mM NH_4HCO_3 followed by the addition of $CaCl_2$ to 1 mM final concentration. Digestion was carried out for 3 h at 37°C with 1:50 (weight:weight) trypsin-to-protein ratio. Salts and reagents were removed by solid-phase extraction using C18 cartridges according to the manufacturer instructions and the resulting peptides were dried in a vacuum centrifuge. The peptides were then resuspended in milliQ water and 500 ng of material was loaded onto in-house packed reversed-phase capillary columns (70-cm × 75 μm i.d.) with 3-μm Jupiter C18. The separation was carried out using a nanoAcquity HPLC system (Waters Corporation) at room temperature. The mobile phase A is 0.1% formic acid in water while mobile phase B is 0.1% formic acid in acetonitrile. The elution was carried out at 300 nL/min with the following gradient: 0–2 min 1% B; 2–20 min 8% B; 20–75 min 12% B; 75–97 min 30% B; 97–100 min 45% B; 100–105 min 95% B; 105–110 min 95% B; 110–140 min 1% B. MS analysis was carried out using a Q Exactive HF (Thermo Fisher Scientific) in data dependent mode. Mass spectrometer settings were as following: full MS (AGC, 1×10^6; resolution, 30,000; m/z range, 350–2000; maximum ion time, 20 ms); MS/MS (AGC, 1×10^5; resolution, 15,000; m/z range, 200–2000; maximum ion time, 200 ms; minimum signal threshold, 2.5×10^4; isolation width, 2 Da; dynamic exclusion time setting, 45 s; collision energy, NCE 30).

All mass spectrometry data were searched using MS-GF+ (Kim and Pevzner, 2014) and MASIC (Monroe et al., 2008) software. MS-GF+ software was used to identify peptides by scoring MS/MS spectra against peptides derived from the whole protein sequence database. MASIC software was used to generate the selected ion chromatographs (SICs) of all the precursors in MSMS datasets and calculate their peak areas as abundance. MASICResultsMerger[2] was used to append the relevant MASIC stats for each peptide hit result in MS-GF+. The MS-GF+ data were then filtered based on 1% false discovery rate (FDR) and less than 5-ppm mass accuracy to generate a list of qualified peptide hit results. The abundances of peptides were determined as the highest peak area identified for the peptide within a sample. RRollup algorithm in InfernoRDN software (Polpitiya et al., 2008) was used to calculate the final protein abundance based on peptide abundance.

[2]https://omics.pnl.gov/software/masic-results-merger

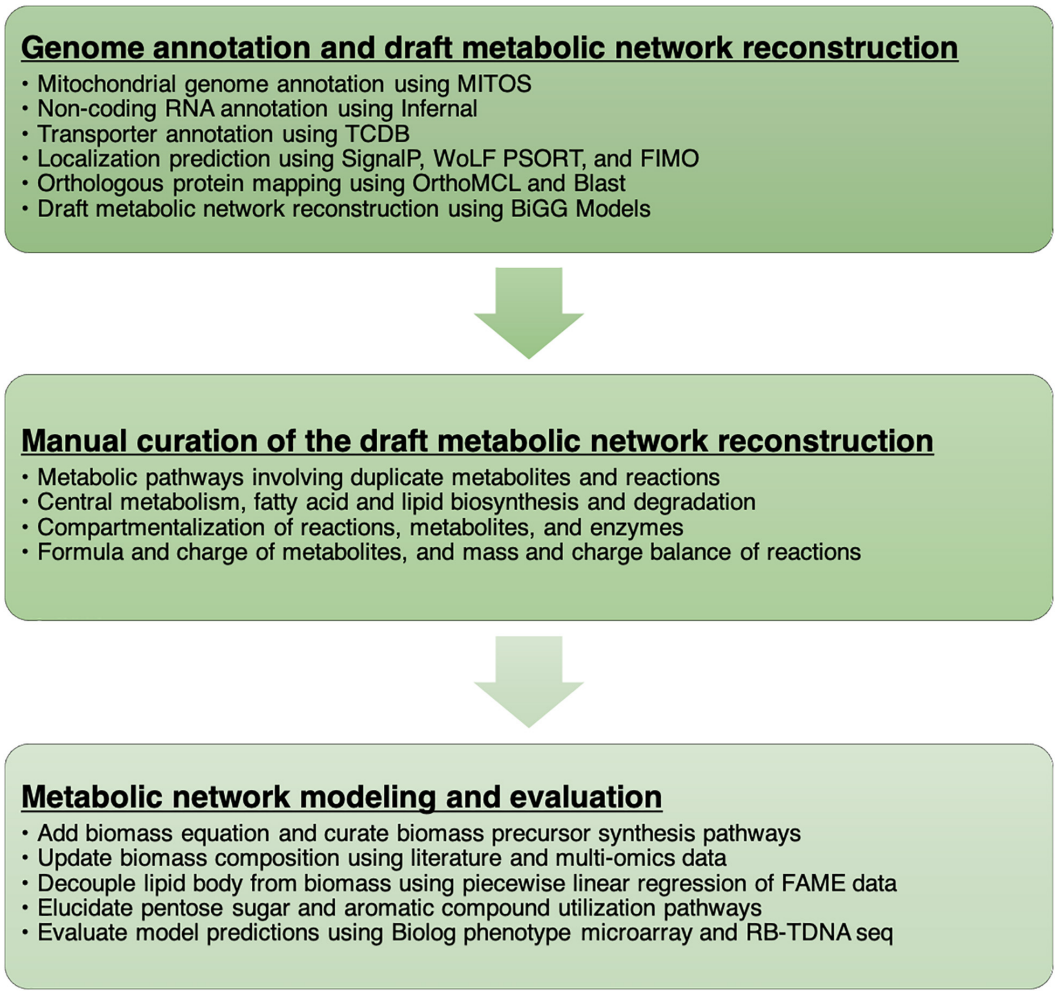


FIGURE 1 | A workflow to develop the metabolic network model of *R. toruloides*.

RESULTS

Metabolic Network Reconstruction Workflow

We documented every step of the metabolic network reconstruction process (**Figure 1**) using Jupyter Notebooks[3] to keep records of changes in metabolic network content and the rationale behind them. We divided the process in multiple notebooks for each stage of reconstruction process (**Supplementary File 1**).

Genome Annotation and Draft Metabolic Network Reconstruction

First, we used orthologous protein groups from OrthoMCL and published metabolic models to build a draft metabolic network reconstruction. Metabolic reactions were taken from published models when any orthologous protein was found in gene-reaction association information. OrthoMCL orthologous groups consist of orthologs and recent paralogs (i.e., gene duplication after speciation and likely to retain similar function) from at least two species (Li et al., 2003). It is generally thought that gene function is more conserved among orthologous genes than between-species paralogs. Recent studies of gene function versus evolutionary history have demonstrated that paralogs can provide more information that previously thought (Stamboulian et al., 2020) but that this added value comes mostly from within species paralogs in taxa with a large corpus of extant biochemical data. Thus, to build our initial metabolic network we transferred functions from orthologous protein groups only. However, during our manual refinement, in rare cases where we had functional data from *R. toruloides* or for filling gaps in pathways for which we have high confidence to exist in *R. toruloides,* we transferred functional prediction from paralogs in other species.

The initial draft reconstruction of *R. toruloides* metabolic network consisted of 1596 proteins, 3804 reactions, and 3589

[3]https://jupyter.org

metabolites. Among the 1596 proteins, 1137 proteins had orthologous *R. toruloides* proteins, but 459 were not yet mapped to *R. toruloides* proteins. Manual investigation of functional annotation and BLAST search were used to determine whether the unmapped proteins were present or absent in *R. toruloides*. For the reactions associated with a protein absent in *R. toruloides*, the absent protein was removed from the gene association if there were isozymes and at least one mapped isozyme was present. A reaction was removed from the metabolic network if the absent protein was a subunit of an enzyme complex and no other isozyme was present. One of the reasons why the number of reactions and metabolites in the initial reconstruction was very large was that protein localization was not yet considered, and same reactions in multiple compartments were present including compartments that are present in other organisms but absent in *R. toruloides*. These reactions were examined and either removed or assigned to appropriate compartments based on the localization prediction and the presence of a signal peptide sequence. It was also found that, although reconstructed from published metabolic models, the initial reconstruction contained many erroneous reactions and metabolites including duplicate reactions or metabolites with different names as well as mass or charge imbalanced reactions.

Manual Curation of the Draft Metabolic Network Reconstruction

Manual curation of the metabolic network reconstruction was performed to first identify and remove duplicate metabolites and reactions. As duplicate or incorrect metabolites and reactions were found, thorough inspection of the metabolic pathway involving them was performed. Functional annotation and localization of each gene in the pathway was confirmed, duplicate or erroneous metabolites and reactions were removed or corrected, and missing metabolites and reactions were either created based on literature and metabolic pathway database or added from other metabolic models in BiGG database. For example, metabolic pathways involving ferricytochrome c were first investigated since duplicate metabolite identifiers exist in BiGG database (ficytc/focytc versus ficytC/focytC). Lactate dehydrogenase and related enzymes involving cytochrome c had many duplicated reactions and incorrect compartment assignments. Cytochrome c oxidase, peroxidase, reductase, NADH:ubiquinone reductase and other mitochondrial electron transport chain reactions were curated. *R. toruloides* is known to have coenzyme Q9 (ubiquinone-9) with nine isoprenyl units (Yamada and Kondo, 1973), and our lipidomics analysis also detected coenzyme Q9. The ubiquinone synthesis reactions from other existing models were modified to include 9 isoprenyl units. Heme biosynthesis reactions were also manually curated since heme O and heme A synthesis reactions in BiGG database had incorrect stoichiometries.

There were many incorrect reactions involved in fatty acid biosynthesis and beta-oxidation, especially for unsaturated fatty acids. Impossible lumped reactions were removed considering the cis and trans configuration of fatty acids, and irrelevant fatty acid reactions (bacteria or plant specific unsaturated fatty acids) were removed. Long-chain fatty acid biosynthesis and fatty acid desaturase reactions were moved from cytosol to endoplasmic reticulum, and short-chain peroxisomal fatty acid beta-oxidation reactions were moved to mitochondria since *R. toruloides* possesses both mitochondrial and peroxisomal beta-oxidation enzymes. Sphingolipid metabolism was extended to include sphingolipid desaturase, methyltransferase, and fungal ceramide biosynthesis and the reactions were placed in endoplasmic reticulum. Localization of reactions involved in phospholipid biosynthesis and remodeling, and triacylglycerol were also updated.

Rhodosporidium toruloides naturally produces and accumulates carotenoids that are derived from mevalonate pathway products. Metabolic reactions and genes in mevalonate pathway and sterol biosynthesis were inspected, and it was found that an ERG27 ortholog is missing in *R. toruloides*. A previous phylogenomics study of sterol synthesis found that a 3-ketosteroid reductase exists in vertebrates and fungi (HSD17B7 in vertebrates and ERG27 in *S. cerevisiae*) but is missing in land plants and other eukaryotic phyla (Desmond and Gribaldo, 2009). More recent studies have suggested that enzymes that oxidizes the C-3 hydroxyl group of sterols to a ketone also reduces the C-3 ketone in tomato (Lee et al., 2019) or sterol-producing bacteria (Lee et al., 2018). On the other hand, in human aldo-keto reductases of the 1C subfamily are involved in 3-ketosteroid reduction and known to be promiscuous (Penning et al., 2015). We found three aldo-keto reductase family proteins in *R. toruloides* (protein ID 13153, 14209, and 14213) that are homologous to human aldo-keto reductases 1C. One of the aldo-keto reductases (14213) had a predicted signal peptide by SignalP and predicted localization in endoplasmic reticulum by WoLF PSORT, and was assigned to reactions catalyzed by the 3-ketosteroid reductase. Carotenoid biosynthesis in endoplasmic reticulum and accumulation in lipid droplets were added based on previous studies (Sun et al., 2017; Ma et al., 2019; Rabeharindranto et al., 2019).

The reactions and genes in the central metabolic pathways were manually checked for their co-factor usage and localization. Reactions in the compartments that are present in other organisms but irrelevant in *R. toruloides* were either moved to appropriate compartments or removed from the model if redundant. Reactions with incomplete gene-to-protein-to-reaction association (e.g., missing subunits) were either removed from the model or updated with corresponding genes if found. Reactions were checked for mass and charge balance, and chemical formula and charge information was updated for all metabolites. When needed, chemical equations were modified based on metabolic databases or new evidence in literature. The curated metabolic reconstruction consisted of 1106 genes, 1934 reactions, and 2010 metabolites (1246 unique metabolites) in nine compartments.

Metabolic Network Modeling and Growth Simulation

A biomass reaction from the *S. cerevisiae* metabolic model and exchange reactions for external metabolites were added

to the metabolic network reconstruction to develop a draft metabolic model that can be used to make growth and flux predictions. In addition, transport reactions for water, carbon dioxide, and oxygen in different compartments were added. We first examined whether each biomass precursor in the biomass reaction can be synthesized in an aerobic glucose minimal medium. A flux balance analysis (FBA) problem maximizing the biomass production was solved, and the shadow prices of metabolites were examined to identify biomass components that could not be synthesized. The cytosolic components in *S. cerevisiae* biomass reaction whose biosynthesis reactions were moved to endoplasmic reticulum were replaced with the respective metabolites in endoplasmic reticulum, and transport reactions for fatty acids, phospholipids, and sterols were added to allow lipid production. In addition, mitochondrial transport reactions for several amino acids and their precursors were added and incorrect reactions in lysine biosynthesis were manually curated to allow synthesis of all components in *S. cerevisiae* biomass reaction. Several reactions generating a free proton gradient across mitochondrial membrane via a loop were identified by mixed-integer programming and removed to prevent unrealistic ATP production.

We used multi-omics and other experimental measurement to update the biomass reaction. The DNA composition was updated using the genome sequence, RNA composition was updated using transcriptomics data, amino acid composition was updated using proteomics data, and lipid composition was updated using fatty acid methyl ester analysis. For lipid composition, we measured fatty acid profiles in multiple media conditions to cover from low lipid to high lipid production states (**Figure 2A**). We observed that the weight percentages of a subset of individual fatty acids linearly increased from a low lipid condition (e.g., YPD) to a high lipid condition (e.g., YNB CN120) whereas the weight percentages of the other fatty acids remained relatively constant. We assumed those fatty acids that were linearly increasing to be the major components in the lipid body, and used segmented linear regression to estimate the fatty acid composition in "lean" cell mass where the majority of lipids are phospholipids, and the fatty acid composition in lipid body where the majority of lipids are triacylglycerols and sterol esters (**Figure 2B**). The fatty acid composition in "lean" cell mass was estimated from the y-intercepts and used in the biomass equation, and the fatty acid composition in lipid body was estimated from the slopes and used in demand reactions for triacylglycerols and sterol esters accumulation in lipid droplet. This allows for the simulation of cell growth and lipid accumulation in lipid body separately, and also enables the simulation of lipid mobilization using sink reactions for triacylglycerol and sterol esters in lipid droplet. Next, we added commonly known trace elements including cofactors and vitamins to the biomass equation and examined using FBA whether they could be synthesized. The reactions involved in folate, thiamine pyrophosphate, quinone, and biotin biosynthesis were manually curated to enable their biosynthesis. It was necessary to add demand reactions for 8-amino-7-oxononanoate and lipoate since we were not able to find all the required enzymes for their synthesis.

We tested the model's capability to predict growth on carbon sources that can be often found in lignocellulosic biomass hydrolyzate – glucose, D-xylose, L-arabinose, and *p*-coumarate. The initial metabolic model was able to predict growth on glucose, but not on D-xylose, L-arabinose, and *p*-coumarate. We examined the existing reactions in the model to identify the missing links within known catabolic pathways. In order to predict growth on D-xylose, the xylose reductase reaction was needed. Two potential xylose reductase encoding genes were found in *R. toruloides* with sequences that are similar to larA and xyrA in *Aspergillus niger*. For growth on L-arabinose, the L-arabinose transporter, L-arabinitol 4-dehydrogenase, and L-xylulose reductase reactions were needed but BLAST of known genes for these reactions resulted in multiple hits with moderate scores. For *p*-coumarate utilization, four reactions in the known *p*-coumarate degradation pathway in bacteria were present in the model. We identified genes that could potentially catalyze the missing reactions in the known *p*-coumarate degradation pathway using functional annotation and BLAST searches. However, additional experimental data was still needed to identify which of these candidates are actually responsible for the missing or added metabolic functions in the D-xylose, L-arabinose, and *p*-coumarate utilization pathways. We therefore performed multi-omics experiments for *R. toruloides* grown in these carbon sources to elucidate the genes and reactions necessary for their utilization.

Multi-Omics Analysis of Lignocellulosic Carbon Utilization in *R. toruloides*

We performed transcriptomics, proteomics, and metabolomics analysis to investigate genes involved in carbon utilization pathways in *R. toruloides*. Cells were grown with glucose, glucose + D-xylose, D-xylose, L-arabinose, or *p*-coumarate as carbon source, and samples were taken during exponential growth phase and stationary phase. An additional sample was taken between exponential and stationary phase for cells grown with glucose + D-xylose to study the co-utilization pattern. Gene expression profiles in cells grown on D-xylose, L-arabinose, or *p*-coumarate was compared to glucose in order to identify genes specifically upregulated or downregulated by each carbon source. RB-TDNA sequencing was also performed in glucose, D-xylose, L-arabinose, *p*-coumarate and other related metabolites. RB-TDNA sequencing uses sequence barcoded random insertions throughout the genome to identify genes required for growth in a given condition. A mixed population of hundreds of thousands of different mutant strains, each bearing an insertion at a different genomic location, is cultured in the condition of interest. The relative abundances of all barcoded strains in the population are simultaneously measured from a single sample by high throughput sequencing. For the more than 6000 genes with three or more independently tracked insertions within their coding sequence, those abundances are aggregated into a single "fitness score" for mutations in each gene in the tested condition. Genes with an essential function in a given condition (e.g., an enzymatic reaction in catabolic pathway) will have negative fitness scores in that condition. Transcriptomics, proteomics, and fitness scores

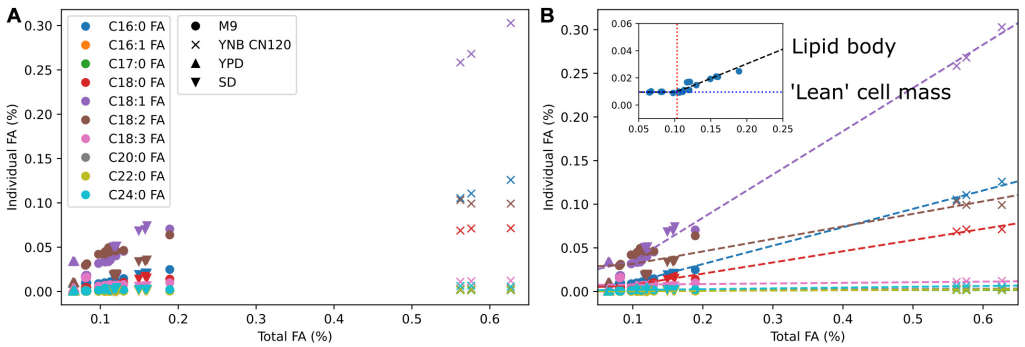


FIGURE 2 | Fatty acid composition of *R. toruloides* in different media conditions and segmented linear regression. **(A)** Fatty acid composition by fatty acid methyl ester analysis in M9, YNB with C to N ratio of 120, YPD, and SD media. Colors indicate different fatty acid species and shapes indicate different media. **(B)** Estimation of fatty acid content in "lean" cell mass and fatty acid content in lipid body using segmented linear regression. Dashed lines are segmented linear regression for each fatty acid k using the equation $y_k = m_k(x–x_0) + b_k$. The inset shows the slope m_k in a black dashed line, the x-intercept x_0 in a red dotted line, and the y-intercept b_k in a blue dotted line for fatty acid k.

were used to assign genes to reactions when annotations were ambiguous or multiple isozymes with substrate promiscuity were present. Metabolomics data was used to identify intermediates in utilization pathways to provide additional support for the proposed pathways.

We first investigated the metabolic pathways and associated genes involved in D-xylose and L-arabinose utilization using the multi-omics data. The functional annotation and multi-omics data suggested that *R. toruloides* uses an alternative pathway involving D-arabinitol and D-ribulose forming ribulose-5-phosphate instead of the known fungal xylose pathway forming D-xylulose-5-phosphate (**Figure 3** and **Table 1**). Two genes in *R. toruloides* NP11 were annotated as D-arabinitol dehydrogenase, RHTO_07702 and RHTO_07844, and used to identify potential arabinitol dehydrogenases in strain IFO0880. Protein ID 9990 was identified as an ortholog of RHTO_07844 by OrthoMCL, and a BLAST search found matches to D-arabinitol 2-dehydrogenases (converting D-arabinitol to D-ribulose) with relatively high identity (over 50%). Consistent with this annotation, Protein ID 9990 had significant fitness defects in many pentose sugars and alcohols including D-xylose, xylitol, D-xylulose, D-arabinitol, L-arabinose, L-lyxose, and L-arabinitol, but not in D-ribulose. Protein ID 9837 was identified as an ortholog of RHTO_07702 by OrthoMCL, and a BLAST search found matches to D-arabinitol dehydrogenase ($NADP^+$), D-arabinitol 2-dehydrogenases, and D-arabinitol 4-dehydrogenase (converting D-xylulose to D-arabinitol) with lower identity (less than 40%). BLAST analysis of *Aspergillus niger* D-arabinitol 4-dehydrogenase (An04g09410) against the *R. toruloides* genome found several hits including protein ID 9837 suggesting its role as D-arabinitol 4-dehydrogenase. However, protein ID 9837 had a weaker fitness defect suggesting that other enzymes participate in the conversion of D-xylulose to D-arabinitol. Among other BLAST hits, protein ID 8905 was upregulated in D-xylose and L-arabinose and had some fitness defect during growth on pentose sugars and alcohols. Therefore, the weak fitness defects for either protein ID 9837 and protein ID 8905 are consistent with genetic redundancy at this step in the xylose utilization pathway. The proposed alternative pathway is supported by our observation that the D-xylulose kinase (protein ID 16850) had very low RNA abundance and no detectable peptides in every condition we tested, and that mutants for protein ID 16850 had no significant fitness defect in any condition tested. Another supporting observation is that the D-ribulose kinase (protein ID 14368) had significant fitness defects in all pentose sugar and alcohol media conditions tested. This pathway is also consistent with recent observations that *R. toruloides* grown on D-xylose transiently accumulates D-arabinitol in the culture media (Jagtap and Rao, 2018). In summary, our omics and genetic data supports an alternative D-xylose and L-arabinose utilization pathway involving a D-ribulose-5-phosphate intermediate rather than a D-xylulose-5-phosphate intermediate (**Figure 3**).

Next, we propose that *R. toruloides* metabolizes *p*-coumarate to protocatechuate by a beta-oxidation-like pathway in the peroxisome (**Figure 4** and **Table 2**). Previously known *p*-coumarate utilization pathway in bacteria such as *P. putida* contains *p*-coumaroyl-CoA hydratase/aldolase or feruloyl-CoA hydratase/lyase that hydrolyzes *p*-coumaroyl-CoA to 3S-(4-hydroxyphenyl)-3-hydroxy-propanoyl-CoA and subsequently produces 4-hydroxybenzoyl-CoA and acetyl-CoA. We found that, in *R. toruloides* grown in *p*-coumarate media, enzymes that are similar to peroxisomal fatty acid beta-oxidation enzymes ACSL (long-chain acyl-CoA synthetase), FOX2 (multifunctional enzyme, 3-hydroxyacyl-CoA dehydrogenase, and enoyl-CoA dehydratase), and POT1 (3-ketoacyl-CoA thiolase) were upregulated. Our fitness data from RB-TDNAseq on *p*-coumarate and previously published RB-TDNAseq data on oleic and ricinoleic acid (Coradetti et al., 2018) shows that these enzymes are distinct from the mitochondrial or peroxisomal fatty acid beta-oxidation enzymes (**Figure 5** and **Table 3**) since they did not have a fitness defect in oleic or ricinoleic acid media. Mitochondrial or peroxisomal fatty acid beta-oxidation genes did not have significant fitness defect in *p*-coumarate or ferulate media (**Figure 5** and **Table 3**). The predicted localization of the enzymes involved in *p*-coumarate indicates

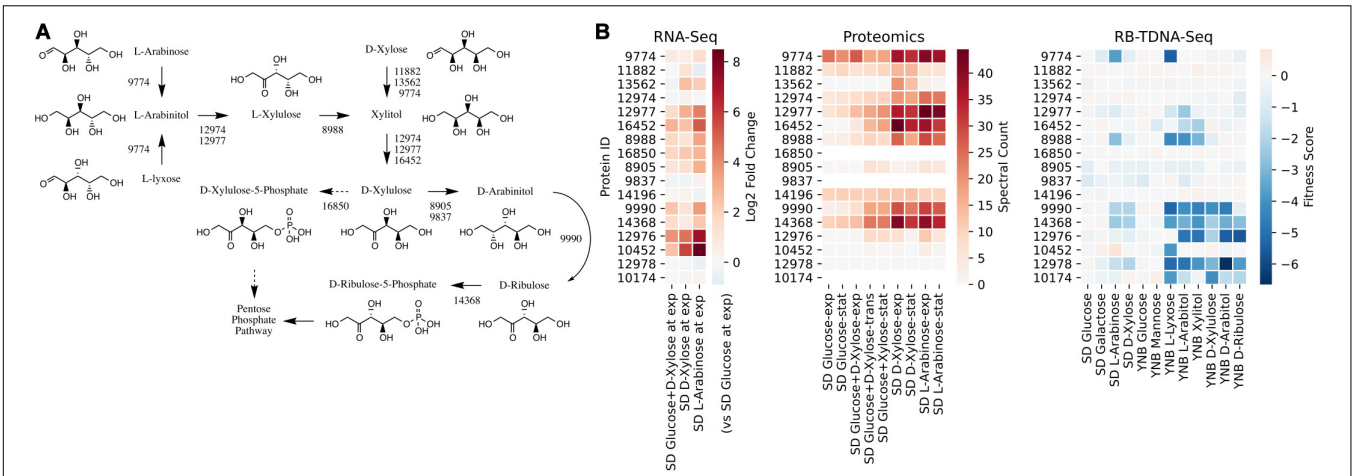


FIGURE 3 | Pentose utilization pathway in *R. toruloides*. **(A)** Pentose sugars and alcohols are converted to D-ribulose-5-phosphate via D-arabinitol dehydrogenases before entering the pentose phosphate pathway. **(B)** Gene expression, protein expression, and fitness scores for pentose utilization pathway genes (exp, exponential phase; trans, transition phase; stat, stationary phase).

TABLE 1 | Genes involved in pentose sugar and alcohol utilization.

Protein ID	Annotation	*S. cerevisiae* best hit	Human best hit
9774	Alcohol dehydrogenase (NADP+)	YPR1	AKR1A
11882	Glycerol 2-dehydrogenase (NADP+)	YPR1	AKR1A
13562	Alcohol dehydrogenase (NADP+)	ADH7	
12974	Zinc-binding alcohol dehydrogenases	SOR1	SORD
12977	L-iditol 2-dehydrogenase	XYL2	SORD
16452	D-xylulose reductase	SOR1	SORD
8988	Sorbose reductase		DHRS4
16850	Xylulokinase	XKS1	XYLB
8905	Reductases with broad range of substrate specificities	IRC24	DHRS4
9837	D-arabinitol dehydrogenase		
9990	D-arabinitol 2-dehydrogenase	SPS19	CBR4
14368	Ribulose kinase and related carbohydrate kinases	YDR109C	FGGY
12976	Predicted transporter (major facilitator superfamily)	STL1	SLC2A
10452	Predicted transporter (major facilitator superfamily)	RGT2	SLC2A
12978	Fungal specific transcription factor Zn(2)-Cys(6) binuclear cluster domain		
10174	Related to C2H2 zinc finger protein	SDD4	

that *p*-coumarate is first degraded to protocatechuate in the peroxisome, and protocatechuate is transported to the cytosol for further degradation via the 3-oxoadipate pathway. Degradation of *p*-coumarate and ferulate via a beta-oxidation like pathway would result in 4-hydroxybenzoate and vanillate, respectively. Protocatechuate and 4-hydroxybenzoate were detected in the intracellular and extracellular metabolomics of cells grown on *p*-coumarate (**Figure 4C**). An enzyme similar to kynurenine 3-monooxygenase BNA4 showed fitness defect in *p*-coumarate, but not in ferulate, which indicates it is likely to be a 3-hydroxybenzoate 4-monooxygenase producing protocatechuate from 4-hydroxybenzoate. 4-hydroxybenzoate may also be transported to mitochondria for quinone biosynthesis. A fitness defect in mitochondrial oxoadipate carrier ODC2 suggests that 3-oxoadipate is transported from cytosol to mitochondria for the final steps in the beta-ketoadipate pathway generating succinyl-CoA and acetyl-CoA which can feed into the TCA cycle. Interestingly, several genes involved in aromatic amino acid metabolism showed a significant fitness defect in *p*-coumarate, although they showed some degree of fitness defect in other media conditions. For example, four genes in the tryptophan degradation pathway to 2-amino-3-carboxymuconate semialdehyde via kynurenine (BNA1, BNA2, BNA4, and BNA5) as well as cytosolic aspartate aminotransferase AAT2 had pronounced fitness defects in *p*-coumarate and ferulate (**Figure 5** and **Table 3**). Since *p*-coumarate and 4-hydroxybenzoate are known to be ubiquinone precursors and they are synthesized from aromatic amino acids, it is possible that high concentration of these compounds affects the regulation of aromatic amino acid pathway genes and fitness defect becomes more pronounced. Taken together, the omics and genetic data support a proposed *p*-coumarate utilization pathway that involves formation of protocatechuate in the peroxisome, followed by ortho-cleavage in the cytosol, and then 3-oxoadipate degradation in the mitochondria (**Figure 4**). Multi-omics analysis and manual curation improved the metabolic model, but their scope was still limited to lignocellulosic carbon utilization pathways. In the next section, we performed a genome-scale evaluation and iteratively improved the model using high-throughput growth phenotyping and functional genomics.

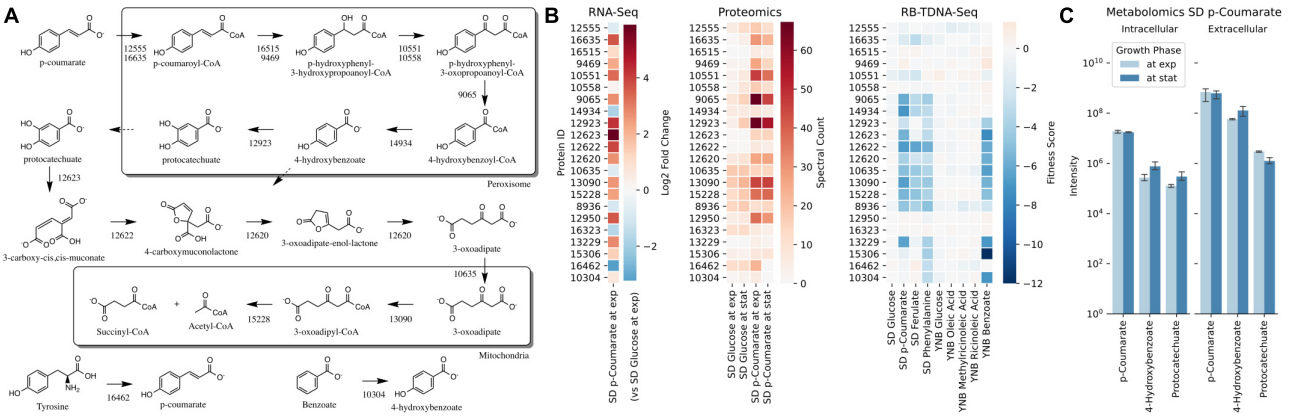


FIGURE 4 | *p*-Coumarate utilization pathway in *R. toruloides*. **(A)** *p*-Coumarate degradation to protocatechuate by a beta-oxidation like pathway in peroxisome, protocatechuate degradation to 3-oxoadipate by the ortho-cleavage pathway in cytosol, and 3-oxoadipate degradation in mitochondria. **(B)** Gene expression, protein expression, and fitness scores for *p*-coumarate utilization pathway genes (exp, exponential phase; stat, stationary phase). **(C)** Intracellular and extracellular measurement of *p*-coumarate and intermediates in *p*-coumarate condition (not detected in glucose, glucose + D-xylose, D-xylose, and L-arabinose conditions).

TABLE 2 | Genes involved in *p*-coumarate utilization.

Protein ID	Annotation	*S. cerevisiae* best hit	*P. putida* best hit	PTS2 (N-terminal)	PTS1 (C-terminal)	Pathway
12555	Long-chain acyl-CoA synthetase	FAA2			AKL*	Peroxisomal *p*-coumarate degradation to protocatechuate
16635	Long-chain acyl-CoA synthetase	PCS60			AKL*	
16515	Enoyl-CoA hydratase/isomerase family				ARL*	
9469	Peroxisomal dehydratase	FOX2			SKL*	
10551	3-oxoacyl-(acyl-carrier protein) reductase	FOX2		6-RLQQVQGQL-14		
10558	3-oxoacyl-(acyl-carrier protein) reductase	FOX2		7-RLSAVSGQL-15		
9065	3-oxoacyl CoA thiolase	POT1				
14934	Alpha/beta hydrolase family				ARL*	
12923	Monooxygenase involved in coenzyme Q (ubiquinone) biosynthesis		pobA		ASL*	
12623	Dioxygenase		pcaH			Protocatechuate degradation via 3-oxoadipate
12622	Lactonase		pcaB			
12620	Carboxymuconolactone decarboxylase family		pcaCD			
13090	3-oxoacid CoA-transferase		pcaIJ			
15228	Acetyl-CoA acyltransferase 1		pcaF			
10635	Mitochondrial 2-oxodicarboxylate transporter	ODC2				
8936	Aspartate aminotransferase, cytoplasmic	AAT2				
12950	Vanillin dehydrogenase	UGA2	ALDH9			
16323	Aldehyde dehydrogenase (NAD+)	HFD1	ALDH3			
13229	Unknown transmembrane protein					
15306	long-chain acyl-CoA synthetase					
16462	Phenylalanine/tyrosine ammonia-lyase					
10304	Cytochrome P450, family 3, subfamily A					

PTS, peroxisomal targeting sequence.

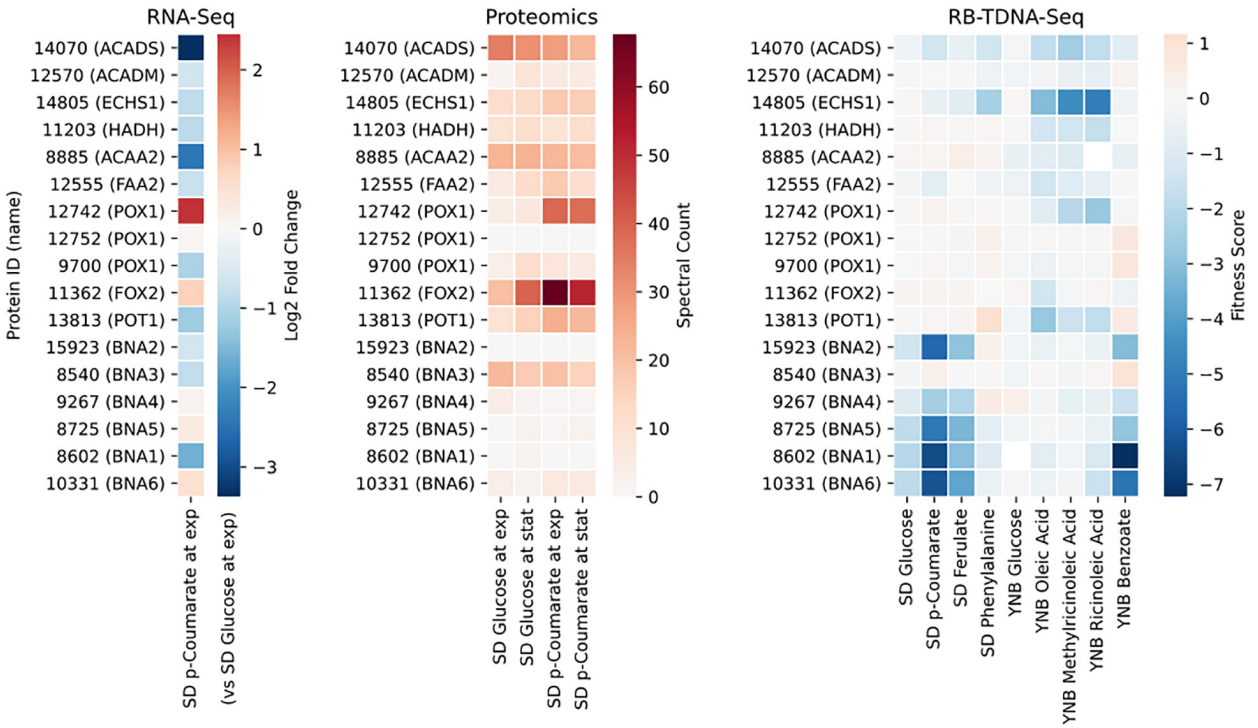


FIGURE 5 | Gene expression, protein expression, and fitness scores for fatty acid beta-oxidation and NAD biosynthesis pathway genes (exp, exponential phase; stat, stationary phase).

TABLE 3 | Genes involved in fatty acid beta-oxidation and NAD biosynthesis.

Protein ID	Annotation	*S. cerevisiae* best hit	Human best hit	Pathway
14070	Short/branched chain acyl-CoA dehydrogenase		ACADS	Fatty acid beta-oxidation
12570	Acyl-CoA dehydrogenase		ACADM	
14805	Enoyl-CoA hydratase	EHD3	ECHS1	
11203	3-hydroxyacyl-CoA dehydrogenase		HADH	
8885	Acetyl-CoA acyltransferase 2	ERG10	ACAA2	
12555	Long-chain acyl-CoA synthetase	FAA2	ACSL1	
12742	Acyl-CoA oxidase	POX1	ACOX1	
12752	Acyl-CoA oxidase	POX1	ACOX1	
9700	Acyl-CoA oxidase	POX1	ACOX1	
11362	Multifunctional beta-oxidation protein	FOX2	HSD17	
13813	Acetyl-CoA acyltransferase 1	POT1	ACAA1	
15923	Indoleamine 2,3-dioxygenase	BNA2	IDO1	NAD biosynthesis
8540	Kynurenine aminotransferase	BNA3	KYAT3	
9267	Kynurenine 3-monooxygenase	BNA4	KMO	
8725	Kynureninase	BNA5	KYNU	
8602	3-hydroxyanthranilate 3,4-dioxygenase	BNA1	HAAO	
10331	Nicotinate-nucleotide pyrophosphorylase (carboxylating)	BNA6	QPRT	

Validation and Reconciliation of Growth Phenotype and Gene Essentiality Predictions

We tested the developed metabolic model's capability to predict growth on different carbon, nitrogen, sulfur, and phosphate sources. Growth phenotype data from Biolog Phenotype MicroArrays were used to evaluate the model predictions (**Figure 6**). Among 384 conditions in Biolog plates (PM1, PM2, PM3B, and PM4A), 116 conditions could be simulated with the model since not all the metabolites were present in the metabolic network. Of these 116 conditions, the model correctly predicted 76 positive and 12 negative growth phenotypes, and incorrectly predicted 6 false positive and 22 false negative growth

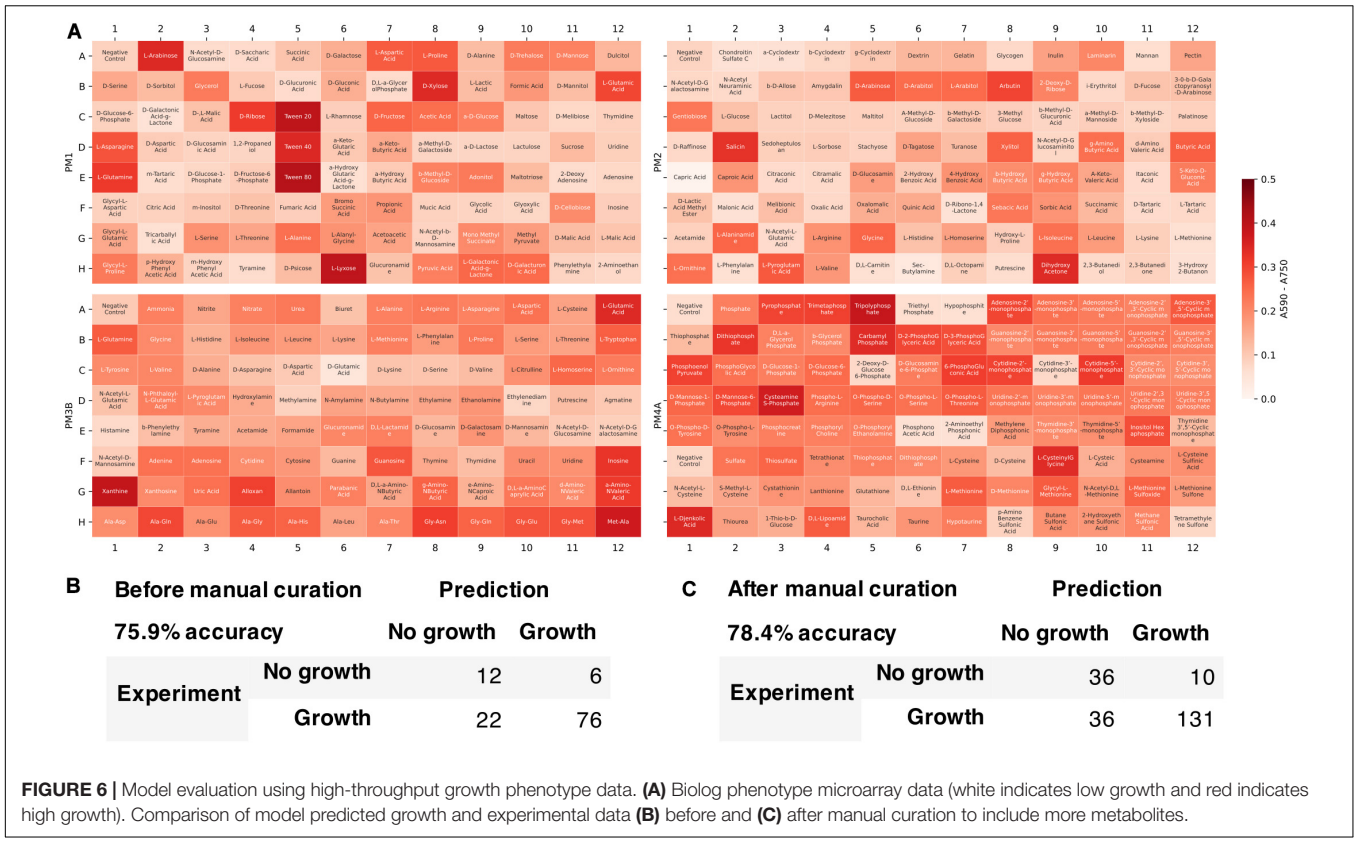


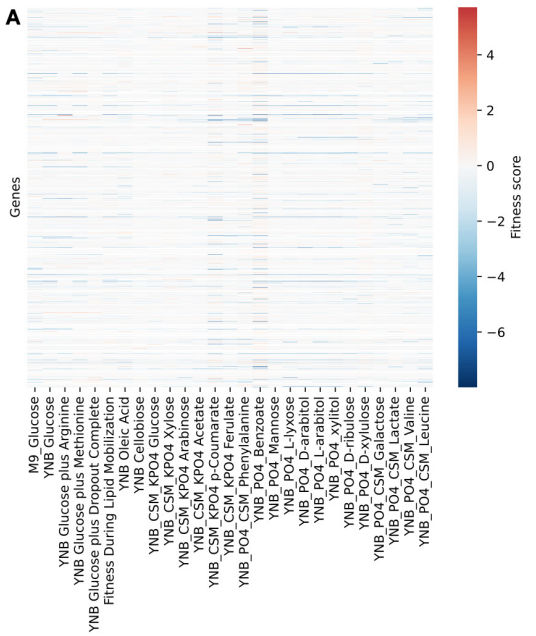
FIGURE 6 | Model evaluation using high-throughput growth phenotype data. **(A)** Biolog phenotype microarray data (white indicates low growth and red indicates high growth). Comparison of model predicted growth and experimental data **(B)** before and **(C)** after manual curation to include more metabolites.

phenotypes. The overall accuracy was 75.9%, comparable to previously published metabolic models of other organisms [e.g., *E. coli iAF1260* model (Feist et al., 2007) with 75.9% accuracy in 170 conditions]. We then manually refined the model to include more metabolites found in Biolog plates and reconcile the inconsistencies. The updated model was able to simulate 213 conditions in Biolog plates with 78.4% accuracy and Matthew's correlation coefficient of 0.493.

We used the fitness scores from RB-TDNAseq to evaluate the model's capability to predict conditionally essential genes in different growth conditions (**Figure 7**). Genes were considered essential if they were classified as essential in our previous RB-TDNAseq study (Coradetti et al., 2018) or fitness score was less than a cut-off value. There were 1147 genes in the model, but 15 genes were mitochondrial and excluded from this analysis since their essentiality was not available from the RB-TDNAseq data. The model predicted gene essentiality for 1132 genes in 27 different growth conditions with 72.7% accuracy and Matthew's correlation coefficient of 0.388. The model was further refined to resolve the inconsistencies and several genes with erroneous ortholog mapping were removed from the model. The refined model had 1142 genes, 2398 reactions, and 2051 metabolites (1205 unique metabolites), and predicted gene essentiality for 1127 non-mitochondrial genes in 27 conditions with 78.6% accuracy and Matthew's correlation coefficient of 0.406 [see **Supplementary File 5** for a comparison with a previously published model (Dinh et al., 2019)]. Among these 1127 genes, 281 genes were essential across all conditions, 772 genes were not essential under any conditions, and 74 genes were essential under only certain conditions. For these 74 conditionally essential genes, the refined model predicted gene essentiality with 78.7% accuracy.

DISCUSSION

In this work, we have developed a genome-scale metabolic network model of *R. toruloides* and utilized the model to study the metabolic pathways for utilizing carbon sources derived from lignocellulosic biomass. The initial metabolic network was reconstructed from high-quality published metabolic network models of other organisms using orthologous protein mapping. There is some risk of incorrect reaction identification from the false positives in ortholog identification due to horizontal gene transfer (HGT), and there are many examples that highlight the importance of HGT in ascomycete yeasts (Goncalves et al., 2018; Shen et al., 2018; Kominek et al., 2019; Devia et al., 2020). However, ortholog identification continues to be a standard practice for the initial reconstruction of genome-scale metabolic networks for non-model organisms and currently we do not have evidence of large scale HGT in *R. toruloides*. As with any model, our model will need to be improved and re-evaluated over time in cases of HGT and as new metabolic pathways are characterized. The developed model contains 1141 genes, 2398 reactions, and 2051 metabolites (1205 unique metabolites) in nine compartments. The lipid body was separated from



B

All model genes		Prediction	
78.6% accuracy		Essential	Not essential
Experiment	Essential	3833	3974
	Not essential	2549	20073

C

Conditionally essential genes in experiment		Prediction	
78.7% accuracy		Essential	Not essential
Experiment	Essential	96	124
	Not essential	302	1476

FIGURE 7 | Model evaluation using high-throughput gene essentiality data. **(A)** RB-TDNA sequencing fitness score data for all genes in the model. Comparison of model predicted gene essentiality and experimental data for **(B)** all model genes and **(C)** conditionally essential genes in experiment.

the biomass equation allowing the independent simulation of lipid accumulation or mobilization in oleaginous yeasts. The separation of the lipid body from biomass eliminated the need for more than one biomass equation depending on the growth condition or lipid content. Multi-omics analysis and metabolic network reconstruction identified unique reactions and enzymes as well as their localization for unique pentose and aromatic compound utilization pathways in *R. toruloides*. The pentose and aromatic compound utilization pathways proposed in this study have not been suggested in previously published multi-omics studies or genome-scale metabolic models of *R. toruloides* (Bommareddy et al., 2015; Dinh et al., 2019; Tiukova et al., 2019a,b; Lopes et al., 2020; Pinheiro et al., 2020). The first genome-scale metabolic model for *R. toruloides* was recently built for strain NP11 (Tiukova et al., 2019b), and a proteomics study of xylose metabolism was conducted by the same research group (Tiukova et al., 2019a). Another genome-scale metabolic model was shortly after published for strain IFO0880 utilizing the functional genomics data (Coradetti et al., 2018). More recent studies utilized these models to study the utilization of different carbon sources, but their focus was primarily on lipid production (Lopes et al., 2020; Pinheiro et al., 2020).The metabolic network model developed in this study was reconstructed and manually curated reproducibly using multi-omics data and electronic notebooks, and validated against high-throughput growth phenotypes in 213 growth conditions and conditional gene essentiality in 27 growth conditions with high prediction accuracies, significantly expanding the breadth and depth of metabolic coverage from previously published models (Dinh et al., 2019; Tiukova et al., 2019b). We believe that the developed metabolic network for *R. toruloides* is most complete and accurate to date, and the multi-omics data and metabolic model presented in this study will be useful for studying and engineering *R. toruloides* for lignocellulosic biomass conversion.

DATA AVAILABILITY STATEMENT

The RNA sequencing data (SRP143805–SRP143836 and SRP143838) and RB-TDNA sequencing data (PRJNA595384) are available at the NCBI Sequence Read Archive. Fitness scores are available at the fungal fitness browser (http://fungalfit.genomics.lbl.gov). The mass spectrometry proteomics data are deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD022377 and 10.6019/PXD02237. The metabolomics data are deposited to the Open Science Framework (OSF) and available at https://osf.io/tnqwx/. Processed multi-omics dataset is available as a **Supplementary File**. The Jupyter notebooks, processed multi-omics dataset, and genome-scale metabolic network model files will also be made available on a GitHub repository (https://github.com/AgileBioFoundry/Rt_IFO0880).

AUTHOR CONTRIBUTIONS

JK, SC, JS, JG, and JM conceived and designed the experiments. JK, SC, JZ, JY, NM, Y-MK, YG, and KB-J performed the experiments and analyzed the data. JK, SC, Y-MK, and YG wrote the initial manuscript. All authors read, revised, and approved the final manuscript.

FUNDING

The work was part of the DOE Agile BioFoundry (http://agilebiofoundry.org), supported by the United States Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, under Award No. DE-NL0030038. This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the United States Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the United States Department of Energy. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative (Proposal ID 49515), and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.612832/full#supplementary-material

Supplementary File 1 | Jupyter notebooks for metabolic network model development and multi-omics data analysis.

Supplementary File 2 | Genome-scale metabolic network model of *R. toruloides* IFO0880 in JSON, SBML, and MAT format and the Memote HTML report.

Supplementary File 3 | Genome annotation of *R. toruloides* IFO0880.

Supplementary File 4 | Multi-omics dataset for *R. toruloides* IFO0880.

Supplementary File 5 | Comparison of gene essentiality predictions by the developed model and a previously published model.

REFERENCES

Adeboye, P. T., Bettiga, M., Aldaeus, F., Larsson, P. T., and Olsson, L. (2015). Catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid by *Saccharomyces cerevisiae* yields less toxic products. *Microb. Cell Fact.* 14:149. doi: 10.1186/s12934-015-0338-x

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Bommareddy, R. R., Sabra, W., Maheshwari, G., and Zeng, A. P. (2015). Metabolic network analysis and experimental study of lipid production in *Rhodosporidium toruloides* grown on single and mixed substrates. *Microb. Cell Fact.* 14:36. doi: 10.1186/s12934-015-0217-5

Bommareddy, R. R., Sabra, W., and Zeng, A. P. (2017). Glucose-mediated regulation of glycerol uptake in *Rhodosporidium toruloides*: insights through transcriptomic analysis on dual substrate fermentation. *Eng. Life Sci.* 17, 282–291. doi: 10.1002/elsc.201600010

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F., Balaji, S., Fu, W., et al. (2011). Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 7:518. doi: 10.1038/msb.2011.52

Coradetti, S. T., Pinel, D., Geiselman, G. M., Ito, M., Mondo, S. J., Reilly, M. C., et al. (2018). Functional genomics of lipid metabolism in the oleaginous yeast *Rhodosporidium toruloides*. *eLife* 7:e32110. doi: 10.7554/eLife.32110

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., et al. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genom.* 19:332. doi: 10.1186/s12864-018-4703-0

Desmond, E., and Gribaldo, S. (2009). Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* 1, 364–381. doi: 10.1093/gbe/evp036

Devia, J., Bastias, C., Kessi-Perez, E. I., Villarroel, C. A., De Chiara, M., Cubillos, F. A., et al. (2020). Transcriptional activity and protein levels of horizontally acquired genes in yeast reveal hallmarks of adaptation to fermentative environments. *Front. Genet.* 11:293. doi: 10.3389/fgene.2020.00293

Dinh, H. V., Suthers, P. F., Chan, S. H. J., Shen, Y., Xiao, T., Deewan, A., et al. (2019). A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. *Metab. Eng. Commun.* 9:e00101. doi: 10.1016/j.mec.2019.e00101

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1777–1782. doi: 10.1073/pnas.0610772104

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Fei, Q., O'Brien, M., Nelson, R., Chen, X., Lowell, A., and Dowe, N. (2016). Enhanced lipid production by *Rhodosporidium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnol. Biofuels* 9:130. doi: 10.1186/s13068-016-0542-x

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., et al. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3:121. doi: 10.1038/msb4100155

Goncalves, C., Wisecaver, J. H., Kominek, J., Oom, M. S., Leandro, M. J., Shen, X. X., et al. (2018). Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. *eLife* 7:e33034. doi: 10.7554/eLife.33034

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704. doi: 10.1093/nar/gkt1183

Hiller, K., Hangebrauk, J., Jager, C., Spura, J., Schreiber, K., and Schomburg, D. (2009). MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal. Chem.* 81, 3429–3439. doi: 10.1021/ac802689c

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Jagtap, S. S., and Rao, C. V. (2018). Production of D-arabitol from D-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Appl. Microbiol. Biotechnol.* 102, 143–151. doi: 10.1007/s00253-017-8581-1

Kim, S., and Pevzner, P. A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5:5277. doi: 10.1038/ncomms6277

Kim, Y. M., and Heyman, H. M. (2018). Mass spectrometry-based metabolomics. *Methods Mol. Biol.* 1775, 107–118. doi: 10.1007/978-1-4939-7804-5_10

Kim, Y. M., Nowack, S., Olsen, M. T., Becraft, E. D., Wood, J. M., Thiel, V., et al. (2015). Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. *Front. Microbiol.* 6:209. doi: 10.3389/fmicb.2015.00209

King, Z. A., Drager, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 11:e1004321. doi: 10.1371/journal.pcbi.1004321

King, Z. A., Lu, J., Drager, 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, D515–D522. doi: 10.1093/nar/gkv1049

Kominek, J., Doering, D. T., Opulente, D. A., Shen, X. X., Zhou, X., DeVirgilio, J., et al. (2019). Eukaryotic acquisition of a bacterial operon. *Cell* 176, 1356–1366.e1310. doi: 10.1016/j.cell.2019.01.034

Kyle, J. E., Crowell, K. L., Casey, C. P., Fujimoto, G. M., Kim, S., Dautel, S. E., et al. (2017). LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics* 33, 1744–1746. doi: 10.1093/bioinformatics/btx046

Lachance, J. C., Lloyd, C. J., Monk, J. M., Yang, L., Sastry, A. V., Seif, Y., et al. (2019). BOFdat: generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput. Biol.* 15:e1006971. doi: 10.1371/journal.pcbi.1006971

Lee, A. K., Banta, A. B., Wei, J. H., Kiemle, D. J., Feng, J., Giner, J. L., et al. (2018). C-4 sterol demethylation enzymes distinguish bacterial and eukaryotic sterol synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5884–5889. doi: 10.1073/pnas.1802930115

Lee, H. J., Nakayasu, M., Akiyama, R., Kobayashi, M., Miyachi, H., Sugimoto, Y., et al. (2019). Identification of a 3beta-hydroxysteroid Dehydrogenase/ 3-ketosteroid reductase involved in alpha-tomatine biosynthesis in tomato. *Plant Cell Physiol.* 60, 1304–1315. doi: 10.1093/pcp/pcz049

Lee, J. J., Chen, L., Shi, J., Trzcinski, A., and Chen, W. N. (2014). Metabolomic profiling of *Rhodosporidium toruloides* grown on glycerol for carotenoid production during different growth phases. *J. Agric. Food Chem.* 62, 10203–10209. doi: 10.1021/jf502987q

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., P. Babaei, P., et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* doi: 10.1038/s41587-020-0446-y

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liu, H., Jiao, X., Wang, Y., Yang, X., Sun, W., Wang, J., et al. (2017). Fast and efficient genetic transformation of oleaginous yeast *Rhodosporidium toruloides* by using electroporation. *FEMS Yeast Res.* 17:fox017.

Liu, X., Zhang, Y., Liu, H., Jiao, X., Zhang, Q., Zhang, S., et al. (2019). RNA interference in the oleaginous yeast *Rhodosporidium toruloides*. *FEMS Yeast Res.* 19:foz031. doi: 10.1093/femsyr/foz031

Lopes, H. J. S., Bonturi, N., Kerkhoven, E. J., Miranda, E. A., and Lahtvee, P. J. (2020). C/N ratio and carbon source-dependent lipid production profiling in *Rhodotorula toruloides*. *Appl. Microbiol. Biotechnol.* 104, 2639–2649. doi: 10.1007/s00253-020-10386-5

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Ma, T., Shi, B., Ye, Z., Li, X., Liu, M., Chen, Y., et al. (2019). Lipid engineering combined with systematic metabolic engineering of *Saccharomyces cerevisiae* for high-yield production of lycopene. *Metab. Eng.* 52, 134–142. doi: 10.1016/j.ymben.2018.11.009

Mo, M. L., Palsson, B. O., and Herrgard, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* 3:37. doi: 10.1186/1752-0509-3-37

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

Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N., and Smith, R. D. (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* 32, 215–217. doi: 10.1016/j.compbiolchem.2008.02.006

Nakayasu, E. S., Nicora, C. D., Sims, A. C., Burnum-Johnson, K. E., Kim, Y. M., Kyle, J. E., et al. (2016). MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *mSystems* 1:e0043-16. doi: 10.1128/mSystems.00043-16

Nogales, J., Palsson, B. O., and Thiele, I. (2008). A genome-scale metabolic reconstruction of *Pseudomonas* putida KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* 2:79. doi: 10.1186/1752-0509-2-79

Nora, L. C., Wehrs, M., Kim, J., Cheng, J. F., Tarver, A., Simmons, B. A., et al. (2019). A toolset of constitutive promoters for metabolic engineering of *Rhodosporidium toruloides*. *Microb. Cell Fact.* 18:117. doi: 10.1186/s12934-019-1167-0

Otoupal, P. B., Ito, M., Arkin, A. P., Magnuson, J. K., Gladden, J. M., and Skerker, J. M. (2019). Multiplexed CRISPR-Cas9-based genome editing of *Rhodosporidium toruloides*. *mSphere* 4:e0099-19. doi: 10.1128/mSphere.00099-19

Park, Y. K., Nicaud, J. M., and Ledesma-Amaro, R. (2018). The engineering potential of *Rhodosporidium toruloides* as a workhorse for biotechnological applications. *Trends Biotechnol.* 36, 304–317. doi: 10.1016/j.tibtech.2017.10.013

Penning, T. M., Chen, M., and Jin, Y. (2015). Promiscuity and diversity in 3-ketosteroid reductases. *J. Steroid Biochem. Mol. Biol.* 151, 93–101. doi: 10.1016/j.jsbmb.2014.12.003

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Pinheiro, M. J., Bonturi, N., Belouah, I., Miranda, E. A., and Lahtvee, P. J. (2020). Xylose metabolism and the effect of oxidative stress on lipid and carotenoid production in *Rhodotorula toruloides*: insights for future biorefinery. *Front. Bioeng. Biotechnol.* 8:1008. doi: 10.3389/fbioe.2020.01008

Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* 11:395. doi: 10.1186/1471-2105-11-395

Polpitiya, A. D., Qian, W. J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G., et al. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556–1558. doi: 10.1093/bioinformatics/btn217

Price, M. N., Ray, J., Iavarone, A. T., Carlson, H. K., Ryan, E. M., Malmstrom, R. R., et al. (2019). Oxidative pathways of deoxyribose and deoxyribonate catabolism. *mSystems* 4:e00297-18. doi: 10.1128/mSystems.00297-18

Rabeharindranto, H., Castano-Cerezo, S., Lautier, T., Garcia-Alles, L. F., Treitz, C., Tholey, A., et al. (2019). Enzyme-fusion strategies for redirecting and improving carotenoid synthesis in *S. cerevisiae*. *Metab. Eng. Commun.* 8:e00086. doi: 10.1016/j.mec.2019.e00086

Rodriguez, S., Denby, C. M., Van Vu, T., Baidoo, E. E., Wang, G., and Keasling, J. D. (2016). ATP citrate lyase mediated cytosolic acetyl-CoA biosynthesis increases mevalonate production in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 15:48. doi: 10.1186/s12934-016-0447-1

Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., et al. (2018). Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175, 1533–1545.e1520. doi: 10.1016/j.cell.2018.10.023

Sigurdsson, M. I., Jamshidi, N., Steingrimsson, E., Thiele, I., and Palsson, B. O. (2010). A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst. Biol.* 4:140. doi: 10.1186/1752-0509-4-140

Stamboulian, M., Guerrero, R. F., Hahn, M. W., and Radivojac, P. (2020). The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction. *Bioinformatics* 36(Suppl._1), i219–i226. doi: 10.1093/bioinformatics/btaa468

Sun, W., Yang, X., Wang, X., Lin, X., Wang, Y., Zhang, S., et al. (2017). Homologous gene targeting of a carotenoids biosynthetic gene in *Rhodosporidium toruloides* by Agrobacterium-mediated transformation. *Biotechnol. Lett.* 39, 1001–1007. doi: 10.1007/s10529-017-2324-3

Sundstrom, E., Yaegashi, J., Yan, J. P., Masson, F., Papa, G., Rodriguez, A., et al. (2018). Demonstrating a separation-free process coupling ionic liquid pretreatment, saccharification, and fermentation with *Rhodosporidium toruloides* to produce advanced biofuels. *Green Chem.* 20, 2870–2879. doi: 10.1039/c8gc00518d

Tiukova, I. A., Brandenburg, J., Blomqvist, J., Sampels, S., Mikkelsen, N., Skaugen, M., et al. (2019a). Proteome analysis of xylose metabolism in *Rhodotorula toruloides* during lipid production. *Biotechnol. Biofuels* 12:137. doi: 10.1186/s13068-019-1478-8

Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M., and Kerkhoven, E. J. (2019b). Genome-scale model of *Rhodotorula toruloides* metabolism. *Biotechnol. Bioeng.* 116, 3396–3408. doi: 10.1002/bit.27162

Wei, S., Jian, X., Chen, J., Zhang, C., and Hua, Q. (2017). Reconstruction of genome-scale metabolic model of *Yarrowia lipolytica* and its application in overproduction of triacylglycerol. *Bioresour. Bioprocess* 4:51. doi: 10.1186/s40643-017-0180-6

Wiebe, M. G., Koivuranta, K., Penttila, M., and Ruohonen, L. (2012). Lipid production in batch and fed-batch cultures of *Rhodosporidium toruloides* from

5 and 6 carbon carbohydrates. *BMC Biotechnol.* 12:26. doi: 10.1186/1472-6750-12-26

Yaegashi, J., Kirby, J., Ito, M., Sun, J., Dutta, T., Mirsiaghi, M., et al. (2017). *Rhodosporidium toruloides*: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnol. Biofuels* 10:241. doi: 10.1186/s13068-017-0927-5

Yamada, Y., and Kondo, K. (1973). Coenzyme q system in the classification of the yeast genera *Rhodotorula* and *Cryptococcus*, and the yeast-like genera *Sporobolomyces* and *Rhodosporidium*. *J. Gen. Appl. Microbiol.* 19, 59–77. doi: 10.2323/jgam.19.59

Zhang, S., Skerker, J. M., Rutter, C. D., Maurer, M. J., Arkin, A. P., and Rao, C. V. (2016). Engineering *Rhodosporidium toruloides* for increased lipid production. *Biotechnol. Bioeng.* 113, 1056–1066. doi: 10.1002/bit.25864

Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., et al. (2012). A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nat. Commun.* 3:1112. doi: 10.1038/ncomms2112

Zhuang, X., Kilian, O., Monroe, E., Ito, M., Tran-Gymfi, M. B., Liu, F., et al. (2019). Monoterpene production by the carotenogenic yeast *Rhodosporidium toruloides*. *Microb. Cell Fact.* 18:54. doi: 10.1186/s12934-019-1099-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

METHODS
published: 09 February 2021
doi: 10.3389/fbioe.2021.612893

Multiomics Data Collection, Visualization, and Utilization for Guiding Metabolic Engineering

Somtirtha Roy[1,2†], *Tijana Radivojevic*[1,2,3†], *Mark Forrer*[2,3,4], *Jose Manuel Marti*[1,2,3], *Vamshi Jonnalagadda*[1,2], *Tyler Backman*[1,3], *William Morrell*[2,3,4], *Hector Plahar*[1,2], *Joonhoon Kim*[3,5], *Nathan Hillson*[1,2,3] *and Hector Garcia Martin*[1,2,3,6*]

[1] *Lawrence Berkeley National Laboratory, Biological Systems and Engineering Division, Berkeley, CA, United States,* [2] *Department of Energy, Agile BioFoundry, Emeryville, CA, United States,* [3] *Joint BioEnergy Institute, Emeryville, CA, United States,* [4] *Sandia National Laboratories, Biomaterials and Biomanufacturing, Livermore, CA, United States,* [5] *Chemical and Biological Processes Development Group, Pacific Northwest National Laboratory, Richland, WA, United States,* [6] *BCAM, Basque Center for Applied Mathematics, Bilbao, Spain*

OPEN ACCESS

Edited by:
Eduard Kerkhoven, Chalmers University of Technology, Sweden

Reviewed by:
Mario Andrea Marchisio, Tianjin University, China
Cleo Kontoravdi, Imperial College London, United Kingdom

***Correspondence:**
Hector Garcia Martin
hgmartin@lbl.gov

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *01 October 2020*
Accepted: *13 January 2021*
Published: *09 February 2021*

Citation:
Roy S, Radivojevic T, Forrer M, Marti JM, Jonnalagadda V, Backman T, Morrell W, Plahar H, Kim J, Hillson N and Garcia Martin H (2021) Multiomics Data Collection, Visualization, and Utilization for Guiding Metabolic Engineering. Front. Bioeng. Biotechnol. 9:612893. doi: 10.3389/fbioe.2021.612893

Biology has changed radically in the past two decades, growing from a purely descriptive science into also a design science. The availability of tools that enable the precise modification of cells, as well as the ability to collect large amounts of multimodal data, open the possibility of sophisticated bioengineering to produce fuels, specialty and commodity chemicals, materials, and other renewable bioproducts. However, despite new tools and exponentially increasing data volumes, synthetic biology cannot yet fulfill its true potential due to our inability to predict the behavior of biological systems. Here, we showcase a set of computational tools that, combined, provide the ability to store, visualize, and leverage multiomics data to predict the outcome of bioengineering efforts. We show how to upload, visualize, and output multiomics data, as well as strain information, into online repositories for several isoprenol-producing strain designs. We then use these data to train machine learning algorithms that recommend new strain designs that are correctly predicted to improve isoprenol production by 23%. This demonstration is done by using synthetic data, as provided by a novel library, that can produce credible multiomics data for testing algorithms and computational tools. In short, this paper provides a step-by-step tutorial to leverage these computational tools to improve production in bioengineered strains.

Keywords: machine learning, flux analysis, metabolic engineering, biofuels, synthetic biology, multiomics analysis

INTRODUCTION

Synthetic biology represents another step in the development of biology as an engineering discipline. The application of engineering principles such as standardized genetic parts (Canton et al., 2008; Müller and Arndt, 2012) or the application of Design-Build-Test-Learn (DBTL) cycles (Petzold et al., 2015; Nielsen and Keasling, 2016) has transformed genetic and metabolic engineering in significant ways. Armed with this new engineering framework, synthetic biology is creating products to tackle societal problems in ways that only biology can enable. Synthetic biology, for example, is being leveraged to produce renewable biofuels to combat climate change

(Peralta-Yahya et al., 2012; Beller et al., 2015; Chubukov et al., 2016), improve crop yields (Roell and Zurbriggen, 2020), combat the spread of diseases (Kyrou et al., 2018), synthesize medical drugs (Ajikumar et al., 2010; Paddon and Keasling, 2014), biomaterials (Bryksin et al., 2014), and plant-based foods (Meat-free outsells beef, 2019).

However, the development of synthetic biology is hindered by our inability to predict the results of engineering outcomes. DNA synthesis and CRISPR-based genetic editing (Ma et al., 2012; Doudna and Charpentier, 2014) allow us to produce and change DNA (the working code of the cell) with unparalleled ease, but we can rarely predict how that modified DNA will impact cell behavior (Gardner, 2013). As a consequence, it is not possible to design a cell to fit a desired specification: e.g., have the cell produce *X* grams of a specified biofuel or an anticancer agent. Hence, metabolic engineering is often mired in trial-and-error approaches that result in very long development times (Hodgman and Jewett, 2012). In this context, machine learning has recently appeared as a powerful tool that can provide the predictive power that bioengineering needs to be effective and impactful (Carbonell et al., 2019; Radivojević et al., 2020; Zhang et al., 2020).

Furthermore, although there is a growing abundance of phenotyping data, the tools to systematically leverage these data to improve predictive power are lacking. For example, transcriptomics data has a doubling time of 7 months (Stephens et al., 2015), and high-throughput techniques for proteomics (Chen et al., 2019) and metabolomics (Fuhrer and Zamboni, 2015) are becoming increasingly available. Often, metabolic engineers struggle to synthesize this data deluge into precise actionable items (e.g., down regulate this gene and knock out this transcription factor) to obtain their desired goal (e.g., increase productivity to commercially viable levels).

Here, we showcase how to combine the following existing tools to leverage omics data and suggest next steps (**Figure 1**): the Inventory of Composable Elements (ICE), the Experiment Data Depot (EDD), the Automated Recommendation Tool (ART), and Jupyter Notebooks. ICE (Ham et al., 2012) is an open source repository platform for managing information about DNA parts and plasmids, proteins, microbial host strains, and plant seeds. EDD (Morrell et al., 2017) is an open source online repository of experimental data and metadata. ART (Radivojević et al., 2020; Zhang et al., 2020) is a library that leverages machine learning for synthetic biology purposes, providing predictive models and recommendations for the next set of experiments. Jupyter notebooks are interactive documents that contain live code, equations, visualizations, and explanatory text (Project Jupyter | Home[1]; IOS Press Ebooks - Jupyter Notebooks) (Kluyver et al., 2016). When combined, this set of tools can effectively store, visualize, and leverage synthetic biology data to enable predictive bioengineering and effective actionable items for the next DBTL cycle. We will demonstrate this with an example in which we leverage multiomics data to improve the production of isoprenol, a potential biofuel (Kang et al., 2019). This multiomics data set

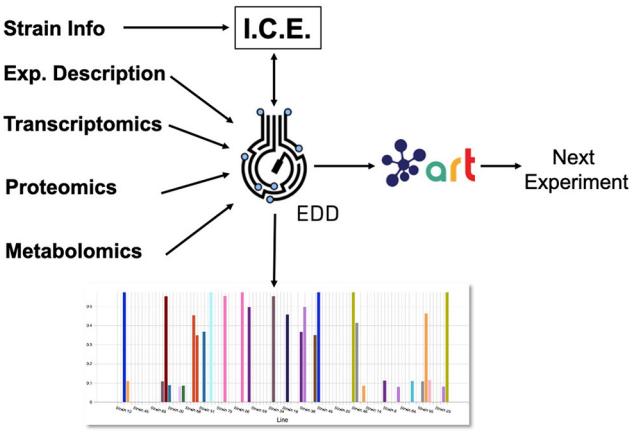


FIGURE 1 | Combining several tools to guide metabolic engineering. The combination of ICE, EDD, and ART provides the ability to store, visualize and leverage multiomics data to guide bioengineering. Here, we showcase how to use this collection of tools to improve the production of isoprenol in *E. coli* for a simulated data set.

is a synthetic data set (i.e., simulated computationally without experimental work) generated through the new Omics Mock Generator (OMG) library. Synthetic data provide the advantage of easily producing large amounts of multimodal data that would be prohibitively expensive to produce experimentally. In this way, we can concentrate this manuscript on the demonstration of computational tools rather than the details and vagaries of data collection. The OMG library provides an easily accessible source of biologically believable data that can be used to test algorithms and tools systematically. Validation for these tools with experimental data has already been provided elsewhere (e.g., Radivojević et al., 2020; Zhang et al., 2020).

METHODS

Synthetic Data Generator Library (OMG)

The Omics Mock Generator (OMG) library is used to provide the synthetic multiomics data needed to test the computational tools described here (**Figure 2**). Since experimental multiomics data are expensive and non-trivial to produce, OMG provides a quick and easy way to produce large amounts of multiomics data that are based on plausible metabolic assumptions. OMG creates fluxes based on Flux Balance Analysis (FBA) and growth rate maximization (Orth et al., 2010), leveraging COBRApy (Ebrahim et al., 2013). OMG can use any genome-scale model, but in this case we have used the iJO1366 *E. coli* genome scale model, augmented with an isoprenol pathway obtained from the iMM904 *S. cerevisiae* model (Notebook A). In order to obtain proteomics data, we assume that the corresponding protein expression and gene transcription are linearly related to the fluxes. We also assume the concentration of metabolites to be loosely related to the fluxes of the reactions that consume or produce them (no fluxes → no metabolite): the amount of metabolite present is assumed to be proportional to the sum of absolute fluxes coming in and out of the metabolite. Therefore, although the data provided by OMG is not real, it is more realistic

[1]Project Jupyter | Home. Available online at: https://jupyter.org/ (accessed September 4, 2020).

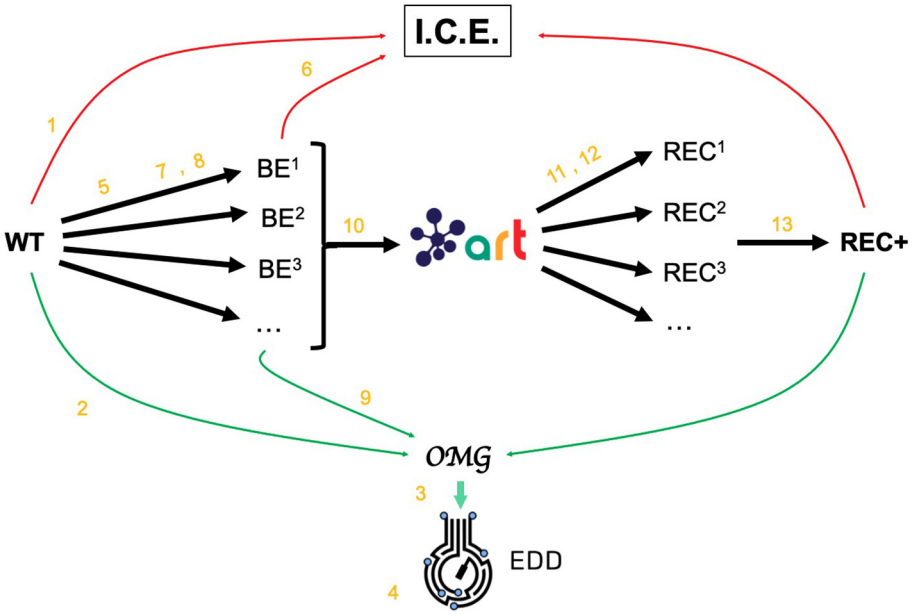


FIGURE 2 | Demonstrating ICE, EDD, and ART using synthetic data. For the purposes of the demonstration of how ICE, EDD, and ART work together, we use a synthetic data set of multiomics data (transcriptomics, proteomics, metabolomics, fluxomics) for several time points created by the Omics Mock Generator (OMG) library (see Methods section). We start with a base strain (wild type, or WT) that is bioengineered according to several designs (i.e., knockout malate dehydrogenase, overexpress citrate synthase) suggested by ART. The results are 95 bioengineered strains (BE^1, BE^2… etc.) for which experimental data (isoprenol production levels) are simulated through OMG and stored in EDD and ICE. These data are then leveraged by ART to recommend, using machine learning, new designs that are expected to improve isoprenol production (REC^1, REC^2, …). These recommendations and production predictions are compared with the ground truth provided by OMG. Each of these steps (in orange) is demonstrated through screencasts and Jupyter notebooks (**Table 1**).

than randomly generated data, providing a useful resource to test the scaling of algorithms and computational tools.

The data generated by OMG was used in this manuscript to test EDD input, output, and visualization, and to provide training data for ART (**Figure 2**). This tool can be a very useful resource for the rapid prototyping of new tools and algorithms.

Generating Flux Time Series Data

Fluxes describe the rates of metabolic reactions in a given organism, and can be easily generated through FBA. FBA assumes that the organism is under selective pressure to increase its growth rate (Orth et al., 2010; Lewis et al., 2012), hence searching for the fluxes that optimize it. FBA relies on genome-scale models, which provide a comprehensive description of all known genetically encoded metabolic reactions (Thiele and Palsson, 2010). FBA produces fluxes by solving through Linear Programming (LP) the following optimization problem:

$$\begin{aligned} &\text{Maximize} V_{\text{biomass}} \text{subject to :} \\ &\Sigma_j S_{ij} V_j = 0 \\ &lb_j \leq V_j \leq ub_j \end{aligned} \quad (1)$$

where S is the stoichiometry matrix of size m^*n (number of metabolites*number of reactions in the model), V_j is the flux for reaction j within the model ($j = 1,...,n$). The lower bound (lb_j) and upper bound (ub_j) provide the minimum and maximum for each reaction. For example, if the input carbon source is glucose and we know that the input in a given time lapse is −15 mmol/gdw/h, the upper bound and lower bound for the exchange reaction corresponding to glucose are set to this value: $-15 < V_{EX\_glc} < -15$. The solution to this optimization problem provides the fluxes that maximize growth rate, and that will be used later on to obtain transcriptomics, proteomics and metabolomics data. However, one must be aware that this optimization problem is underdetermined, and there are multiple solutions that satisfy exchange flux constraints.

We create time series of fluxes by doing a batch simulation based on FBA (see OMG library and Notebook A). We assume a given concentration for extracellular metabolites (e.g., 22 mM of glucose, or 18 mM of ammonium) and, for each time point, we run FBA for the model and update the extracellular metabolite concentration based on the exchange fluxes coming from the simulation (see Notebook A). For example, if an exchange flux of V_EX_glc_D = −15 mmol/gdw/hr is obtained for the model, the corresponding glucose concentration is adjusted as follows:

$$[glc\_D]_{new} = [glc\_D]_{old} - 15\, mmol/gdw/hr \cdot \Delta t \cdot [cell]$$

Where $[glc\_D]_{old}$ is the old glucose concentration, $[glc\_D]_{new}$ is the updated concentration, Δt is the time change and $[cell]$ is the cell concentration. In practice, we assume that the cell density increase is better described by an exponential than a

linear relationship, and $\Delta t \cdot [cell]$ is substituted by $\Delta[cell]/\mu$, where μ is the growth rate: $\mu = 1/[cell] \cdot \Delta[cell]/\Delta t$. The simulation proceeds until the carbon source (e.g., glucose) is exhausted. The result of the simulation is a set of fluxes, cell concentration and extracellular metabolite concentration for each time point (see **Figure 3**). The fluxes will be the base for calculating transcriptomics, proteomics, and metabolomics data, as shown below.

Generating Proteomics Data

The flux values obtained from FBA are subsequently used to generate proteomics data, which describe the concentration of the protein catalyzing a given reaction within the host organism. The protein concentration for each time point is derived from the corresponding fluxes through a linear relationship:

$$P_j = V_j/k + \beta \tag{2}$$

which is loosely inspired in the Michaelis-Menten equation (Heinrich and Schuster, 1996), where V_j is the flux of reaction j, P_j is the concentration of the protein catalyzing the reaction j, and k is a linear constant arbitrarily set to 0.1. The symbol β is an added random noise which is set to 5% of the signal.

Generating Transcriptomics Data

The aforementioned proteomics values are subsequently used to generate transcriptomics data, which describe the abundance of RNA transcripts linked to a given protein within the host organism. For simplicity, the transcript data is assumed to have a linear relationship with the proteomics data:

$$T_j = P_j/q + \gamma \tag{3}$$

where T_j is the abundance of RNA transcripts linked to the reaction j, and q is a linear constant arbitrarily set to 0.833. As above, γ is a random noise addition set to 5% of the signal data. This calculation is performed for each time point.

Generating Metabolomics Data

The flux values obtained from the FBA are also used to generate the metabolomics data, which describe the concentration of a given metabolite within the host organism. While finding the metabolite concentrations compatible with a given metabolic flux, protein concentrations, and transcript levels is a non-trivial endeavor, here we attempt to produce metabolite profiles that are not obviously unreasonable. Hence, we want concentrations of zero for metabolites that are connected to fluxes that are null, and non-zero in any other case. The easiest way to achieve this is by averaging the absolute value of all the fluxes producing or consuming the desired metabolite:

$$M_i = \Sigma_j |S_{ij} V_j| \, / n \tag{4}$$

where M_i is the concentration of the metabolite i, j is a reaction that involves metabolite i, and n is the total number of reactions in which metabolite i participates. This calculation is performed for each time point.

Generating Training Data for Machine Learning and Testing Predictions

We leverage the OMG library to create training data to showcase the use of ART to guide bioengineering (**Figure 2**). We will first create multiomics (transcriptomics, proteomics, metabolomics), cell concentration, and extracellular metabolite concentration data for the wild type *E. coli* strain (WT). Then we will use ART to suggest initial WT modifications (designs) so as to create enough data to train ART to be predictive. Those initial designs will be used by OMG to simulate isoprenol production data for bioengineered strains that include the genetic modifications indicated by the initial designs. ART will be trained on these data, and then used to suggest strain modifications that are predicted to increase isoprenol production. We will then compare ART predictions for isoprenol production with the "observed" results produced by using those designs to simulate bioengineered strains through OMG. In sum, OMG results are used as ground truth to be leveraged in testing ART's performance.

This process involves the following phases:

1. **Choosing input and response variables.** Since the objective is to improve production of isoprenol, we use isoprenol concentration as the response variable. By inspecting the *E. coli* network we choose the following fluxes connected to acetyl-CoA, which is the source for the isoprenol pathway: ACCOAC, MDH, PTAr, CS, ACACT1r, PPC, PPCK, PFL. These fluxes then form the set of input variables for ART (Radivojević et al., 2020).
2. **Representation of different strain designs (i.e., genetic modifications).** We will consider only two types of modifications for each flux: knock-out (KO) and doubling the flux (UP). This choice results in three categories for each of the fluxes, which additionally include no modification (NoMod). We denote these categories by 0, 1, and 2 for KO, NoMod, and UP, respectively. Considering eight fluxes and three options for design of each, the total number of possible designs is $3^8 = 6{,}561$.
3. **Choose training data size**. We choose the initial training data to consist of 96 nonequivalent designs (instances), including the WT strain. This choice mimics one 96-well plate run. Different designs (instances) here represent different engineered strains. These 96 designs represent about 1.46% of all design space.
4. **Generate initial designs.** Initial designs are generated using ART's feature for generating recommendations for the *initial* cycle, by setting its input parameter `initial_cycle` to True. This ART functionality relies on the Latin Hypercube method (McKay et al., 1979), which spaces out draws in a way that ensures the set of samples represents the variability of the full design space. Another parameter needed for ART is `num_recommendations`, which we set to 95 (see point 3 above). See the ART publication (Radivojević et al., 2020) for a list of other optional parameters. As the current version of ART deals only with continuous variables for the initial cycle,

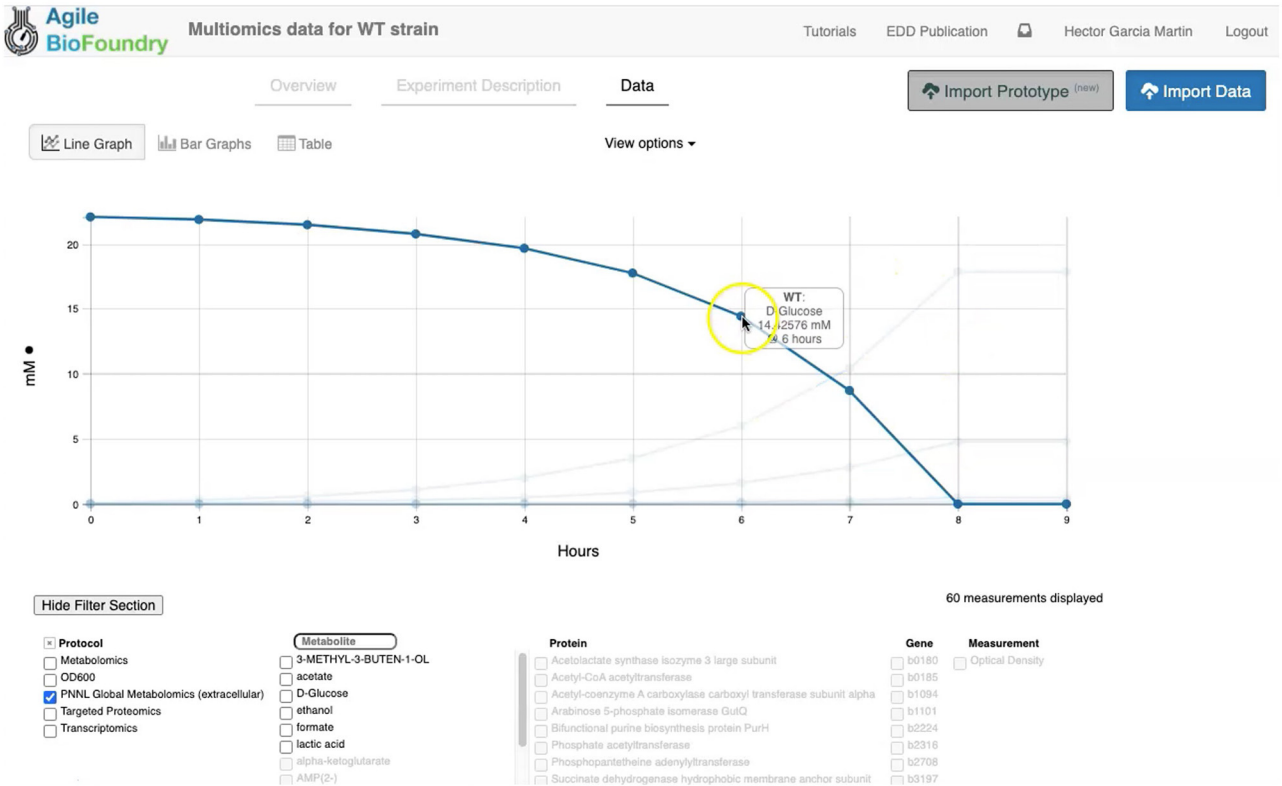


FIGURE 3 | Visualizing data in EDD. EDD provides data visualization in the form of bar and line charts. The lower menu provides filtering options to facilitate comparison of lines. More sophisticated visualization can be achieved by pulling the data from EDD through the REST API.

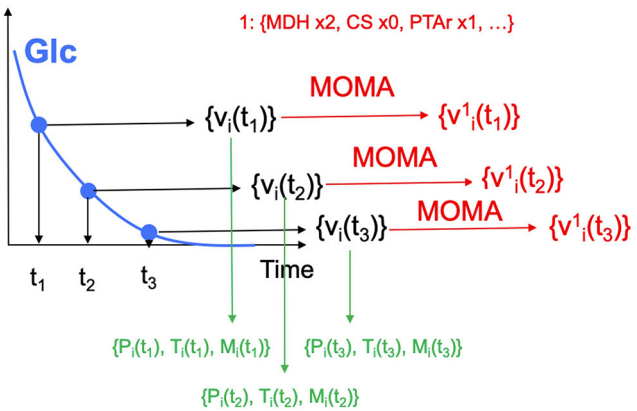


FIGURE 4 | Generating multiomics time series data. For each time point, we generate fluxes by solving an FBA problem, until glucose is fully consumed. MOMA is used in conjunction with the *design* (e.g., increase MDH flux 2-fold, knock CS out, maintain PTAr) to predict fluxes for the strain bioengineered according to the design.

ART's recommendations will be drawn from interval [0, 1], which is the default interval if no specific upper and lower bounds are provided in a separate file. We then transform each of those values into one of the defined categories {0, 1, 2} by applying the function f(x) = 3*floor(x). Finally, we add a WT strain design {1,1,…,1}. See Jupyter Notebook B for the details.

5. **Generate production data for the initial designs.** The initial designs from ART are used as input to the OMG library, generating our "ground truth" for the isoprenol production levels for each of the initial designs. This represents the strain construction and the corresponding phenotyping experiments, which are simulated through OMG's mechanistic modeling. In order to simulate how production is affected by the genetic changes suggested in the designs (e.g., knock out MDH, upregulate PFL and do not change CS), we used MOMA (Segrè et al., 2002) for each of the time points in the flux series (**Figure 4**). Details can be found in Notebook C.
6. **Training ART with initial production data**. ART uses the initial designs and their corresponding productions from phases 4 and 5 to train, enabling it to predict isoprenol production for designs it has not seen. Details can be found in Jupyter Notebook D.
7. **Generate next-cycle design recommendations using ART**. Once trained, ART generates 10 recommendations that are expected to improve isoprenol production. Predictions are generated for each recommendation in the form of probability distributions (**Figure 5**). Details can be found in Notebook D.
8. **Compare ART predictions to ground truth**. Finally, we take ART's design recommendations from phase 7

```
art.recommend(data=data)
art.recommendations
```

	ACCOAC	MDH	PTAr	CS	ACACT1r	PPC	PPCK	PFL	3-METHYL-3-BUTEN-1-OL
0	2.0	2.0	1.0	2.0	2.0	2.0	0.0	1.0	0.573584
1	2.0	2.0	2.0	2.0	2.0	2.0	0.0	1.0	0.572193
2	1.0	2.0	1.0	2.0	2.0	2.0	0.0	1.0	0.572167
3	2.0	1.0	1.0	2.0	2.0	2.0	0.0	1.0	0.572039
4	2.0	2.0	1.0	2.0	2.0	1.0	0.0	1.0	0.571967
5	2.0	2.0	2.0	2.0	2.0	1.0	0.0	1.0	0.571282
6	2.0	2.0	1.0	2.0	2.0	2.0	0.0	2.0	0.570968
7	2.0	2.0	2.0	2.0	2.0	2.0	0.0	2.0	0.570915
8	1.0	2.0	2.0	2.0	2.0	2.0	0.0	1.0	0.570771
9	2.0	1.0	2.0	2.0	2.0	2.0	0.0	1.0	0.570769

It turns out that all recommendations indicate that the CS and ACACT1r reaction fluxes should double and PPCK should be knocked out.

We can visualize the predicted distribution for all recommendations:

```
pred_col_name = 'Mean predicted Isoprenol [mM]'
plot.recomm_distribution(art, art.find_current_best(), [pred_col_name])
```

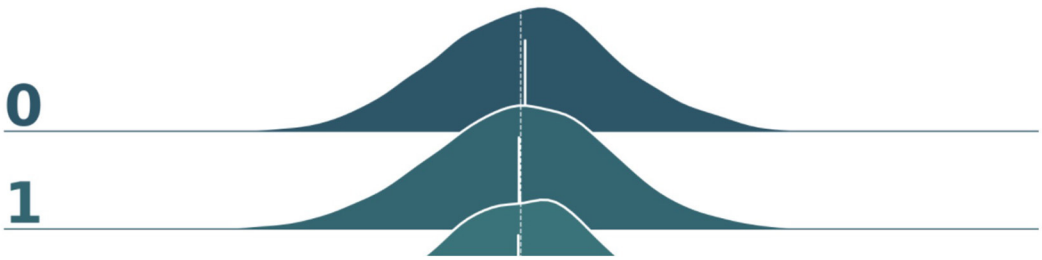


FIGURE 5 | Using machine learning to predict production and recommend new designs. The ART library takes a DataFrame containing input designs (i.e., which fluxes to overexpress, 2, keep the same, 1, or knock out, 0) and isoprenol production (response). The trained model recommends new designs that have the highest production. The recommendations come with predictions of production in a probabilistic fashion: i.e., the probability of production of 10, 15, 25, 40 mMol, etc.

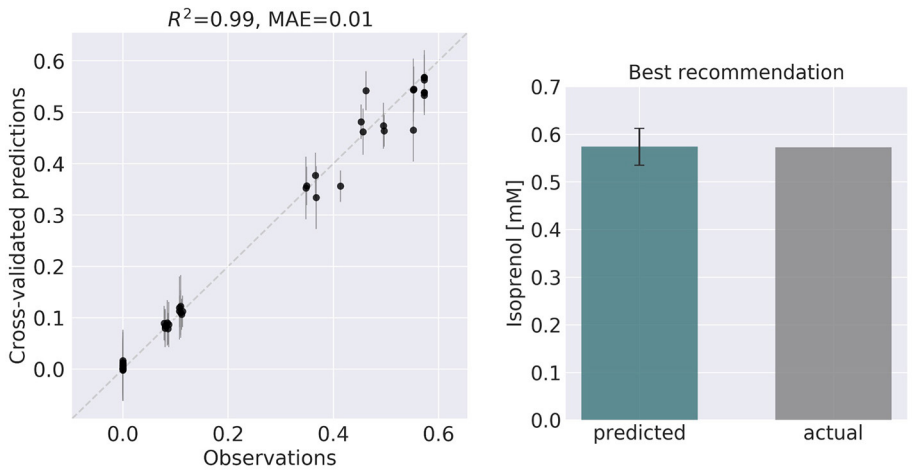


FIGURE 6 | ART recommendations display production levels production very similar to predictions. Left panel compares cross-validated predictions for isoprenol production from ART versus the values obtained through the OMG library for the training data set. Cross-validation keeps a part of the data set hidden from the training to compare against predictions, providing a good idea of the quality of predictions for new data sets. The right panel compares the predicted production for the recommended strain (#97) vs the actual production as generated through the OMG library. The comparison indicates a very good agreement between the prediction and observation.

and use OMG to simulate the corresponding ground truth isoprenol productions, similarly as in phase 5. We compare these isoprenol productions (observed production) vs. the machine learning recommendations from ART (**Figure 6**). Details can be found in Jupyter Notebook E.

Data Formats for Generic Input Files

The data formats for the generic input files in EDD involve five columns. The first column specifies the line (e.g., WT). The second one is the measurement type identifier, which involves a standardized choice: Pubchem IDs for metabolites, UniProt IDs for proteins, and Genbank gene IDs for transcripts (e.g., CID: 715, acetate). The third column is the time point (e.g., 0 h). The fourth column is a value for the corresponding identifier (e.g., 2.4). The final column is the unit corresponding to this value: FPKM for transcriptomics data, proteins/cell for proteomics data, and mg/L or mM for metabolomics data. Optical Density (OD) has no units.

RESULTS AND DISCUSSION

In the following sections, we will provide a step-by-step example of how to use this suite of tools (ICE, EDD, ART, OMG, and Jupyter notebooks) to guide metabolic engineering efforts and increase isoprenol production in *E. coli* for a synthetic data set (**Figure 2**, and "Methods" section). Using synthetic (simulated) data allows for a more effective demonstration, since there is no obstacle to creating time series multiomics data involving transcriptomics, proteomics, metabolomics, and fluxomics. For the purposes of this demonstration, creating such a data set using real experiments would be prohibitively expensive and limiting.

Storing Strain Information in ICE

Our first step involves storing the initial strain information in the Agile BioFoundry (ABF) instance of ICE: https://public-registry.agilebiofoundry.org/ (see Screencast 1 in **Tables 1**, **2**). Storing strain information in ICE provides a standard way to document the design phase and make this information available for later use. ICE provides access controls so that the strains can be created as the experiment progresses, and then made public later (e.g., at publication time). We will initially store the information for the base strain (or wild type, WT). After creating an account and logging in, we click on "Create Entry" and choose "Strain." We then fill the relevant strain information (e.g., "Name," "Biosafety Level," "Description," "Sequence," etc.). Finally, we will click on "Submit" to create the strain entry. The strain is now available on the ICE instance and is assigned a part number that will be used to enter experimental data into EDD as the next step. Strain information can also be easily downloaded from ICE through the GUI (Screencast 5).

Importing Data Into EDD

Importing the multiomics data set into EDD allows for standardized data storage and retrieval. The use of the Experiment Data Depot (EDD) allows for the seamless integration of data generated from different analytical methods (e.g., mass spectroscopy, sequencing, HP-LC). EDD focuses on storing the *biologically interpretable data*: i.e., data that can be immediately interpreted by a biologist without requiring detailed knowledge of the analytical measurement technique (e.g., metabolite concentrations rather than GC-MS traces, or transcripts per cells rather than individual genomic sequencing reads). Furthermore, the use of standardized schemas is fundamental for downstream analysis such as machine learning or mechanistic modeling. Indeed, it is estimated that 50–80% of a data scientist's time is spent with the type of data wrangling that EDD avoids by providing an ontology (For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights - The New York Times) (Lohr, 2014). This ontology reflects the objects and processes most often encountered in metabolic engineering experiments (see Figure 2 in Morrell et al., 2017). For example, a typical metabolic engineering project starts by obtaining several *strains* from a strain repository (e.g., DMZ, ATCC, ICE). Those strains are cultured in different flasks under different conditions (different media, induction levels etc.), which we call *lines* because they are a combination of strain and culture conditions, and represent a different line of enquiry or question being asked (e.g., does this strain under this condition improve production?). The final steps usually involve making measurements relevant to the experiment's ultimate goal. These measurements could be the concentration of the metabolite that the strains are engineered

TABLE 1 | Workflow for the paper and demonstrative Jupyter notebooks and screencasts.

Step	Description	Demonstration
1	WT strain import into ICE	Screencast 1
2	WT multiomics data generation	Notebook A
3	EDD import of WT multiomics data	Screencast 2
4	EDD visualization of WT multiomics data	Screencast 3
5	Initial designs generation by ART	Notebook B
6	Bulk ICE import of bioengineered (BE) strains	Screencast 4
7	ICE export of BE strains	Screencast 5
8	Isoprenol production data generation for BE strains	Notebook C
9	Bulk EDD import and visualization of BE strains production data	Screencast 6
10	EDD export of BE production data	Screencast 7 + Notebook D
11	ART predictions and recommendations	Screencast 8 + Notebook D
12	Using the ART frontend	Screencast 9
13	Comparing ART predictions with ground truth	Notebook E

*All screencasts and notebooks are enumerated in **Tables 2**, **3**.*

TABLE 2 | Screencasts.

Screencast	Description
1	WT strain import into ICE
2	EDD import of WT multiomics data
3	EDD visualization of WT multiomics data
4	Bulk ICE import of bioengineered (BE) strains
5	ICE export of BE strains
6	Bulk EDD import and visualization of BE strains production data
7	EDD export of BE production data
8	ART predictions and recommendations
9	Using the ART frontend

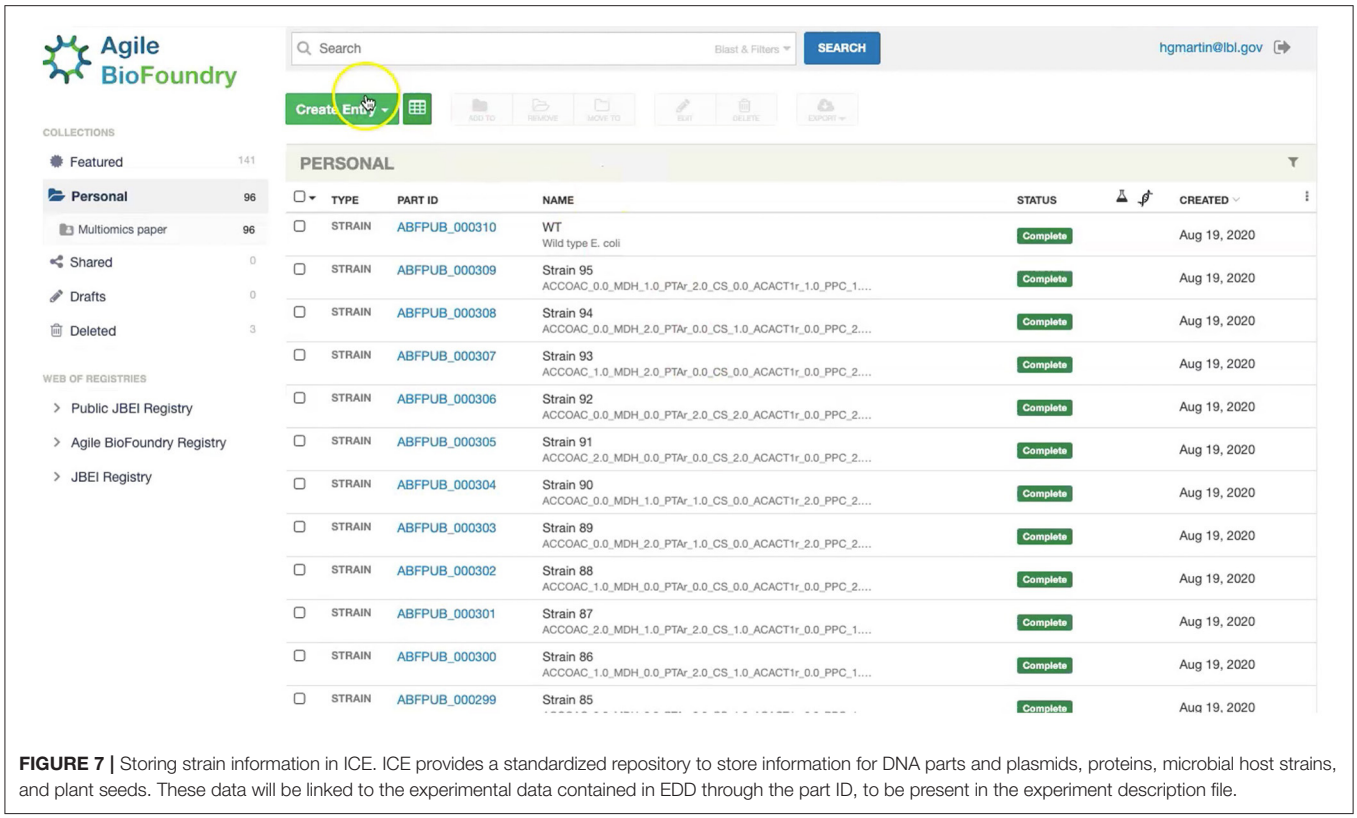


FIGURE 7 | Storing strain information in ICE. ICE provides a standardized repository to store information for DNA parts and plasmids, proteins, microbial host strains, and plant seeds. These data will be linked to the experimental data contained in EDD through the part ID, to be present in the experiment description file.

to produce (e.g., isoprenol in our example), or a transcriptomics or proteomics analysis that describes the amount of gene transcription or protein expression. Each of these measurements is obtained by applying a *protocol* (e.g., proteomics) to a given line, resulting in an *assay*. Performing an assay results in a set of *measurement data*: e.g., the number of grams of acetate per liter in the media, or the number of proteins per cell. In this ontology, assays can include one or more time points. As a general rule, destructive assays (e.g., proteomics through LC-MS) include one time point per assay, and non-destructive assays (e.g., continuous measurement of cell optical density through an optode in a fermentation platform such as a biolector) include several time points in the same assay. All this information is collected in a *study*, which is used to describe a single continuous experiment (e.g., using measuring isoprenol production for all bioengineered strains).

The first step in the data import involves creating a study and uploading an "Experiment description" file, which collects all the experimental design and metadata (see Screencast 2). The "Experiment description" file describes the strains being used through a "Part ID" number tied to a strain repository, such as ICE (**Figure 7**). This file also contains metadata relevant to the experiment (e.g., temperature, culture shaking speed, culture volume etc.). The "Experiment description" file should not include any result data. Often, the distinction between data and metadata is crystal clear, but it can be blurred in the case of concatenated experiments: e.g., the hydrolysate sugar concentration for a plant deconstructed with ionic liquid can be data for a deconstruction experiment, but metadata for an experiment focused on the fermentation of that hydrolysate through a bioengineered strain. As a rule of thumb, metadata involves the information that is known before the experiment, and data is the information that is only obtained by performing the experiment. Once the "Experiment description" file is added, you should be able to see all the experimental design data under the "Experiment description" tab in EDD.

The second step in data input involves uploading the data files for each assay. In order to do so, click on "Import data," and you will access EDD's new streamlined import. This new import emphasizes clarity and usability, and starts by prompting for the "Data category" to be uploaded (**Figure 8**). Data categories involve broad umbrellas of data types such as: transcripts, proteins, metabolites, or other data. The next choice involves the actual specific protocol used for acquiring the data. There are many types of, e.g., proteomics protocols which differ in extraction protocols, as well as the type of mass spectrometer used and its setting (chromatography column, gradient time, etc.). We encourage the use of formal protocol repositories which provide DOIs (digital object identifiers), such as protocols.io (Teytelman et al., 2016), to encourage reproducibility. Protocols can be added by the system administrator in charge of the EDD instance used. The next choice involves the type of file used to input the data. All protocols include a *generic* file type which is the simplest possibility (see "Data Formats for Generic Input Files" section). More complex file types can be easily added through scripts that map into this generic file type. The file upload completes the

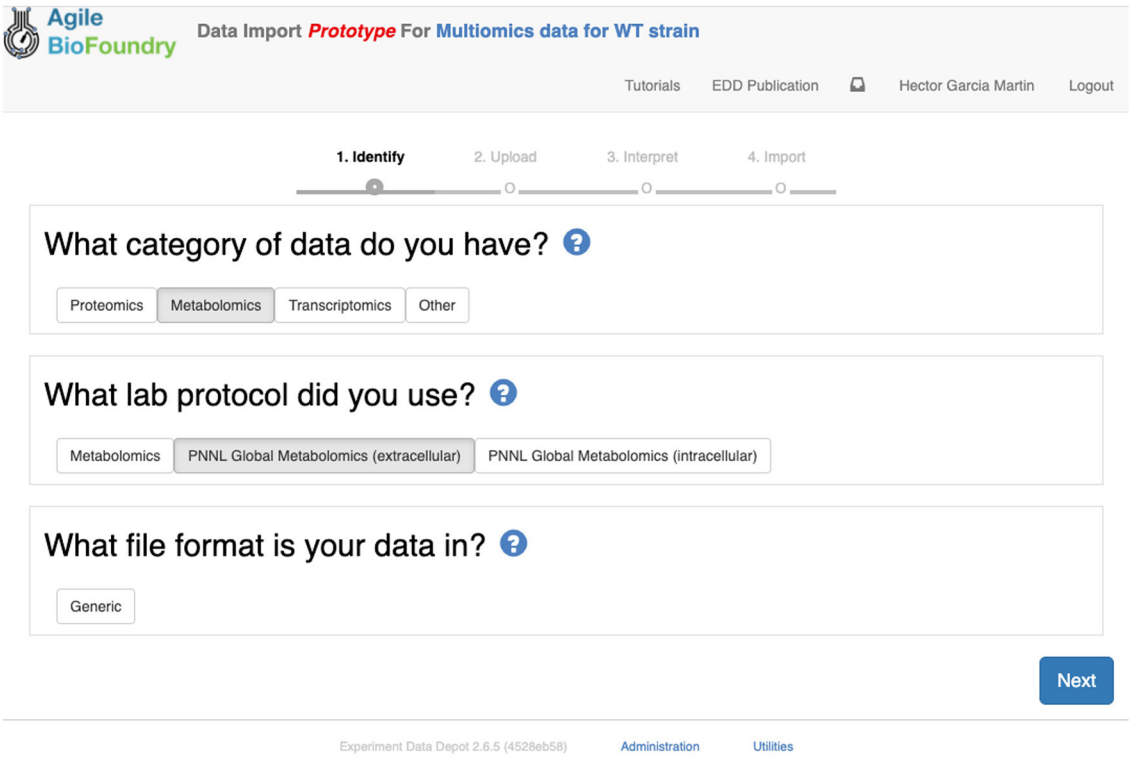


FIGURE 8 | Importing data into EDD. The new data import into EDD is divided into three parts: an initial choice of the data category, the protocol used to gather the data, and the file format used for the data. Once these are chosen, the data is uploaded for future visualization and use with, e.g., machine learning algorithms or mechanistic models.

data import, and these data can be now found under the "Data" tab in the study. Similar to entering strains in ICE, data and metadata import into EDD creates incremental documentation of the experiment in a form that can easily be published later by modifying the study's access controls.

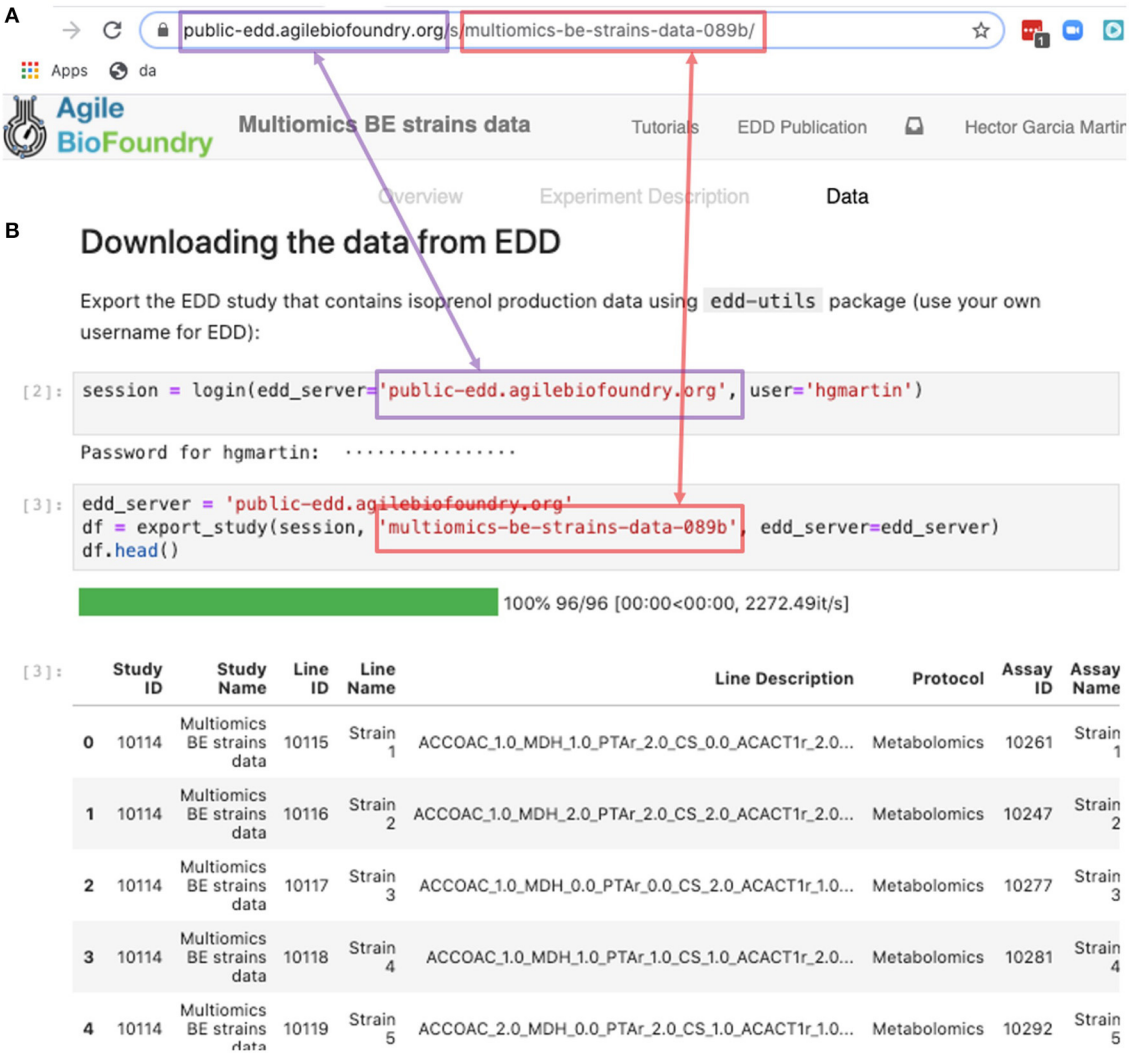
We import data into EDD twice in our metabolic engineering workflow (**Figure 2**): steps 3 and 9 in **Table 1**. In the first case (Screencast 2), we upload data created through the OMG library for the wild type (WT, Notebook A). Later (step 8) we use OMG to simulate isoprenol production data for the 95 bioengineered strains proposed by ART (Notebook C), and upload that data into EDD (step 9, Screencast 6).

Visualizing Data in EDD

EDD provides data visualization for the comparison of multiomics data sets though line and bar graphs. For example, you can easily compare the synthetic data sets created for this manuscript, which include the cell density, extracellular metabolites, transcriptomics, proteomics and metabolomics data for the base strain (**Figure 2**). Once in EDD, the data can be viewed by clicking on the "Data" tab in the corresponding EDD study (see Screencast 3, **Figure 3**). The default view is the "Line Graph" view, which displays the data as times series, with the time dimension on the *x*-axis and the measurements on the *y*-axis. Each different measurement unit is given an axis (e.g., mg/L for metabolites, proteins/cell for proteomics, FPKM for transcriptomics). The filters at the bottom of the screen allow the users to choose the data they want to concentrate on: e.g., only transcriptomics, only proteomics, only metabolites, only metabolite "octanoate," or only proteins "Galactokinase" and "Maltoporin," or only gene "b0344" and protein "Enolase." It is also possible to view the data in the form of bar graphs clustered by measurement, line, or time. The "Table" tab shows a quick summary of all data available in the study.

EDD provides basic visualization capabilities to check data quality and compare different lines and data types. Users who want a customized visualization or figure can download the data into a pandas DataFrame (pandas - Python Data Analysis Library) (McKinney, 2015) through the REST API (see next section), and use Python to leverage any of the available visualization libraries: e.g., Matplotlib (Yim et al., 2018) or Seaborn (Seaborn: Statistical Data Visualization — Seaborn 0.10.1 Documentation) (Waskom et al., 2020).

EDD visualization is demonstrated for two very different cases in two steps of our metabolic engineering workflow (**Figure 2**): steps 4 (Screencast 3) and 9 (Screencast 6). The first case involves visualizing multiomics data for a single strain (WT, **Figure 3**), whereas the second case involves the visualization of shallow data (final isoprenol production at a final point) for 96 different strains.



	Study ID	Study Name	Line ID	Line Name	Line Description	Protocol	Assay ID	Assay Name
0	10114	Multiomics BE strains data	10115	Strain 1	ACCOAC_1.0_MDH_1.0_PTAr_2.0_CS_0.0_ACACT1r_2.0...	Metabolomics	10261	Strain 1
1	10114	Multiomics BE strains data	10116	Strain 2	ACCOAC_1.0_MDH_2.0_PTAr_2.0_CS_2.0_ACACT1r_2.0...	Metabolomics	10247	Strain 2
2	10114	Multiomics BE strains data	10117	Strain 3	ACCOAC_1.0_MDH_0.0_PTAr_0.0_CS_2.0_ACACT1r_1.0...	Metabolomics	10277	Strain 3
3	10114	Multiomics BE strains data	10118	Strain 4	ACCOAC_1.0_MDH_1.0_PTAr_1.0_CS_1.0_ACACT1r_2.0...	Metabolomics	10281	Strain 4
4	10114	Multiomics BE strains data	10119	Strain 5	ACCOAC_2.0_MDH_0.0_PTAr_2.0_CS_1.0_ACACT1r_1.0...	Metabolomics	10292	Strain 5

FIGURE 9 | Exporting data from EDD into an executable Jupyter notebook for downstream processing. The EDD study web address **(A)** provides the server (magenta) and the slug (red) to export the study data in the form of a pandas DataFrame into a Jupyter notebook **(B)**. Once in a DataFrame format in a Jupyter notebook, a plethora of Python libraries are available for visualization, mechanistic modeling or machine learning.

Exporting Data From EDD

Data can be exported from EDD in two ways: a manual CSV file download, or a REpresentational State Transfer (REST) (Masse, 2011) Application Programming Interface (API). The REST API is the preferred method since it is easy to use, convenient, and flexible.

CSV export works through the Graphical User Interface (GUI) found in the "Table" tab. By selecting the desired measurements and clicking on "Export Data," the user can access a menu that provides options for layout and metadata to be included, as well as a visual example of the export. A CSV file is then generated by clicking on "Download."

The REST API provides a way to download the data in a form that can be easily integrated into a Jupyter notebook (see **Figure 9**, Screencast 7, and **Table 3**). A Jupyter notebook is a document that contains live code, equations, visualizations and explanatory text (Project Jupyter | Home[1]; Jupyter Notebooks) (Kluyver et al., 2016). The edd-utils package uses EDD's REST API to provide a DataFrame inside of your Jupyter notebook to visualize and manipulate as desired. A DataFrame is a structure from the popular library Pandas (Python ANd Data

TABLE 3 | Jupyter notebooks.

Jupyter notebook	Description
A	WT multiomics data generation
B	Initial designs generation by ART
C	Isoprenol production data generation for BE strains
D	EDD export and ART recommendations
E	Comparing ART predictions with ground truth

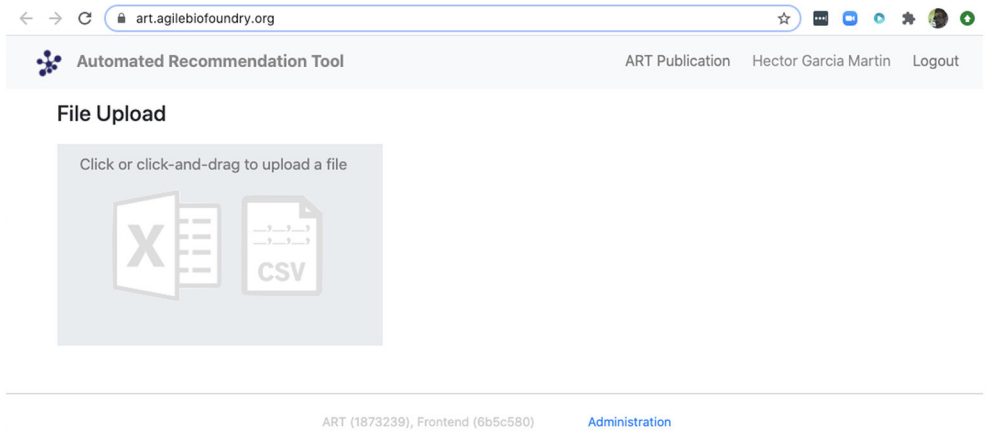


FIGURE 10 | ART also provides a frontend that does not require coding. The frontend can be found at https://art.agilebiofoundry.org/ and provides the main functionality of the ART library (**Figure 5**) in an intuitive interface. The frontend also provides a REST API that users with coding experience can leverage to use Berkeley Lab's compute resources for running ART, or to trigger ART runs automatically from other code.

AnalysiS) (pandas - Python Data Analysis Library) (McKinney, 2015), that focuses on providing tools for data analysis in python. The combination of Pandas with Jupyter notebooks provides reproducible workflows and the capability to do automated data analyses (see Screencast 8). The users can run the `export_study()` function from the edd-utils package (see code availability) to download a study from a particular EDD instance. The study is identified by its slug: the last part of the internet address corresponding to the EDD study (**Figure 9**). The EDD instance is identified by its internet address (e.g., http://public-edd.agilebiofoundry.org/ or https://public-edd.jbei.org/). Anyone with an approved account can use the EDD instances hosted at those addresses, or anyone can create their own EDD instance by downloading and installing the open source EDD software (see code availability). EDD includes access controls, e.g., for preventing dissemination of experimental data prior to publication.

EDD export is demonstrated in a single step in our workflow (**Figure 2**): the export of production data for the bioengineered strains in step 10 (Screencast 7). These data will be used to train the machine learning algorithms in the Automated Recommendation Tools, and recommend new designs.

Recommending New Designs Through ART

The data stored in EDD and ICE can be used to train machine learning methods and recommend new experiments. We will now show, for example, how to use the data we uploaded into EDD to suggest how to improve the final production of isoprenol, by using the Automated Recommendation Tool (ART) (Radivojević et al., 2020). ART is a tool that combines machine learning and Bayesian inference to provide a probabilistic predictive model of production, as well as recommendations for next steps (**Figure 5**, see technical details in Radivojević et al., 2020). In this case, we will use as input the genetic modifications on the strain (e.g., knockout ACCOAC, overexpress MDH, maintain CS, etc.) and we will try to predict final isoprenol production (see "Generating training data for machine learning and testing predictions" section).

Firstly, we will adapt the DataFrame obtained previously (step 10 in **Table 1**) to provide training data for ART. All we need to provide ART is the input (genetic modifications) and response (final isoprenol production) for each of the 96 instances. We do this by expanding the line description to include the design details into several new columns detailing the specified genetic modification for each reaction (Screencast 8, Notebook D). Next, we specify ART's parameters: input variables, response variable, number of recommendations, etc. ART uses these data and parameters to train a predictive model, which is able to recapitulate quite effectively the observed isoprenol production for the training data set (**Figure 6**). Recommendations for designs that are predicted to increase production are also provided by ART (**Figure 5**).

We can also use ART's web-based graphical frontend to produce recommendations, if we prefer not to use code (**Figure 10**). ART's frontend can be found at https://art.agilebiofoundry.org/, and requires creating an account (Screencast 9). Once the account is created and approved, we can provide the same DataFrame we created above in CSV or Excel format (Notebook D). We can then select the input variables and response variables, select the objective (maximization, minimization or reach a target value), a threshold for declaring success, and then press "Start." The job is sent to the server and an email is sent to the user once it is finished.

ART's recommendations suggest that knocking out the PPCK flux and either maintaining or overexpressing all other seven fluxes should increase production of isoprenol from 0.46 mMol to 0.57 mMol (23% increase, **Figure 5**). The machine learning model suggests that these are the best combinations based on the predictive probability. We can check this result through the OMG library, which we consider our ground truth (**Figure 2**).

Indeed, isoprenol production for this design is 0.57 mMol vs. the predicted 0.57 ± 0.02 (**Figure 6**). Hence, ART has been able to predict which combination of designs would produce a production increase. This is a non-trivial endeavor, since only 11% of designs actually improve production, according to the synthetic data provided by OMG.

CONCLUSION

In conclusion, we have shown that the combination of tools presented here (ICE + EDD + ART + OMG + Jupyter notebooks) provide a standardized manner to store data so it can be leveraged to produce actionable recommendations. We have shown how to use ICE to store strain information, EDD to store experiment data and metadata, and ART to leverage these data to suggest new experiments that improve isoprenol production. By combining these tools we have shown how to pinpoint genetic modifications that improve production of isoprenol, a potential biofuel, by 23% (from 0.46 mMol to 0.57 mMol, **Figure 5**), in a simulated data set (through OMG). The same procedures are applicable in the case of real experimental data. In sum, this set of tools provides a solution for the data deluge that bioengineering is currently experiencing, and a way to build on preexisting data to fruitfully direct future research.

CODE AVAILABILITY

All the code used in this paper can be found in the following github repositories:

- The Jupyter Notebooks associated with this paper, as well as the input files: https://github.com/AgileBioFoundry/multiomicspaper
- OMG: https://github.com/JBEI/OMG
- ICE: https://github.com/JBEI/ice
- EDD: https://github.com/JBEI/EDD
- The edd-utils package: https://github.com/JBEI/edd-utils
- ART: https://github.com/JBEI/ART

DATA AVAILABILITY STATEMENT

Data are available in the multiomics github repository. See Jupyter notebooks (**Table 3**) for exact file location. In addition, the collection for initial designs (screencast 5) is stored in a public ICE instance at https://public-registry.agilebiofoundry.org/folders/8. EDD entries for WT and BE strains (screencasts 2 and 6) are stored at https://public-edd.agilebiofoundry.org/s/multiomics-data-for-wt-strain-c450/ and https://public-edd.agilebiofoundry.org/s/multiomics-be-strains-data-089b/, respectively.

AUTHOR CONTRIBUTIONS

SR, TR, JM, and HGM created and developed the concept for the paper. SR, JM, TR, and HGM developed the OMG library. TR and HGM created the Jupyter Notebooks. HGM created the screencasts. MF and WM developed the ART frontend. MF contributed error checking and bug fix code to ART. MF, NH, TR, and HGM designed and tested the ART frontend. SR, TR, MF, WM, JK, and TB tested the screencasts and notebooks. SR, TR, MF, VJ, NH, and HGM wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was part of the DOE Agile BioFoundry (http://agilebiofoundry.org) and the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the U. S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, and the Office of Science, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. This research is also supported by the Basque Government through the BERC 2018–2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017-0718.

ACKNOWLEDGMENTS

The authors thank Zak Costello and Jacob Coble for their contributions during the initial development of the ART frontend.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.612893/full#supplementary-material

REFERENCES

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70–74. doi: 10.1126/science.1191652

Beller, H. R., Lee, T. S., and Katz, L. (2015). Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Nat. Prod. Rep.* 32, 1508–1526. doi: 10.1039/C5NP00068H

Bryksin, A. V., Brown, A. C., Baksh, M. M., Finn, M. G., and Barker, T. H. (2014). Learning from nature - novel synthetic biology approaches for biomaterial design. *Acta Biomater.* 10, 1761–1769. doi: 10.1016/j.actbio.2014.01.019

Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26, 787–793. doi: 10.1038/nbt1413

Carbonell, P., Radivojevic, T., and García Martín, H. (2019). Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synth. Biol.* 8, 1474–1477. doi: 10.1021/acssynbio.8b00540

Chen, Y., Guenther, J. M., Gin, J. W., Chan, L. J. G., Costello, Z., Ogorzalek, T. L., et al. (2019). Automated "cells-to-peptides" sample preparation workflow for high-throughput, quantitative proteomic assays of microbes. *J. Proteome Res.* 18, 3752–3761. doi: 10.1021/acs.jproteome.9b00455

Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martín, H. G. (2016). Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Syst. Biol. Appl.* 2:16009. doi: 10.1038/npjsba.2016.9

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096. doi: 10.1126/science.1258096

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Fuhrer, T., and Zamboni, N. (2015). High-throughput discovery metabolomics. *Curr. Opin. Biotechnol.* 31, 73–78. doi: 10.1016/j.copbio.2014.08.006

Gardner, T. S. (2013). Synthetic biology: from hype to impact. *Trends Biotechnol.* 31, 123–125. doi: 10.1016/j.tibtech.2013.01.018

Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., and Keasling, J. D. (2012). Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* 40:e141. doi: 10.1093/nar/gks531

Heinrich, R., and Schuster, S. (1996). *The Regulation of Cellular Systems*. Boston, MA: Springer US.

Hodgman, C. E., and Jewett, M. C. (2012). Cell-free synthetic biology: thinking outside the cell. *Metab. Eng.* 14, 261–269. doi: 10.1016/j.ymben.2011.09.002

Kang, A., Mendez-Perez, D., Goh, E.-B., Baidoo, E. E. K., Benites, V. T., Beller, H. R., et al. (2019). Optimization of the IPP-bypass mevalonate pathway and fed-batch fermentation for the production of isoprenol in *Escherichia coli*. *Metab. Eng.* 56, 85–96. doi: 10.1016/j.ymben.2019.09.003

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). *Jupyter Notebooks - A Publishing Format for Reproducible Computational Workflows*. IOS Press. Available online at: http://ebooks.iospress.nl/publication/42900 (accessed August 6, 2020).

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., et al. (2018). A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* 36, 1062–1066. doi: 10.1038/nbt.4245

Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305. doi: 10.1038/nrmicro2737

Lohr, S. (2014). For big-data scientists, 'janitor work' is key hurdle to insights. *New York Times* 17:B4.

Ma, S., Tang, N., and Tian, J. (2012). DNA synthesis, assembly and applications in synthetic biology. *Curr. Opin. Chem. Biol.* 16, 260–267. doi: 10.1016/j.cbpa.2012.05.001

Masse, M. (2011). *REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces*. Cambridge: O'Reilly Media, Inc.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245. doi: 10.1080/00401706.1979.10489755

McKinney, W. (2015). *Pandas, Python Data Analysis Library*. Available online at: http://pandas.pydata.org (accessed September 3, 2020).

Meat-free outsells beef (2019). Meat-free outsells beef. *Nat. Biotechnol.* 37, 1250–1250. doi: 10.1038/s41587-019-0313-x

Morrell, W. C., Birkel, G. W., Forrer, M., Lopez, T., Backman, T. W. H., Dussault, M., et al. (2017). The experiment data depot: a web-based software tool for biological experimental data storage, sharing, and visualization. *ACS Synth. Biol.* 6, 2248–2259. doi: 10.1021/acssynbio.7b00204

Müller, K. M., and Arndt, K. M. (2012). Standardization in synthetic biology. *Methods Mol. Biol.* 813, 23–43. doi: 10.1007/978-1-61779-412-4_2

Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614

Paddon, C. J., and Keasling, J. D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* 12, 355–367. doi: 10.1038/nrmicro3240

Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328. doi: 10.1038/nature11478

Petzold, C. J., Chan, L. J. G., Nhan, M., and Adams, P. D. (2015). Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* 3:135. doi: 10.3389/fbioe.2015.00135

Radivojević, T., Costello, Z., Workman, K., and Garcia Martin, H. (2020). A machine learning Automated Recommendation Tool for synthetic biology. *Nat. Commun.* 11:4879. doi: 10.1038/s41467-020-18008-4

Roell, M.-S., and Zurbriggen, M. D. (2020). The impact of synthetic biology for future agriculture and nutrition. *Curr. Opin. Biotechnol.* 61, 102–109. doi: 10.1016/j.copbio.2019.10.004

Segrè, 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

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., et al. (2015). Big data: astronomical or genomical? *PLoS Biol.* 13:e1002195. doi: 10.1371/journal.pbio.1002195

Teytelman, L., Stoliartchouk, A., Kindler, L., and Hurwitz, B. L. (2016). Protocols.io: virtual communities for protocol development and discussion. *PLoS Biol.* 14:e1002538. doi: 10.1371/journal.pbio.1002538

Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi: 10.1038/nprot.2009.203

Waskom, M., Gelbart, M., Botvinnik, O., Ostblom, J., Hobson, P., Lukauskas, S., et al. (2020). *mwaskom/seaborn: v0.11.1 (December 2020)* [Computer software]. Zenodo. doi: 10.5281/ZENODO.592845

Yim, A., Chung, C., and Yu, A. (2018). *Matplotlib for Python Developers: Effective Techniques for Data Visualization With Python*. Birmingham, AL: Packt Publishing Ltd.

Zhang, J., Petersen, S. D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., et al. (2020). Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. *Nat. Commun.* 11:4880. doi: 10.1038/s41467-020-17910-1

Conflict of Interest: NH declares financial interests in TeselaGen Biotechnologies, and Ansa Biotechnologies.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

REVIEW
published: 23 February 2021
doi: 10.3389/fbioe.2021.613307

Omics-Driven Biotechnology for Industrial Applications

Bashar Amer[1,2] *and Edward E. K. Baidoo*[1,2,3]*

[1] Lawrence Berkeley National Laboratory, Joint BioEnergy Institute, Emeryville, CA, United States, [2] Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [3] U.S. Department of Energy, Agile BioFoundry, Emeryville, CA, United States

Biomanufacturing is a key component of biotechnology that uses biological systems to produce bioproducts of commercial relevance, which are of great interest to the energy, material, pharmaceutical, food, and agriculture industries. Biotechnology-based approaches, such as synthetic biology and metabolic engineering are heavily reliant on "omics" driven systems biology to characterize and understand metabolic networks. Knowledge gained from systems biology experiments aid the development of synthetic biology tools and the advancement of metabolic engineering studies toward establishing robust industrial biomanufacturing platforms. In this review, we discuss recent advances in "omics" technologies, compare the pros and cons of the different "omics" technologies, and discuss the necessary requirements for carrying out multi-omics experiments. We highlight the influence of "omics" technologies on the production of biofuels and bioproducts by metabolic engineering. Finally, we discuss the application of "omics" technologies to agricultural and food biotechnology, and review the impact of "omics" on current COVID-19 research.

Keywords: systems biology, genomics, transcriptomics, metabolomics, proteomics, multi-omics, metabolic engineering, biotechnology

OPEN ACCESS

Edited by:
Eduard Kerkhoven, Chalmers University of Technology, Sweden

Reviewed by:
Hyun Uk Kim, Korea Advanced Institute of Science and Technology, South Korea
Adam M. Feist, University of California, San Diego, United States

***Correspondence:**
Edward E. K. Baidoo
eebaidoo@lbl.gov

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *01 October 2020*
Accepted: *11 January 2021*
Published: *23 February 2021*

Citation:
Amer B and Baidoo EEK (2021) Omics-Driven Biotechnology for Industrial Applications. Front. Bioeng. Biotechnol. 9:613307. doi: 10.3389/fbioe.2021.613307

INTRODUCTION

Biotechnology employs biological processes, organisms, or systems to yield products and technologies that are improving human lives (Bhatia, 2018). The use of biological systems to manufacture bioproducts of commercial relevance (i.e., biomanufacturing) is a key component of the biotechnology industry. This manufacturing approach is used by the energy, material, pharmaceutical, food, agriculture, and cosmetic industries (Zhang et al., 2017). The bioproducts made from biomanufacturing processes are typically metabolites and proteins, which can be obtained from cells, tissues, and organs. The biological systems producing these bioproducts can be natural or modified by genetic engineering (Zhang et al., 2017), metabolic engineering (optimizing metabolic networks and pathways for increased production of metabolites and/or by-products), synthetic biology (applying engineering principles to the chemical design of biological systems), and protein engineering (optimizing protein design to develop valuable proteins) (Zhang et al., 2017).

Modern biotechnology-based biomanufacturing started in the early twentieth century with the production of short-chain alcohols and ketones, amino acids, organic acids, and vitamin C by microbial mono-culture fermentation (Zhang et al., 2017). This was followed by the production of antibiotics via microbial fermentation in the 1940s and protein drug production in animal cells and microorganisms in the 1980s (Zhang et al., 2017). With the advent of translational research

(e.g., stem cell research) in the 2000s, the bioproducts can now be engineered cells, tissues, and organs (e.g., by stem cell engineering) (Roh et al., 2016; Zhang et al., 2017). In addition to this, advancements in synthetic biology and metabolic and protein engineering have been applied to renewable energy research in the development of advanced biofuel and hydrogen production by engineered microorganisms (Zhang et al., 2017). Research efforts are underway at bioenergy research centers (e.g., US DOE Bioenergy Research Centers) to make biofuels more affordable by coproducing them with renewable bioproducts. This is of great importance, as environmental, geopolitical, and economic factors are reshaping our view of global energy and manufacturing demands (Clomburg et al., 2017). The tools (and ideology) from these approaches have also been leveraged by the food industry to produce artificial food products via synthetic biocatalysts in a sustainable way (Zhang et al., 2017).

Synthetic biology and metabolic engineering can benefit from systems biology approaches, which in turn rely on "omics" technologies to characterize and understand metabolic networks. The considerable amount of knowledge obtained from omics-driven systems biology experiments can be used in the development of synthetic biology tools and the advancement of metabolic engineering. This facilitates the manipulation of complex biological systems toward establishing robust industrial biomanufacturing platforms (Baidoo and Teixeira Benites, 2019).

In this review, we aim to examine the influence of "omics" technologies on biotechnology research. "Omics" techniques were compared to understand their relevance and applicability to biotechnology research, especially in the context of microbial biotechnology, with the aim of facilitating the experimental design of individual "omics" and multi-omics studies. Finally, we compared the trends in "omics" utilization during the last two decades to determine their progression in biotechnology research.

A COMPARISON OF THE MAJOR "OMICS" TECHNOLOGIES

Biological engineering requires the accurate prediction of phenotype from genotype. Thus, testing and validating modified and synthesized genomes (i.e, genomics) as well as the study of the transcriptome (the complete set of RNA transcripts), which is expressed from the genome (i.e., transcriptomics), are crucial to evaluating genome engineering. Proteomics and metabolomics have also gained a lot of attention due to their provision of metabolic information pertaining to both function and phenotype (Baidoo, 2019). Forty years ago, scientists realized that the flow of biochemical information in biological systems is not unidirectional from the genome to metabolome, but rather a set of interactions between the "omes" (Roberts et al., 2012). Therefore, a multi-omics approach is necessary for the elucidation of chemical structure, function, development, adaptation, and evolution of biological systems for deeper understanding of the principles of life (Baidoo, 2019) (**Figure 1**).

In comparison to metabolites and proteins, genes are less chemically heterogeneous. Each gene is made up of DNA that is composed of only four basic nucleotides (i.e., guanine, adenine, cytosine and thymine), whereas each protein is

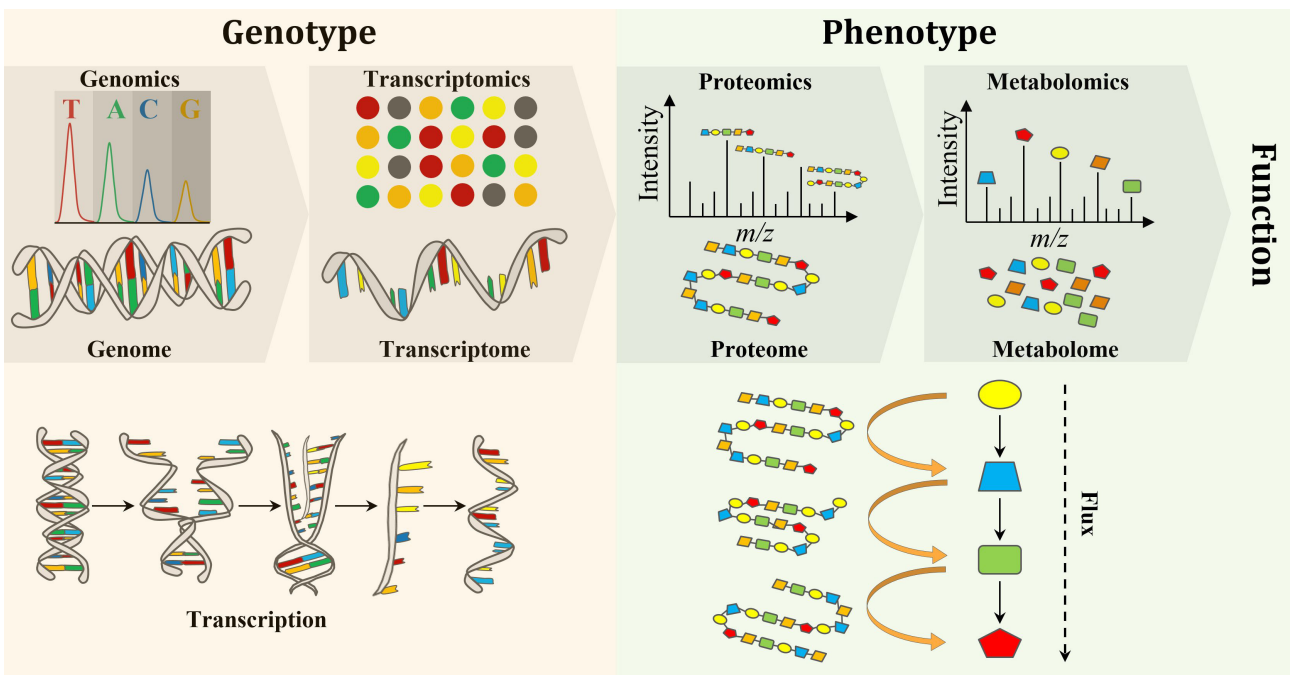


FIGURE 1 | An overview of the flow of molecular information from genes to metabolites to function and phenotype, and the interactions between the "omes" and the "omics" techniques used to measure them.

composed of a mixture of 32 amino acids, while metabolites are much more variable in their chemical structures (Wang et al., 2010). Therefore, it is analytically less challenging to perform genomics and transcriptomics, when compared to proteomics and metabolomics (Aizat et al., 2018). Consequently, genomics and transcriptomics provide the most comprehensive and robust platforms for biotechnology applications. Over the past few decades, research has shown that genomics and transcriptomics cannot solely provide a complete description of complex biological systems as genetic information can produce more questions than answers. For instance, genomics can describe genes and their interactions (measure genotype) but cannot explain phenotypes. Thus, the attention is turned to the utilization of other "omics" techniques, such as proteomics and metabolomics, which can bridge the gap between genetic potential and final phenotype to facilitate a greater understanding of biological systems (Smith and Figeys, 2006; Wilmes et al., 2015). While transcriptomics (transcription) and proteomics (translation) provide information on gene expression, the latter directly links genotype to phenotype. In addition to providing phenotypic information, the metabolome provides an instant response to genetic and/or environmental perturbations and, therefore, provides a snap shot of the actual metabolic and physiological state of a cell (Tang, 2011). However, metabolomics alone is not able to measure changes at the gene level and correlate them with the observable properties of organisms, the phenotypes, which are produced by the genotype in the first place (Fiehn, 2001). Therefore, a comprehensive understanding of an organism on a molecular level requires the integration of "omics" data in order to discover new molecules and pathways (Wang et al., 2010) (**Figure 1**). Integration of "omics" data helps to assess the flow of information from one "omics" level to the other and, therefore, links genotype to phenotype (Subramanian et al., 2020). Furthermore, the combination of "omics" techniques is important to address open biological questions (i.e., data driven research) that accelerate our understanding of the system as a whole and boost the use of systems metabolic engineering tools in industrial settings (Zhao et al., 2020).

Genomics and Transcriptomics

The construction of predictable and preferred phenotypes is crucial in synthetic biology; therefore, tight and tunable control of gene expression is highly desirable. Biological engineering, moreover, is greatly benefiting from the recent innovations in genomics and genome editing technologies, which offer advanced tools to re-engineer naturally evolved systems and to build new systems as well. In addition, advances in *de novo* synthesis and *in vivo* gene targeting enable efficient testing of model-driven hypotheses. Furthermore, genomics allows the high-throughput DNA sequencing and large-scale bimolecular modeling of metabolic and signaling networks in natural and engineered strains (Pagani et al., 2012).

Genomics and Transcriptomics Analysis

One of the challenges facing traditional genomics (and other "omics") analyses is that not all microorganisms can be cultured in a laboratory setting. Furthermore, isolated strains might behave differently in culture than in their natural environments. Therefore, there was an urgent need to develop cultivation independent methods to study microbial communities (VerBerkmoes et al., 2009). Metatranscriptomics can reveal the diversity of active genes within microbial communities (e.g., 16S rRNA sequencing for reconstructing phylogenies) (Bashiardes et al., 2016).

Metagenomics studies the structure and function of genetic material in complex samples of multi-organisms as well as of entire microbial communities without a cultivation step and can offer a solution for such challenges and facilitate the discovery of novel genes, enzymes, and metabolic pathways. Metagenomics analyses are classified as sequence-based and function-based screening, which are used to discover and identify, respectively, novel natural genes and compounds from environmental samples (Chistoserdova, 2010; Gilbert and Heiner, 2015; Kumar Awasthi et al., 2020). For example, metagenomics is actively used in agricultural research to understand the microbial communities in the soil system (Durot et al., 2009), to examine various microbes that can stimulate the cycling of macro- and micro-nutrients, and the release of essential enzymes, which enhance crop production (Cupples, 2005).

Nanopore sequencers are massively parallel sequencing technologies. Oxford Nanopore Technologies (ONT), in particular, provides a single molecule sequencer using a protein nanopore that realizes direct sequencing without DNA synthesis or amplification (Brown and Clarke, 2016; Roumpeka et al., 2017). The ONT sequencer can determine DNA/RNA modifications and sequence an ultra-long read limited by the input nucleotide length (Kono and Arakawa, 2019). However, ONT reads require polishing and great care needs to be taken when contigs are polished individually to avoid the removal of true, natural sequence diversity due to cross mapping of reads in repeat regions. Therefore, it was found that it is crucial to apply long-range information technologies (e.g., 10x genomics, Hi-C, synthetic long reads) and to develop new algorithms to simplify the extensive assembly and polishing workflow (Somerville et al., 2018).

Sort-seq is a single-cell sequencing platform, which combines flow cytometry, binned fluorescence-activated cell sorting (FACS), next-generation sequencing (NGS), and statistical inference to quantify the dynamic range of many biosensor variants in parallel (Rohlhill et al., 2017; Batrakou et al., 2020; Koberstein, 2020). FACS, which enables the sorting of single cells, allows the enrichment of specific cells to generate high-resolution gene expression and transcriptional maps (Kambale et al., 2019). NGS and RNA sequencing (RNA-seq) technologies enable the large-scale DNA and RNA sequencing of the entire genome and transcriptome, respectively, providing an unbiased and comprehensive view of biological systems towards understanding genomic function (Frese et al., 2013; Alfaro et al., 2019; Stark et al., 2019). Examples of NGS platforms include the Illumina HiSeq, Genome Analyzer Systems, 454 Genome Sequencer FLX Titanium System, the Helicos HeliScope, the SOLiD sequencing platform and the Ion Torrent Sequencing platform. In addition, there are other techniques used to measure

the interaction between proteins and DNA, such as chromatin immunoprecipitation (ChIP). ChIP followed by NGS sequencing (ChIP-seq) has high potential for detailing the binding sites of various transcription factors and assaying protein–DNA interaction at a whole-genome level (Roukos, 2012).

Bar-seq (barcode analysis by sequencing) is another high-throughput quantitative DNA sequencing technique that enables the parallel phenotyping of pools of thousands of mutants and monitoring thousands of gene-chemical interactions (Smith et al., 2010; Robinson et al., 2014). Techniques, such as bar-seq can lower the complexity of data obtained from a large number of sequence reads, thus, making NGS more efficient and affordable (Smith et al., 2009).

New computational tools have enabled researchers to perform fast and accurate analysis of big genomics data. Extracted genomic information has been used to model metabolic processes and signaling networks across the entire cell, generating many new testable hypotheses (Lewis et al., 2012; Esvelt and Wang, 2013). Due to the robustness of genomic measurements, there are numerous genomics databases and data analysis tools available (Roumpeka et al., 2017).

Proteomics

Proteomics focuses on the analysis of proteins and peptides produced by cells at different stages of development and life cycle and in biological systems under a given growth condition. Proteomics is also used to elucidate the temporal dynamics of protein expression levels or post-translational modification (PTM) (VerBerkmoes et al., 2009).

Proteomics Sample Preparation

High biological sample diversity and complexity and the dynamic range of protein levels present in such samples are the main challenges that proteomics encounter. These factors, in addition to the large number of proteins, complicate the analysis of low abundance proteins. The development of automated sample preparation workflows are becoming more common for high-throughput, quantitative proteomic assays of microbes. One automated workflow was able to quantify >600 peptides with a 15.8% median coefficient of variation, demonstrating the robustness of this approach (Chen et al., 2019). Another high-throughput automatable workflow was developed to increase the yield of lysis of several representative bacterial and eukaryotic microorganisms via vigorous bead-beating with silica and glass beads in presence of detergents (Hayoun et al., 2019). Interestingly, a universal, high-throughput and a detergent-free sample preparation protocol was developed this year for peptide generation from various microbes [i.e., *Escherichia coli* (*E. coli*), *Staphylococcus aureus* and *Bacillus cereus*]. The protocol holds the potential to dramatically simplify and standardize sample preparation while improving the depth of proteome coverage especially for challenging samples (Doellinger et al., 2020).

Proteomics Data Acquisition

Protein Identification and Structural Elucidation

Most proteomics workflows are based on a bottom-up approach, where protein is extracted, digested (e.g., trypsin digestion) into proteolytic peptides, then analyzed via MS (Kleiner et al., 2017). When liquid-chromatography is coupled to mass spectrometry (LC-MS), both qualitative and quantitative data analysis of proteins are improved. Moreover, the application of multidimensional LC separation prior to MS protein analysis provides enhanced MS sensitivity by reducing sample complexity and increasing the number of chromatographic peaks that can be resolved in a single analytical run (Hinzke et al., 2019; Duong et al., 2020). Targeted proteomics via LC-tandem MS (LC-MS/MS) is a commonly used MS method, where the analysis focuses on a subset of biological proteins of interest (Marx, 2013). When a bottom-up approach is applied to all proteins within a given biological system, it is called shotgun (untargeted) proteomics (Wolters et al., 2001; Nesvizhskii and Aebersold, 2005). Top-down proteomics, conversely, is based on the analysis (via LC-MS or LC-MS/MS) of intact proteins, and thus, provides unique information about the molecular structure of proteins (e.g., PTM) (Catherman et al., 2014). However, it is not always possible to separate intact proteins, especially large proteins, prior to MS analysis in a top-down approach. Besides that, top-down is less sensitive and has a lower throughput than a bottom-up approach (Catherman et al., 2014).

Accurate determination of protein structure helps to define their roles and functions in biological systems. However, many folded proteins have complex structures, which complicates their structural elucidation (Yates, 2019). Therefore, cryogenic electron microscopy and ion-mobility-MS are used to determine the structures of such proteins such proteins (Yates, 2019). Moreover, a combination of MALDI, high resolution MS (i.e., orbitrap and ion trap MS) and a UV–Vis-based reduction assay is used to elucidate peptide modification via the analysis of specific fragmentation of synthesized peptides, which might have inhibitory effects on various diseases (Rühl et al., 2019).

The identification of PTM peptides can be difficult in the case of labile modifications (e.g., phosphorylation and S-nitrosylation) that might break down during MS/MS fragmentation. Such modifications require soft fragmentation and high-resolution methods to identify and determine the location of a PTM. Electron-transfer dissociation is considered to be the favorable choice for the identification of liable PTM as it transfers electrons to multi-protonated proteins or peptides, which leads to N-Cα backbone bond cleavage (Chen et al., 2017).

Metaproteomics is the large-scale characterization of the entire protein complement of environmental microbiota at a given point in time to determine structure (Wilmes and Bond, 2004; Kleiner et al., 2017), metabolism and physiology of community components (Kleiner et al., 2012). The recent advancement in LC and high-resolution MS have enabled the identification and quantification of more than 10,000 peptides and proteins per sample in metaproteomics (Kleiner, 2019). Metaproteomics can also measure interactions between community components (Hamann et al., 2016) and assess substrate consumption (Bryson et al., 2016; Kleiner et al., 2018).

Protein Quantification

Besides identification, MS-based technologies became the tools of choice for the quantification of proteins in an organism

(Karpievitch et al., 2010). Stable isotope labeling is one approach that can be used to quantify proteins by measuring the relative abundance of labeled protein to non-labeled protein (VerBerkmoes et al., 2009). However, the variation in ionization efficiency among peptides and proteins and the low recovery of some peptides (e.g., hydrophobic peptides adhere to surfaces) can affect the accuracy of their direct quantification. Recent advances in MS acquisition rate, detection, and resolution have addressed much of the sensitivity concerns of MS-based quantification for proteomics (Iwamoto and Shimada, 2018). MS sensitivity was further enhanced with the application of micro-flow (Krisp et al., 2015; Bian et al., 2020) and nano-flow (Wilson et al., 2015) LC-MS. Another major advancement for global protein quantification was the introduction of isobaric tags or multiplexed proteomics, which in a single experiment enables the quantification of proteins across multiple samples (Pappireddi et al., 2019). Tandem-mass-tags are examples of commonly used isobaric tags for instance in human cerebrospinal fluids (Dayon et al., 2008).

Proteomics Data Analysis

Proteomics data analysis tools are generally used for protein identification (via bioinformatics) and quantification, and bioinformatics techniques tools used to process the proteomics data. A few examples of data analysis tools that are used for the identification of peptides and proteins include Mascot (Eng et al., 1994), Swiss-Prot (Bairoch and Boeckmann, 1994), Sequest (Perkins et al., 1999), Tandem (Craig and Beavis, 2004), Skyline (MacLean et al., 2010), Uni-Prot,[1] UniNovo (Jeong et al., 2013), and SWPepNovo (Li et al., 2019). Such algorithm-based software were developed to match the MS collected data from peptide/protein analysis to their base peptides/proteins and with *in silico* predicted intact masses and fragmentation patterns (Urgen Cox and Mann, 2011). Moreover, they determine both the mass and exact location of any possible modifications (Hansen et al., 2001; Savitski et al., 2006). Common bioinformatics techniques tools for proteomics data analysis include CRONOS (Waegele et al., 2009), COVAIN (Sun and Weckwerth, 2012), SIGNOR (Perfetto et al., 2016), KEGG (Kanehisa et al., 2017), and STRING v11 (Szklarczyk et al., 2019).

Metabolomics

Metabolomics, which is the measurement of small molecule substrates, intermediates, and/or end products of cellular metabolism (i.e. metabolites), provides an immediate and dynamic response to genetic and/or environmental perturbations in a biological system (Fiehn, 2002; Ellis and Goodacre, 2012; Zhao et al., 2020). Targeted and untargeted metabolomics are used to quantify a group of defined metabolites and determine all measurable metabolites in a biological sample, respectively (Scalbert et al., 2009). MS-based metabolomics, like proteomics, normally employs separation [e.g. LC and gas chromatography (GC)] or capillary electrophoresis (CE) before MS detection (Fiehn, 2002). Whereas, MALDI-MS conducts high-throughput screening without separation.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical technique for high-throughput metabolic fingerprinting and provides more reliable metabolite structure (e.g., via 2D NMR) identification than MS (Giraudeau, 2020). However, although NMR offers unambiguous structure determination of unknown metabolites via ^{1}H- and ^{13}C-NMR, MS-based methods comprise widely accessed metabolomics techniques due to higher sensitivity and lower instrumentation cost (Chatham and Blackband, 2001). Furthermore, NMR is semi-quantitative whereas MS is quantitative, thus, NMR and MS are highly complementary techniques. In addition, the diverse physiochemical properties (e.g., solubility, reactivity, stability, and polarity) of the metabolome limits our ability to analyze all metabolites from a biological system with a single or even a limited-set of analytical techniques (Fiehn, 2002). Therefore, multiple methods are used for comprehensive metabolome characterization.

Metabolomics Sample Preparation

Metabolites are constantly going under reformation and transformation in biochemical reactions within a cell and/or being thermally degraded (and in some cases oxidized) at ambient conditions (Scalbert et al., 2009). Therefore, quick and efficient metabolic quenching protocols are required to accurately quantify metabolic information. Not surprisingly, researchers tend to develop metabolic quenching methods in conjunction to and metabolite extraction protocols. Doran et al. (2017), for example, proposed an acidic-based metabolic quenching to aqueous-alcohol metabolite extraction. This protocol yielded low metabolite leakage and high extraction recovery in *Acidithiobacillus ferrooxidans*. Complex biological sample matrices can also suppress metabolite MS detection. Thus, clean up strategies, such as solid phase extraction (SPE) and solid phase micro-extraction (SPME) can reduce the complexity of sample matrices prior to LC-MS and GC-MS analysis, thereby increasing the quantitative capability of metabolomics methods (Yang et al., 2011). The last 5 years witnessed the development of high-throughput 96-well plate SPE (Li et al., 2015) and 96-well automated SPME (Mousavi et al., 2015) for the simultaneous extraction of metabolites and lipids from biological samples.

In addition, robotics and microfluidics tools can be applied to high-throughput synthetic biology applications by automating cell preparation and metabolite extraction to increase coverage (Yizhak et al., 2010; Koh et al., 2018; Vavricka et al., 2020). Automated liquid handler technologies, therefore, are important for high-throughput sample preparation as they ensure good quality and reproducibility of sample extraction and processing for unbiased measurement of metabolic differences (e.g., based on disease states or interventions stimuli) (Liu et al., 2019).

Metabolomics Data Acquisition

The development of nanoelectrospray-ionization and direct infusion nanoelectrospray high-resolution MS have led to a considerable increase in the dynamic range and detection sensitivity of metabolites from tissues and biofluids in

[1] *UniProt*. Available online at: https://www.uniprot.org/ (accessed September 30, 2020).

human studies (Chekmeneva et al., 2017; Southam et al., 2017). Generally, nanospray technology is more sensitive than electrospray, but suffer from low robustness. However, nanoelectrospray employs low level of nebulization and flow rate to achieve high sensitivity without compromising robustness (Guo et al., 2016). Besides, the application of ion mobility and high resolution MS has improved the identification of isomers, thereby enabling a more accurate assessment of their biological roles (Ren et al., 2018; Rathahao-Paris et al., 2019). Moreover, new developments in orbitrap MS systems have improved metabolites annotation and coverage in GC- and LC-MS studies (Simirgiotis et al., 2017; Misra et al., 2018; Manier et al., 2019; Stettin et al., 2020).

Although GC-MS requires more sample preparation steps when derivatizing hydrophilic non-volatile metabolites, it is more robust than LC-MS. Moreover, method development is easier for GC-MS than for LC-MS. GC-MS also achieves better identification of untargeted metabolites due to standardized ionization conditions, which makes it possible to set up a universal compound identification library/database, such as NIST. While CE achieves the highest separation efficiency, CE-MS is the least robust and sensitive of the three separation techniques.

Real-time metabolomics enables the simultaneous and high-throughput analysis of microbial metabolites without the need for time-consuming sample preparation steps (Link et al., 2015; Boguszewicz et al., 2019; Nguyen et al., 2020). However, the lack of chromatographic or electrophoretic separation in this approach reduces the quantitative capability of this technique (Baidoo and Teixeira Benites, 2019). While MALDI can be used for high-throughput metabolite screening, MALDI imaging MS has emerged as a powerful tool for analyzing tissue specimen in an unprecedented detail. MALDI imaging MS has made significant contributions to the understanding of the biology of disease and its perspectives for pathology research and practice, as well as in pharmaceutical studies (Aichler and Walch, 2015; Mahajan and Ontaneda, 2017; Schulz et al., 2019).

Metabolomics technologies are regularly applied to metabolic flux analysis (MFA, i.e., ^{13}C) studies (Baidoo and Teixeira Benites, 2019). MFA determines the rates of *in vivo* metabolic reactions. Thus, enabling an understanding of carbon and energy flow throughout the metabolic network in a cell. Overall, MFA accelerates the discovery of novel metabolic pathways and enzymes for improved synthetic bioproduction (Feng et al., 2010; Ando and García Martín, 2019; Babele and Young, 2019; Vavricka et al., 2020). However, the availability and high cost of stable isotope compounds can limit MFA capability (Gonzalez and Pierron, 2015).

Metabolomics Data Analysis

Multivariate data analysis methods, such as principal component analysis (PCA) and partial least squares (PLS) analysis are used to analyze large quantities of metabolic profiling data (i.e., reveal clustering-based on features). In addition, there is a need for advanced pathway analysis tools to interpret metabolomics data to solve some of the most challenging biological paradoxes and reveal optimal conditions for biological systems. Such techniques enable systems biology researchers to utilize metabolomics data as a resource that contributes to an iterative cycle of hypothesis generating and hypothesis testing phases (Kell, 2004; Vavricka et al., 2020). To address all of this, more attention is being paid to the area of big data and machine learning. Thus, the state-of-the-art understanding of cell metabolism can be improved and further combined with mechanistic models to automate synthetic biology and intelligent biomanufacturing (Oyetunde et al., 2018). To this end, recent advancements in metabolomics tools for data analysis, storing and sharing have been developed [e.g., WebSpecmine (Cardoso et al., 2019), SIRIUS 4 (Dührkop et al., 2019), MetaboAnalyst 4.0 (Chong et al., 2018), and SECIM (Kirpich et al., 2018)]. Knowledge of biology (e.g., regulation, metabolism, physiology, etc.) is still, however, necessary for efficient experimental design and accurate data interpretation in order to understand and accurately characterize biological systems.

Multi-Omics for Systems Biology

The recent advancement in omics technologies has improved the analysis efficiency by reducing cost and time, but also by collecting informative and meaningful multi-omics data. Thus, facilitating the implementation of multi-omics techniques in systems biology studies. However, integrating multi-omics platforms is still an ongoing challenge due to their inherent data differences (Saito and Matsuda, 2010; Yizhak et al., 2010; Brunk et al., 2016; Koh et al., 2018; Pinu et al., 2019; Vavricka et al., 2020). For example, genomics data are qualitative, accurate and reproducible, while other "omics" data, such as proteomics and metabolomics are both qualitative and quantitative, not as reproducible, and noisy (Kuo et al., 2002; MacLean et al., 2010; Guo et al., 2013; Gross et al., 2018). Further, multi-omics data is normally pre-treated by various data treatment methods (e.g., deconvolution, normalization, scaling, and transformation) and software before being integrated. Multi-omics studies also require experts in their respective "omics" fields (as well as IT support) to validate multi-omics data. While this provides greater data interpretation accuracy it does, however, complicate data acquisition and analysis.

Recently, Pinu et al. discussed some recommendations to overcome the major challenge facing the implementation of multi-omics techniques in systems biology, which is the differences among their inherent data. The aim of their recommendations is to make researchers aware of the importance of having a proper experimental design in the first place. Thus, the appropriate biological samples should be carefully selected, prepared, and stored before planning any "omics" study. Afterward, researchers should carefully collect quantitative multi-omics data and associated meta-data and select better tools for integration and interpretation of the data. Finally, develop new resources for the deposition of intact multi-omics data sets (Pinu et al., 2019). It is also necessary to select or develop methods that keep the optimum balance between high recovery and low degradation of extracted biological features.

As scientists are becoming more aware of the importance of multi-omics analysis, a number of tools, databases, and

methods are being developed for the aim of integrating multi-omics data sets. These tools perform advanced statistics (e.g., multivariate data analysis) and data illustration (e.g., correlation maps). Examples of databases used for multi-omics analysis include ECMDB 2.0 (Sajed et al., 2016), *Saccharomyces* Genome Database (MacPherson et al., 2017), YMDB 2.0 (Ramirez-Gaona et al., 2017), GenBank (Benson et al., 2013), KEGG (Kanehisa and Subramaniam, 2002), and many others. A recent review by Subramaniam et al. showed that common multi-omics data integration and interpretation tools were able to derive new insights from data, conduct disease subtyping, and obtain diagnostic biomarker prediction (Subramanian et al., 2020).

Table 1 provides a comprehensive comparison of the major "omics" technologies. The aim of this comparison is to facilitate the experimental design of individual "omics" and multi-omics studies by highlighting the general characteristics of each technology.

OMICS-GUIDED BIOTECHNOLOGY

"Omics" technologies are becoming increasingly involved in the development of biotechnological processes for the production of many substantial products. The use of "omics" technologies to characterize and understand biological systems has enabled researchers to select and predict phenotypes (Abid et al., 2018; Babar et al., 2018), which aids the optimization of biotechnological processes toward enhanced production (in quality and quantity) of commercially relevant products (**Figure 2**). This section discusses the application of "omics" in the development of biofuels and bioproducts, agricultural biotechnology, food biotechnology, and bio-therapeutics. In addition, this section discusses the involvement of "omics" technologies in the development of bio-therapeutics for COVID-19.

Omics-Guided Metabolic Engineering of *E. coli* and Yeast Toward the Production of Primary and Secondary Metabolism-Based Biofuels and Bioproducts

Microbial production of bio-based chemicals represents an appealing and more sustainable alternative to traditional petrochemicals (Opgenorth et al., 2019) and has led to a growing catalog of natural products and high-value chemicals (Carbonell et al., 2018). The use of lignocellulosic biomass offers an economical approach to generate biofuels and bioproducts (Fatima et al., 2018). However, to achieve consistent conversion of low-cost input material into value-added products (Yan and Fong, 2018) at industrial levels requires systematic engineering workflows.

The Design-Build-Test-Learn (DBTL) cycle is becoming an increasingly adopted frame-work for metabolic engineering experiments (Opgenorth et al., 2019). It represents a systematic and efficient approach to strain development efforts in biofuels and bio-based products (Opgenorth et al., 2019). Growing interest in the DBTL cycle for metabolic engineering is largely due to improving capabilities in synthetic biology (e.g., synthetic biology tools, DNA synthesis, and genome editing), "omics" technologies, and Learning methods (Carbonell et al., 2018; Opgenorth et al., 2019; Robinson et al., 2020). The DBTL cycle uses synthetic biology to Design and Build genetic constructs in microbial hosts, after which the information gained from "omics" technologies, during the Test phase of the cycle, is passed on to Learning processes (**Figure 3**). What is Learned (e.g., COBRA models) is then fed back to new cycles of design to advance the engineering biology goal (Vavricka et al., 2020) for further strain development and optimization. Thus, facilitating the rapid optimization of microbial strains for production of any chemical compound of interest (Carbonell et al., 2018). Arguably the weakest link in the DBTL cycle workflow is the Learning process since mathematical models (of the engineered bioproduct, pathway, biological system, or biome) are only as good as their assumptions (Liu and Nielsen, 2019). Consequently, both high quality and large "omics" data sets are necessary to improve training models, ensuring increased accuracy and robustness of the Learning process.

Recently, Geiselman et al. utilized the DBTL cycle to engineer *Rhodosporidium toruloides*, an oleaginous yeast species with the ability to grow on lignocellulosic feedstock, to produce the diterpene ent-kaurene, a potential therapeutic, by the native mevalonate pathway and the non-native production of the diterpene precursor geranylgeranyl diphosphate (GGPP). Multi-omics data, in the first round of the DBTL cycle, suggested a limited availability of GGPP. In successive DBTL cycles, an optimal GGPP synthase (GGPPS) was selected, whose expression was balanced with the addition of kaurene synthase from *Gibberella fujikuroi* and a mutant version of farnesyl diphosphate (FPP) synthase from *Gallus gallus* that produces GGPP under strong promoters. The higher ent-kaurene titer achieved was the first demonstration of the production of a non-native diterpene from lignocellulosic hydrolysate in *Rhodosporidium toruloides* (Geiselman et al., 2020). Additionally, Opgenorth et al. used the DBTL cycle approach to optimize 1-dodecanol production in *E. coli* MG1655 strains by modulating ribosome binding sites and acyl-ACP/acyl-CoA reductase on a single operon. The proteomics and metabolomics data collected during the first DBTL cycle were used to train the Learning algorithms, with protein profiles being used to suggest the second DBTL cycle, which led to a 21% increase in 1-dodecanol titer. While this resulted in a 6-fold increase in what was previously reported, the study, however, highlighted the need for more accurate protein expression predictive tools and the importance of genomic sequencing checks on plasmids in cloning and production strains to establishing robust microbial cell factories (Opgenorth et al., 2019).

Adaptive laboratory evolution (ALE) studies the evolutionary forces and adaptive changes influencing microbial strain phenotypes, performance, and stability in order to acquire production strains containing beneficial mutations and positive traits (Dragosits and Mattanovich, 2013; Sandberg et al., 2019; Gibson et al., 2020). Microorganisms are cultured in a desired growth environment for an extended period of time, allowing natural selection to enrich for mutant strains with improved fitness (Johansen, 2018; Sandberg et al., 2019). Therefore, the

TABLE 1 | A comparison of the major "omics" techniques.

Technique	Genomics	Transcriptomics	Proteomics	Metabolomics	Multi-omics
Study	- Genome (the complete set of genes in a biological system) - Genome sequence - Gene functions and interactions - Metagenomics	- Transcriptome (the complete set of RNA transcripts in a biological system) - Gene transcription (i.e., gene expression) - Transcriptome sequence - Metatranscriptomics via 16S rRNA	- Proteome (the complete set of proteins in a biological system) - Protein translation (i.e., gene expression) - Post-translational modification (PTM) state of proteins - Metaproteomics	- Metabolome (the complete set of metabolites in a biological system) - Metabolites (i.e., substrates, intermediates or end products of cellular metabolism) - Pathway flux (i.e., concentration and/or metabolic flux analysis)	- Integrated information of genes, transcriptomes, proteins, and metabolites
Advantages	- Evaluate genome modification in engineered vs. naturally evolved systems	- Evaluate gene function via mRNA transcripts - 16S rRNA sequencing for reconstructing phylogenies	- Assess gene function - Evaluate protein translation - Evaluate PTM - Identify diagnostic biomarkers - Provides phenotypic information	- Assess gene function - Identify metabolic pathway bottlenecks - Identify diagnostic biomarkers (e.g., productivity biomarkers) - Evaluate protein function - Provides phenotypic information	- Identify diagnostic biomarkers with a high degree of accuracy - Provides a comprehensive knowledge and understanding of biological systems
Challenges/ Disadvantages	- Cannot solely provide complete description of complex biological systems (i.e., cannot describe phenotypes)	- Cannot solely provide complete description of complex biological systems (i.e., cannot describe phenotypes) - Insufficient information due to PTM - Cross contamination and cross hybridization	- High instrument cost - Difficult protein/peptide quantification - Inaccurate analysis of labile PTM - Can be expensive as it requires advanced tools (e.g., mass spectrometry) - Low abundance proteins are difficult to analyze - Cross contamination during enzymatic proteolysis (same peptide may come from different proteins) - Difficult to cover whole proteome due to large number of proteins	- High instrument cost - The metabolome is very chemically diverse - Metabolites can have short half-lives due to instability and/or bio-transformations - Low abundance metabolites are difficult to analyze - Challenging sample preparation (e.g., metabolite extraction, and matrix clean up) - Difficult to identify source of metabolite production and consumption in microbial communities	- High data volume and complexity - This approach can be very expensive - Requires good/rigorous experimental design that accounting for all parameters pertaining to individual and combined "omics" technologies - Requires advanced data integration and analysis tools, and specialists from each discipline
Relative throughput	- Highest (fast DNA sequencing)	- High (fast RNA sequencing)	- Moderate	- Moderate	- Depends on selected "omics"
Ideal for	- Testing model-driven hypotheses (targeted approach)	- Testing model-driven hypotheses (targeted approach)	- Targeted (i.e., bottom-up) - Untargeted (i.e., shotgun) - Identifying pathway bottlenecks	- Targeted analysis - Untargeted analysis - Identifying pathway bottlenecks	- Understanding biological systems - Identify diagnostic biomarkers - Identify bioproduction bottleneck
Pathway analysis	- No	- No	- Yes (e.g., via protein abundance/quantity and PTM)	- Yes (e.g., via the assessment of protein function and metabolite production)	- Yes (integrated proteomics and metabolomics)
Relative coverage	- Very comprehensive	- Comprehensive	- Good	- Moderate	- Depend on selected "omics" techniques
Information gained	- Genotype	- Genotype	- Phenotype	- Phenotype	- Connects genotype to phenotype
Type of data	- Qualitative	- Qualitative and quantitative	- Qualitative and quantitative	- Qualitative and quantitative	- Qualitative and quantitative
Relative ease of sample preparation	- Easiest (chemically homogeneous, and stable, thus, easy sample preparation, storage and analysis)	- Easy to moderate	- Moderate (proteins are chemically heterogeneous, and moderately stable, thus, challenging sample preparation, and moderately hard to store)	- Difficult (metabolites are physio-chemically heterogeneous, and unstable (e.g., thermally labile), thus, challenging sample preparation (e.g., metabolite extraction, and matrix clean up), and storage)	- Most difficult
Relative ease of data acquisition	- Easiest	- Easy	- Moderate to difficult	- Difficult (no single analytical technique, nor multiple analytical techniques (e.g., GC-MS, LC-MS, CE-MS, and NMR) can cover the whole metabolome)	- Depends on selected "omics" techniques
Key acquisition tools	- Next generation sequencing (NGS) - PCR - RFLP-PCR	- RNA sequencing (RNA-seq)	- LC-MS (e.g., orbitrap, and TOF) - MALDI	- LC-MS - GC-MS - CE-MS - NMR	- Combination of "omics" acquisition tools
Relative ease of data analysis	- Easiest (fast and accurate data analysis)	- Easy	- Difficult (protein identification and quantification are challenging steps)	- Difficult (data need pre-treatment (e.g., normalization, scaling, and transformation) before analysis)	- Most difficult (integration of various "omics" data is challenging)
Key data analysis tools	- EBI - GEO - ArrayExpress - GenBank	- DESeq2 - DEXseq	- Mascot, Sequest - Tandem - Skyline - Uni-prot, Swiss-prot	- R and Matlab based tools - SIMCA - WebSpecimine, SIRIUS 4 - MetaboAnalyst 4.0, SECIM - COLMAR, MzMine	- ECMDB 2.0 - YMDB 2.0 - GenBank
Overall relative ease of analysis	- Easiest	- Easy	- Moderate to difficult	- Difficult	- Most difficult
Reproducibility	- High	- Good	- Moderate	- Low to moderate	- Depend on selected "omics" techniques

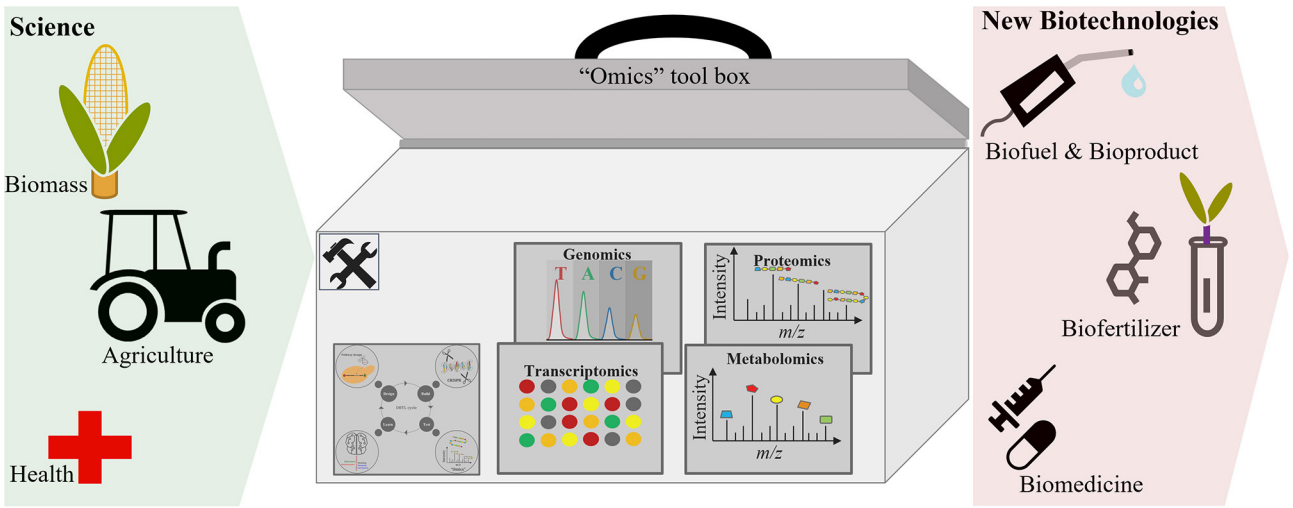


FIGURE 2 | "Omics" approaches facilitate the development of new biotechnologies.

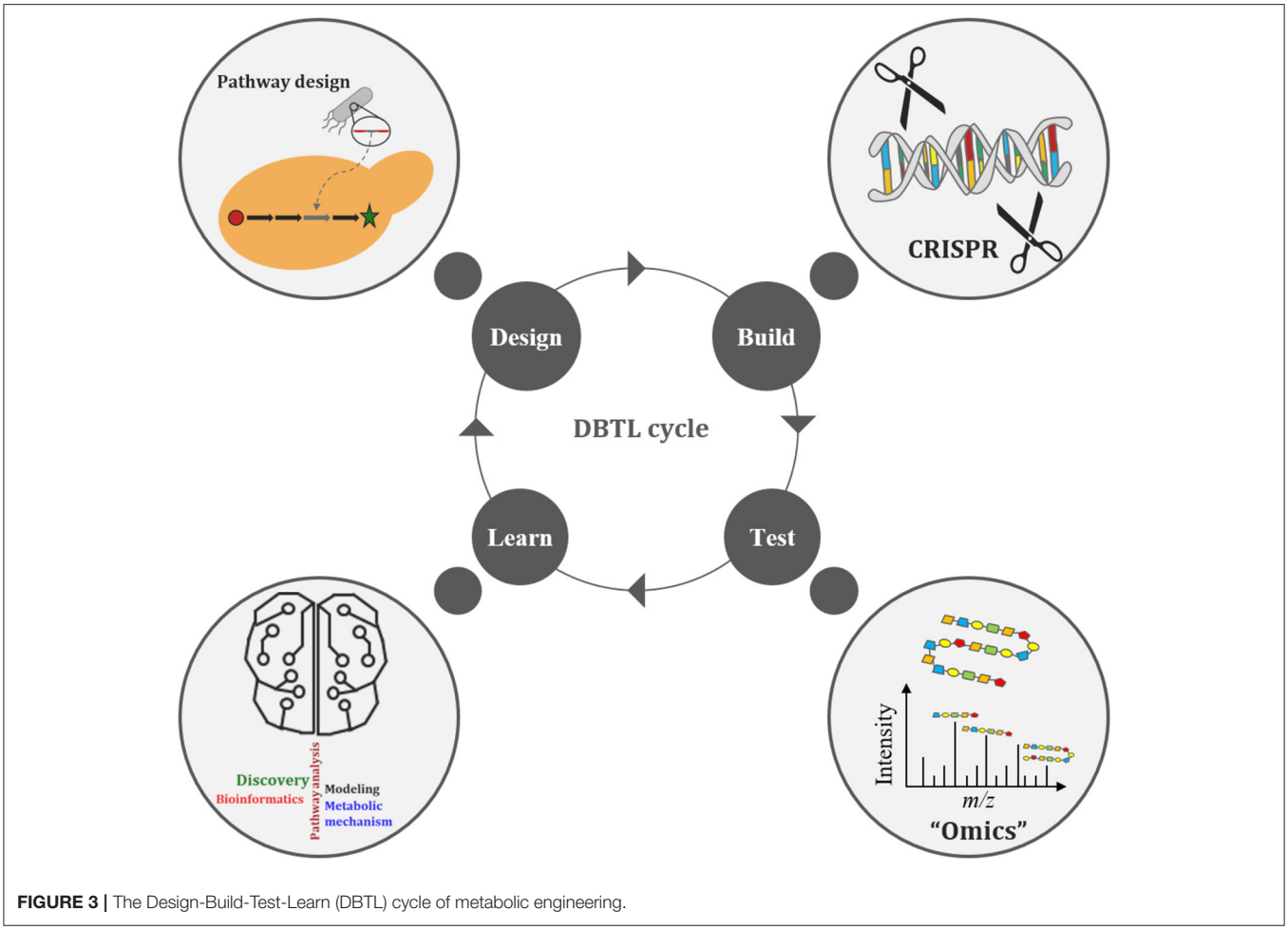


FIGURE 3 | The Design-Build-Test-Learn (DBTL) cycle of metabolic engineering.

throughput of ALE will depend on the experimental design. Furthermore, the task of identifying all beneficial mutations of an ALE experiment remains a major challenge for the field (Phaneuf et al., 2020). ALE is often used to optimize microbial growth rate, increase strain tolerance, improve stress regulation and adaptation, improve substrate utilization and uptake, increase

product titer/rate/yield, as well as for biological discovery via systems biology, evolutionary modeling, and genome dynamics (Bergh, 2018; Dourou et al., 2018; Sun et al., 2018; Wang et al., 2018; Yan and Fong, 2018; Sandberg et al., 2019; Phaneuf et al., 2020). ALE experiments allow researchers to learn how to improve multiple strain properties simultaneously (Sandberg et al., 2019).

ALE has become a valuable tool in metabolic engineering for strain development and optimization by reliably facilitating microbial fitness improvements, via both predictable and non-intuitive mechanisms (Sandberg et al., 2019). ALE can be employed in the DBTL cycle in the Build step to recover strains with fitness issues or to optimize strains (Sandberg et al., 2019). Furthermore, ALE can be used in the Design step to enrich for mutant strains with improved fitness and can also replace both Design and Build steps in situations where a desired phenotype is tied to selection without the need for engineering (Sandberg et al., 2019; Lee and Kim, 2020). While ALE can precede the Test and Learn steps in the DBTL cycle, the information gained from these steps can also be utilized by ALE to produce strains with better properties. In this way, ALE may benefit from using "omics" technologies during the Test phase of the DBTL cycle (Horinouchi et al., 2018; Long and Antoniewicz, 2018; Walker et al., 2019; Wu et al., 2020).

In most microbial metabolic engineering studies, however, the Learn phase of the DBTL cycle is often addressed by hypothesis-driven user intuition that is often based on empirical evidence (Liu and Nielsen, 2019). As with the DBTL cycle, genomic sequence information in the traditional approach is invariably utilized in the initial stages of a study. However, genomics has come a long way in the last decade. Bar-seq can now be used to study population dynamics of *Saccharomyces cerevisiae* (*S. cerevisiae*) deletion libraries during bioreactor cultivation, enabling the identification of factors that impact the diversity of a mutant pool (Wehrs et al., 2020). Whereas, a sort-seq-guided engineering approach can be used to identify key mutated promotors for tuning expression levels, thereby facilitating the dynamic regulation of microbial growth as well as dynamic pathway regulation (Rohlhill et al., 2017). While transcriptomics yields gene expression data (i.e., activity of target genes, gene sequence data, and gene expression levels), proteomics and metabolomics approaches are increasingly being used for pathway analysis studies as they can measure protein translation and activity, respectively (Volke et al., 2019). Proteomics-guided approaches have been used to engineer polyketide biosynthesis platforms for aromatic compounds in yeast (Jakočiunas et al., 2020) and *in vitro* production of adipic acid (Hagen et al., 2016). In addition to this, metabolomics enables the assessment of pathway flux, carbon source diversion, and cofactor imbalance, which all contribute to the identification of pathway bottlenecks (Nielsen and Jewett, 2007; Zhao et al., 2020, Volke et al., 2019). Luo et al. used a metabolomics guided approach to characterize cannabinoid production in engineered *S. cerevisiae* and identified cannabinoid analogs produced by several promiscuous pathway genes (Luo et al., 2019). Furthermore, metabolomics analysis aided the design (Kang et al., 2016) and optimization (Kang et al., 2019) of a novel isopentenyl diphosphate-bypass mevalonate pathway in *E. coli* for C5 alcohol production. With a combined genetic, biochemical and fermentation approach, Uranukul et al. utilized the native glycolytic pathway in *S. cerevisiae* to produce monoethylene glycol, an important commodity chemical , and upon further metabolic engineering and process optimization were able to achieve 4.0 g/L (Uranukul et al., 2019). The integration of proteomics and metabolomics promises accurate assessment of pathway flux due to proper accounting of protein abundance. When pathway data is obtained in addition to transcriptomics data and/or large scale targeted/untargeted proteomics or metabolomics data, the impact of the engineered pathway on cellular metabolism and physiology can be determined. Exploring the interplay between heterologously expressed pathways and endogenous metabolism could reveal factors affecting strain variation, identify perturbed metabolic nodes, and produce new engineering targets (Chen, 2016).

The Application of "Omics" Technologies to Agricultural and Food Biotechnology

Recent advances in agricultural biotechnology have led to new plant varieties being engineered by recombinant DNA technology and grown by farmers to respond to market demands and environmental challenges (https://www.usda.gov/topics/biotechnology). "Omics" technologies are being applied to agricultural biotechnology to enhance desirable phenotypic traits (e.g., color, taste, drought tolerance, pesticide resistance, etc.) (Aliferis and Chrysayi-tokousbalides, 2011; Van Emon, 2016). While "omics" plays a major role in improving crop quality, consistency, and productivity, they have also led to the development of food crops with enhanced nutritional composition (Van Emon, 2019) (**Figure 2**). Moreover, omics-driven systems biology provides an understanding of the interactions between the "omes" and mechanisms involved and provide links between genes and traits (Van Emon, 2016).

As arable land is being farmed more heavily, soil is becoming more susceptible to loss of structure, organic matter, minerals, and erosion. Thus, efforts are being made, via agricultural biotechnology, to maintain a sustainable supply of nutrients essential to the growth of crop plants. An integral part of this approach is the use of biofertilizers, which are the preparations containing specialized living organisms (i.e., microbial inoculants) that can fix, mobilize, solubilize, or decompose nutrient sources, and are applied through seed or soil to enhance nutrient uptake by plants (Mohanram and Kumar, 2019). The sustainable enhancement in food production from less available arable land relies on the balanced utilization of inorganic minerals, organic matter, and biofertilizer sources of plant nutrients to augment and maintain soil fertility and productivity. Widespread adoption of this approach, however, has been hindered by varying responses of microbial inoculants across fields and crops (Mohanram and Kumar, 2019). As a result, there is an urgent need to understand the mechanisms underlying the interdependencies between soil microbial communities and the host plant and their impact on crop productivity. These interactions are played out in the rhizosphere, which encompasses the region of soil that is

directly influenced by root secretions and associated microbial communities (Zhalnina et al., 2018). In a recent comparative genomics and exo-metabolomics study, specific rhizosphere bacteria were shown to have a natural preference for certain aromatic organic acids exuded by plants, suggesting that plant exudation traits and microbial substrate uptake traits interact to yield the patterns of microbial community assembly (Zhalnina et al., 2018). Furthermore, the application of genomics and transcriptomics to the study of luxury phosphate uptake (i.e., the ability of microalgae to take up more phosphorus than necessary for immediate growth) revealed a range of Pi transporters in various microalgae and their expression patterns in relation to the availability of P (Yang et al., 2018; Mohanram and Kumar, 2019; Solovchenko et al., 2019). At present, "omics" technologies are being used to understand complex rhizospheric intercommunications, which is crucial to the development and choice of biofertilizer and, by extension, the construction of rhizopheres that promote stable plant growth, better crop productivity and yield (Mohanram and Kumar, 2019).

In the related field of food biotechnology, transcriptomics and metabolomics analysis showed that *Bacillus pumilus* LZP02 promote the growth of rice roots by enhancing carbohydrate metabolism and phenylpropanoid biosynthesis (Liu et al., 2020). Further, the application of "omics" to starch bioengineering is increasing our understanding of the specific contributions of the most important enzymes for starch biosynthesis. This has enhanced our ability to predict how starch-related phenotypes can be modified, thus ensuring further progress in the research field of rice starch biotechnology (Nakamura, 2018). "Omics" are solving issues surrounding food quality and traceability, to safeguard the origin of food, and discover biomarkers of potential food safety problems (Ferranti, 2018). In the wine industry, the wine microbiome associated with the fermentation of must has a great influence on factors transforming grapes to wine, including flavor and aroma. "Omics" characterization of the complex interactions between these microbes, the substrate and environment, is key to shaping wine production (Sirén et al., 2019). Finally, combining "omics" technologies with genome editing of food microbes can be used to generate enhanced probiotic strains, develop novel bio-therapeutics and alter microbial community structure in food matrices (Pan and Barrangou, 2020).

The Use of "Omics" Technologies in the Development of Therapeutics for COVID-19

The coronavirus disease 2019 (COVID-19) characterized by the Severe acute respiratory syndrome coronavirus 2 [i.e., SARS-CoV-2, which binds to the ACE2 receptor in the lung and other organs (Ahmed et al., 2020)] has caused a global pandemic and slowed much of the world's economy. To date (December 5th, 2020), there are more than 64 million confirmed cases and 1.5 million confirmed deaths world-wide (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). Thus, there is a pressing need for an effective countermeasure to mitigate the spread of the pandemic (van Doremalen et al., 2020).

Consequently, efforts are underway to fast-track the development and production of safe and effective vaccines against SARS-CoV-2. Prior knowledge of SARS and Middle East respiratory syndrome (MERS) has enabled scientists to target the spike protein as the viral antigen (via the ACE2 receptor). Moreover, the release of the SARS-CoV-2 genome sequence in January 2020 made it possible to expedite the development of next generation mRNA [e.g., *mRNA-1273 from NIH/Moderna* (Jackson et al., 2020) *and BNT162 from Pfizer/BioNTech* (Mulligan et al., 2020)] and DNA [e.g., ChAdOx1 nCoV-19 from University of Oxford/Vaccitech/AstraZeneca (Folegatti et al., 2020; van Doremalen et al., 2020)] vaccine platforms that encode for the antigen. Once injected into a host, the former (which is encased in lipid nano-particles) remains in the cytoplasm while the latter (which is encased in an attenuated adenovirus vector) enters its nucleus. The host cell translates these genetic materials into the spike protein, which decorates the surface of the cell and elicits an adaptive immune response mediated by T cells (e.g., $CD4^+$ and $CD8^+$) and B cells (i.e., antibodies). These vaccines were reported to be efficacious against SARS-CoV-2 in recent clinical trials, which underscores the importance of genomics to this new era of vaccine development.

Gordon et al. produced a SARS-CoV-2 protein interaction map via a proteomics-based approach to reveal targets for drug repurposing. They cloned, affinity tagged and expressed 26 of the 29 SARS-CoV-2 proteins in human cells and identified the associated proteins via proteomics analysis. A total of 66 human proteins or host factors were identified as possible drug targets of 69 compounds, from which two sets of these pharmacological agents showed antiviral activity. This work highlights the potential of host-factor-targeting agents, when acting alone or in combination with drugs that target viral enzymes, to be used as therapeutic treatments for COVID-19 (Gordon et al., 2020). Furthermore, computational immunoproteomics studies have the potential to guide lab-based investigations to evaluate specificity of diagnostic products, to forecast on potential adverse effects of vaccines and to reduce the use of animal models (Tilocca et al., 2020).

Recently, metabolomics was able to distinguish COVID-19 patients from healthy controls via the analysis of 10 plasma metabolites. Additionally, lipidomics data from this study suggests monosialodihexosyl ganglioside enriched exosomes could be involved in pathological processes related to COVID-19 pathogenesis (Song et al., 2020). Recent proteomics and metabolomics studies in COVID-19 patient sera suggest that SARS-CoV-2 infection causes metabolic dysregulation of macrophage and lipid metabolism, platelet degranulation, complement system pathways, and massive metabolic suppression (Shen et al., 2020); with the plasma metabolomic signatures appearing to be similar to those described for sepsis syndrome (Langley et al., 2013, 2014; Migaud et al., 2020). Furthermore, transcriptomics results indicate higher expression of genes related to oxidative phosphorylation both in peripheral mononuclear leukocytes and bronchoalveolar lavage fluid, suggesting a critical role for mitochondrial activity during SARS-CoV-2 infection (Gardinassi et al., 2020). Understanding the clinical presentation of COVID-19 as well as metabolomic,

proteomic, and genetic profiles could lead to the discovery of diagnostic, prognostic and predictive biomarkers, ensuring the development of more effective medical therapy (Ahmed et al., 2020). Moreover, identifying metabolic biomarkers of severe vs. mild disease states in the lung during respiratory infections could lead to the development of novel therapeutics that modulate symptom and disease severity (Bernatchez and McCall, 2020; Shen et al., 2020). It is, therefore, critical to develop new approaches to early assess which cases will likely become clinically severe (Shen et al., 2020).

TRENDS IN "OMICS" RELATED BIOTECHNOLOGY RESEARCH

The aim of this section is to present the trends in "omics" techniques utilization in biotechnology research (i.e., food, natural products, agriculture, pharmaceutical, materials, and bioenergy) during the course of last two decades. Such trends, which are based on the number of annual publications obtained from a search in Web of Science (www.webofknowledge.com) topics, are used to illustrate the progression of "omics" technologies in biotechnology. In Web of Science, the topics searched are as follows: title, abstract, author keywords, and keywords plus. It is worth noting that the trends in this section showing a reduction in the number of publications for the current year (i.e., 2020), which is expected to be a result of the COVID-19 related shut-down that has affected scientific labs worldwide and the mid-2020 collection of the data. Furthermore, it is also important to note that search entries, such as food, and bioenergy can generate publications based on the contributions of both closely and, to a lesser extent, loosely related topics. Additionally, it is worth noting that not all "omics" research data (e.g., industrial-based studies) is being published and made available to the public. Therefore, the trends data (**Figures 4**, **5**) represents a qualitative rather than a quantitative measure of "omics" utilization.

Omics-based technologies serve as the connective tissue that links biotechnology to these fields of research. For example, the advancement in genomics technologies have improved biotechnology platforms, which have led to developments in pharmaceutical, bioenergy, food, materials, and agriculture research (Oksman-Caldentey and Saito, 2005; Crommelin et al., 2013; Misra et al., 2013; de Pablo et al., 2019; Rexroad et al., 2019). **Figure 4** shows a steady growth in the number of genomics publications in biotechnology research, which might be due to the advancement in DNA sequencing, resulting in reduced cost and increased throughput (Pagani et al., 2012). Transcriptomics also appears to show a slight increase in the application of biotechnology-based research during the last decade (**Figure 4**).

Interestingly, while the utilization of proteomics in biotechnology seems to be significantly increased in the period of 2004–2006, it has generally plateaued during the following years (**Figure 4**). However, **Figure 4** suggests a growing trend in the application of metabolomics studies to biotechnology-based research during the last decade. Additionally, metabolic flux analysis has also shown a slight growing trend in the last decade.

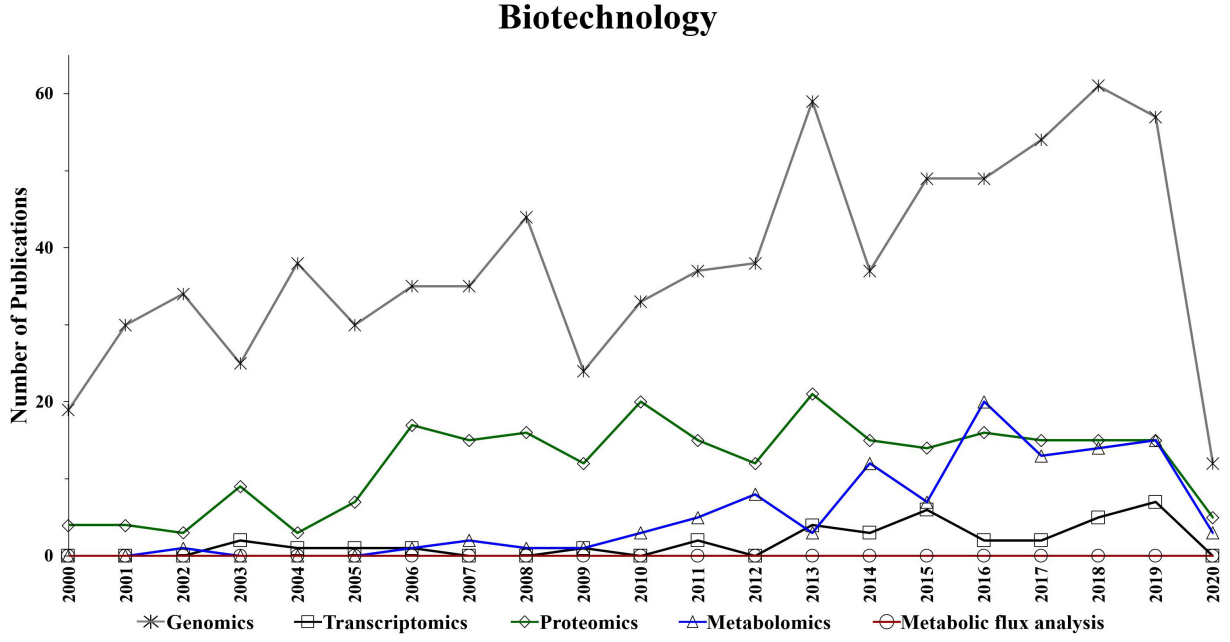


FIGURE 4 | Number of annual publications that utilize "omics" technologies for biotechnology during the period of 2000–2020. Search criteria: the individual "omics" technology was selected. The search was conducted using the Web of Science platform.

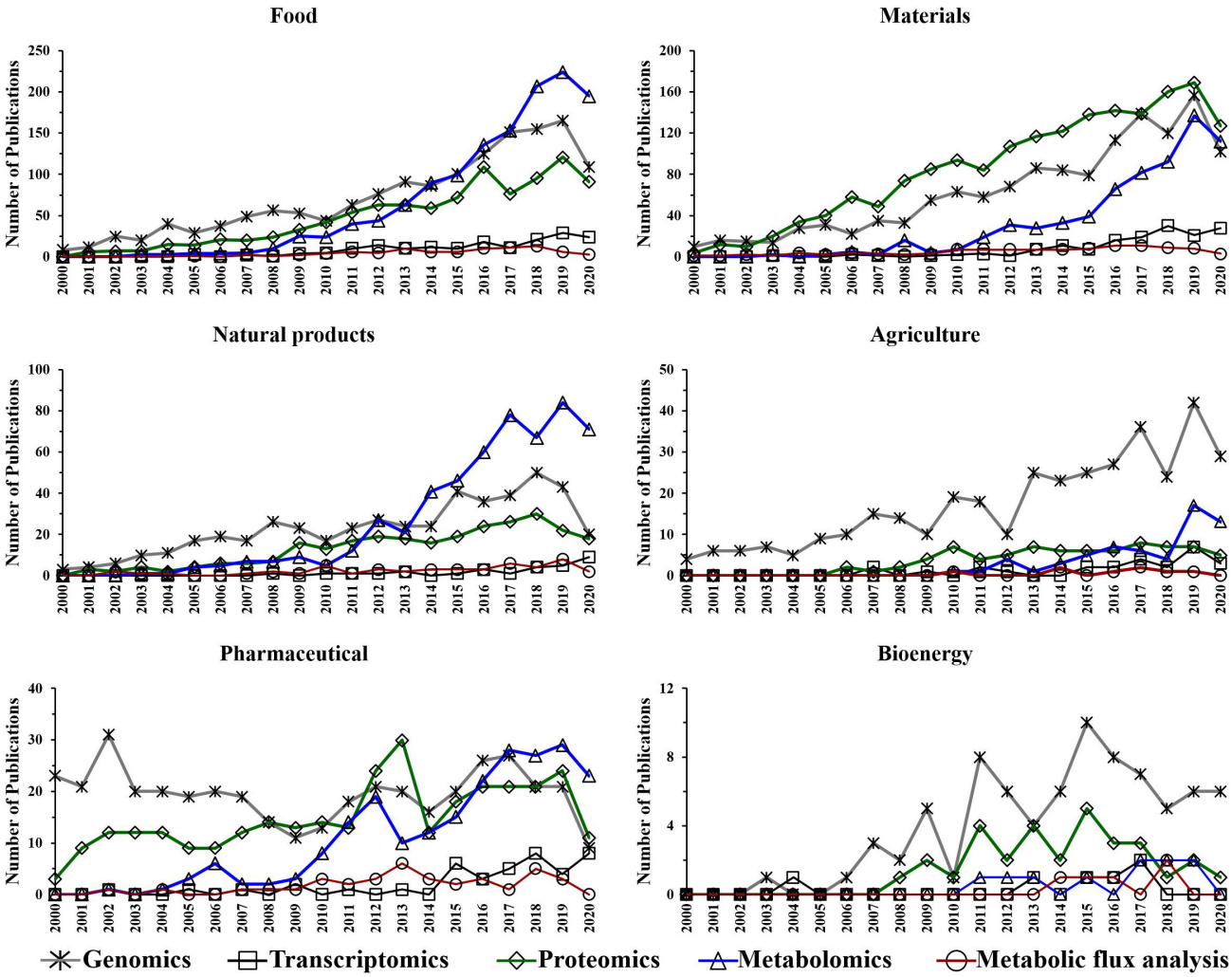


FIGURE 5 | Number of annual publications that utilize "omics" technologies for biotechnology research areas during the period of 2000–2020. Search criteria: the individual "omics" technology and the individual biotechnology research area were selected. The search was conducted using the Web of Science platform.

The performed search also suggests an increased utilization of genomics, proteomics and metabolomics during the last two decades in the following research fields: food, materials, and natural products (**Figure 5**). This is not surprising as there is a growing need for more phenotypic information. Consequently, scientists are using these "omics" techniques to facilitate their research. The improvements in proteomics and metabolomics analytical capabilities may also have contributed to the potential growth in their utilization over the last decade for those fields. Genomics has shown an upward trend in the number of publications for agriculture over the last 20 years (**Figure 5**).

The utilization of metabolic flux analysis in all research areas has shown a prospective slight growth trend during the last decade (**Figure 5**). The number of metabolic flux analysis publications, however, is relatively low in the perspective areas for similar reasons to that described for biotechnology.

CONCLUSIONS

The DBTL cycle is becoming an increasingly adopted framework in metabolic engineering experiments as it provides a systematic and efficient approach to strain development. However, the DBTL cycle is limited by the Learning process since it requires high quality and large "omics" data sets to increase the accuracy and robustness of Learn methods. The DBTL cycle relies heavily on "omics" technologies during the testing phase of the cycle, and can be integrated into ALE experiments. "Omics" technologies have played major roles in the metabolic engineering of biofuels, bioproducts, and crop development. Proteomics and metabolomics are routinely applied to the analysis of engineered biosynthetic pathways in microbial hosts. Genomics sequencing information appears to be a key component in the development of next generation mRNA and DNA vaccines against virus's such as SARS-CoV-2. Whereas, transcriptomics, proteomics, and metabolomics analyses are

being used to guide the development of therapeutic drugs for COVID-19. In the last 20 years, genomics has shown a steady growth in the number of biotechnology publications, however, the emergence of transcriptomics, proteomics, and metabolomics in this field of research is a testament to the development of robust "omics" technologies and methods.

AUTHOR CONTRIBUTIONS

BA conducted the data curation, data analysis, and literature review. EEKB conducted the data analysis and literature review. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

REFERENCES

Abid, F., Zahid, M. A., Abedin, Z. U., Nizami, S. B., Abid, M. J., Kazmi, S. Z. H., et al. (2018). *Omics Approaches in Marine Biotechnology: The Treasure of Ocean for Human Betterments*. London: Elsevier Inc.

Ahmed, Z., Zeeshan, S., Foran, D. J., Kleinman, L. C., Wondisford, F. E., and Dong, X. Q. (2020). Integrative clinical, genomics and metabolomics data analysis for mainstream precision medicine to investigate COVID-19. *BMJ Innov.* 7, 1–5. doi: 10.1136/bmjinnov-2020-000444

Aichler, M., and Walch, A. (2015). MALDI imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab. Investig.* 95, 422–431. doi: 10.1038/labinvest.2014.156

Aizat, W. M., Ismail, I., and Noor, N. M. (2018). "Recent development in omics studies," in *Advances in Experimental Medicine and Biology*, eds W. M. Aizat, H-H. Goh and S. N. Baharum (New York, NY: Springer New York LLC), 1–9. doi: 10.1007/978-3-319-98758-3_1

Alfaro, M. P., Sepulveda, J. L., and Lyon, E. (2019). "Molecular testing for targeted therapies and pharmacogenomics," in *Accurate Results in the Clinical Laboratory*, eds A. Dasgupta and J. L. Sepulveda (Amsterdam: Elsevier), 349–363. doi: 10.1016/b978-0-12-813776-5.00022-4

Aliferis, K. A., and Chrysayi-tokousbalides, M. (2011). Metabolomics in pesticide research and development : review and future perspectives. *Metabolomics* 7, 35–53. doi: 10.1007/s11306-010-0231-x

Ando, D., and García Martín, H. (2019). "Genome-scale 13C fluxomics modeling for metabolic engineering of *Saccharomyces cerevisiae*," in *Methods in Molecular Biology*, ed E. E. K. Baidoo (New York, NY: Humana Press Inc.), 317–345. doi: 10.1007/978-1-4939-8757-3_19

Babar, M. M., Afzaal, H., Pothineni, V. R., Zaidi, N. U. S. S., Ali, Z., Zahid, M. A., et al. (2018). "Omics approaches in industrial biotechnology and bioprocess engineering," in *Omics Technologies and Bio-engineering: Volume 2: Towards Improving Quality of Life*, eds D. Barh and V. Azevedo (London: Elsevier), 251–269. doi: 10.1016/B978-0-12-815870-8.00014-0

Babele, P. K., and Young, J. D. (2019). Applications of stable isotope-based metabolomics and fluxomics toward synthetic biology of cyanobacteria. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 13:e1472. doi: 10.1002/wsbm.1472

Baidoo, E. E. K. (2019). *Microbial Metabolomics: A General Overview*. New York, NY: Humana Press.

Baidoo, E. E. K., and Teixeira Benites, V. (2019). Mass spectrometry-based microbial metabolomics: techniques, analysis, and applications. *Methods Mol. Biol.* 1859, 11–69. doi: 10.1007/978-1-4939-8757-3_2

Bairoch, A., and Boeckmann, B. (1994). The SWISS-PROT protein sequence data bank: current status. *Nucleic Acids Res.* 22, 3578–3580.

Bashiardes, S., Zilberman-Schapira, G., and Elinav, E. (2016). Use of metatranscriptomics in microbiome research. *Bioinform. Biol. Insights* 10, 19–25. doi: 10.4137/BBI.S34610

Batrakou, D. G., Müller, C. A., Wilson, R. H. C., and Nieduszynski, C. A. (2020). DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. *Nat. Protoc.* 15, 1255–1284. doi: 10.1038/s41596-019-0287-7

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, 36–42. doi: 10.1093/nar/gks1195

Bergh, B. Van Den (2018). Experimental design, population dynamics, and diversity in microbial experimental evolution. *Microbiol. Mol. Biol. Rev.* 82, 1–54. doi: 10.1128/MMBR.00008-18

Bernatchez, J. A., and McCall, L. I. (2020). Insights gained into respiratory infection pathogenesis using lung tissue metabolomics. *PLoS Pathog.* 16:e1008662. doi: 10.1371/journal.ppat.1008662

Bhatia, S. (2018). "History, scope and development of biotechnology," in *Introduction to Pharmaceutical Biotechnology*, eds S. Bhatia and D. Goli (Bristol: IOP Publishing Ltd), 1–61. doi: 10.1088/978-0-7503-1299-8ch1

Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y. C., Meng, C., et al. (2020). Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. *Nat. Commun.* 11, 1–12. doi: 10.1038/s41467-019-13973-x

Boguszewicz, Bieleń, A., Mrochem-Kwarciak, J., Skorupa, A., Ciszek, M., Heyda, A., et al. (2019). NMR-based metabolomics in real-time monitoring of treatment induced toxicity and cachexia in head and neck cancer: a method for early detection of high risk patients. *Metabolomics* 15, 1–16. doi: 10.1007/s11306-019-1576-4

Brown, C. G., and Clarke, J. (2016). Nanopore development at Oxford Nanopore Rapid, semi-automated protein terminal characterization using ISDetect. *Nat. Publ. Gr.* 34, 481–482. doi: 10.1038/nbt.3622

Brunk, E., George, K. W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., et al. (2016). Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell Syst.* 2, 335–346. doi: 10.1016/j.cels.2016.04.004

Bryson, S., Li, Z., Pett-Ridge, J., Hettich, R. L., Mayali, X., Pan, C., et al. (2016). Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. *mSystems* 1, 1–16. doi: 10.1128/msystems.00027-15

Carbonell, P., Jervis, A. J., Robinson, C. J., Yan, C., Dunstan, M., Swainston, N., et al. (2018). An automated design-build-test-learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.* 1:66. doi: 10.1038/s42003-018-0076-9

Cardoso, S., Afonso, T., Maraschin, M., and Rocha, M. (2019). WebSpecmine: a website for metabolomics data analysis and mining. *Metabolites* 9:237. doi: 10.3390/metabo9100237

Catherman, A. D., Skinner, O. S., and Kelleher, N. L. (2014). Top down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* 445, 683–693. doi: 10.1016/j.bbrc.2014.02.041

Chatham, J. C., and Blackband, S. J. (2001). Nuclear magnetic resonance spectroscopy and imaging in animal research. *ILAR J.* 42, 189–208. doi: 10.1093/ilar.42.3.189

Chekmeneva, E., Dos Santos Correia, G., Chan, Q., Wijeyesekera, A., Tin, A., Young, J. H., et al. (2017). Optimization and application of direct infusion nanoelectrospray HRMS method for large-scale urinary metabolic phenotyping in molecular epidemiology. *J. Proteome Res.* 16, 1646–1658. doi: 10.1021/acs.jproteome.6b01003

Chen, B. J., Lam, T. C., Liu, L. Q., and To, C. H. (2017). Post-translational modifications and their applications in eye research. *Mol. Med. Rep.* 15, 3923–3935. doi: 10.3892/mmr.2017.6529

Chen, G. Q. (2016). Omics meets metabolic pathway engineering. *Cell Syst.* 2, 362–363. doi: 10.1016/j.cels.2016.05.005

Chen, Y., Guenther, J. M., Gin, J. W., Chan, L. J. G., Costello, Z., Ogorzalek, T. L., et al. (2019). Automated "cells-to-peptides" sample preparation workflow for high-throughput, quantitative proteomic assays of microbes. *J. Proteome Res.* 18, 3752–3761. doi: 10.1021/acs.jproteome.9b00455

Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* 32, 1351–1359. doi: 10.1007/s10529-010-0306-9

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 46, W486–W494. doi: 10.1093/nar/gky310

Clomburg, J. M., Crumbley, A. M., and Gonzalez, R. (2017). Industrial biomanufacturing: the future of chemical production. *Science* 355:aag0804. doi: 10.1126/science.aag0804

Craig, R., and Beavis, R. C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467. doi: 10.1093/bioinformatics/bth092

Crommelin, D. J. A., Sindelar, R. D., and Meibohm, B. (2013). *Pharmaceutical Biotechnology*: *Fundamentals and Applications, 4th Edn.* New York, NY: Springer.

Cupples, A. M. (2005). Principles and applications of soil microbiology, second edition. *J. Environ. Qual.* 34, 731–732. doi: 10.2134/JEQ2005.0731DUP

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., et al. (2008). Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem.* 80, 2921–2931. doi: 10.1021/ac702422x

de Pablo, J. J., Jackson, N. E., Webb, M. A., Chen, L. Q., Moore, J. E., Morgan, D., et al. (2019). New frontiers for the materials genome initiative. *NPJ Comput. Mater.* 5, 1–23. doi: 10.1038/s41524-019-0173-4

Doellinger, J., Schneider, A., Hoeller, M., and Lasch, P. (2020). Sample preparation by easy extraction and digestion (SPEED)–a universal, rapid, and detergent-free protocol for proteomics based on acid extraction. *Mol. Cell. Proteomics* 19, 209–222. doi: 10.1074/mcp.TIR119.001616

Doran, M. L., Mykytczuk, N., Bieniek, A., Methé, A., and Merritt, T. J. S. (2017). Evaluation of quenching and extraction procedures for performing metabolomics in *Acidithiobacillus ferrooxidans*. *Metabolomics* 13, 1–10. doi: 10.1007/s11306-017-1298-4

Dourou, M., Aggeli, D., Papanikolaou, S., and Aggelis, G. (2018). Critical steps in carbon metabolism affecting lipid accumulation and their regulation in oleaginous microorganisms. *Appl. Microbiol. Biotechnol.* 102, 2509–2523. doi: 10.1007/s00253-018-8813-z

Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution–principles and applications for biotechnology. *Microb. Cell Fact.* 12:64. doi: 10.1186/1475-2859-12-64

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302. doi: 10.1038/s41592-019-0344-8

Duong, V. A., Park, J. M., and Lee, H. (2020). Review of three-dimensional liquid chromatography platforms for bottom-up proteomics. *Int. J. Mol. Sci.* 21:1524. doi: 10.3390/ijms21041524

Durot, M., Bourguignon, P. Y., and Schachter, V. (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Rev.* 33, 164–190. doi: 10.1111/j.1574-6976.2008.00146.x

Ellis, D. I., and Goodacre, R. (2012). Metabolomics-assisted synthetic biology. *Curr. Opin. Biotechnol.* 23, 22–28. doi: 10.1016/j.copbio.2011.10.014

Eng, J. K., McCormack, A. L., and Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–989. doi: 10.1016/1044-0305(94)80016-2

Esvelt, K. M., and Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* 9:641. doi: 10.1038/msb.2012.66

Fatima, S., Hameed, A., Noman, M., Ahmed, T., Shahid, M., Tariq, M., et al. (2018). Lignocellulosic biomass : a sustainable bioenergy source for future lignocellulosic biomass : a sustainable bioenergy source for the future. *Protein Pept. Lett.* 25, 1–16. doi: 10.2174/0929866525666180122144504

Feng, X., Page, L., Rubens, J., Chircus, L., Colletti, P., Pakrasi, H. B., et al. (2010). Bridging the gap between fluxomics and industrial biotechnology. *J. Biomed. Biotechnol.* 2010:460717. doi: 10.1155/2010/460717

Ferranti, P. (2018). The future of analytical chemistry in foodomics. *Curr. Opin. Food Sci.* 22, 102–108. doi: 10.1016/j.cofs.2018.02.005

Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* 2, 155–168. doi: 10.1002/cfg.82

Fiehn, O. (2002). Metabolomics - The link between genotypes and phenotypes. *Plant Mol. Biol.* 48, 155–171. doi: 10.1023/A:1013713905833

Folegatti, P. M., Ewer, K. J., Aley, P. K., Angus, B., Becker, S., Belij-Rammerstorfer, S., et al. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* 396, 467–478. doi: 10.1016/S0140-6736(20)31604-4

Frese, K. S., Katus, H. A., and Meder, B. (2013). Next-generation sequencing: From understanding biology to personalized medicine. *Biology (Basel)* 2, 378–398. doi: 10.3390/biology2010378

Gardinassi, L. G., Souza, C. O. S., Sales-Campos, H., and Fonseca, S. G. (2020). Immune and metabolic signatures of COVID-19 revealed by transcriptomics data reuse. *Front. Immunol.* 11:1636. doi: 10.3389/fimmu.2020.01636

Geiselman, G. M., Zhuang, X., Kirby, J., Tran-Gyamfi, M. B., Prahl, J. P., Sundstrom, E. R., et al. (2020). Production of ent-kaurene from lignocellulosic hydrolysate in *Rhodosporidium toruloides*. *Microb. Cell Fact.* 19, 1–12. doi: 10.1186/s12934-020-1293-8

Gibson, B., Dahabieh, M., Krogerus, K., Jouhten, P., Magalhães, F., Pereira, R., et al. (2020). Adaptive laboratory evolution of ale and lager yeasts for improved brewing efficiency and beer quality. *Annu. Rev.* 11, 23–44. doi: 10.1146/annurev-food-032519-051715

Gilbert, D., and Heiner, M. (2015). *Advances in Computational Methods in Systems Biology*. Amsterdam: Elsevier.

Giraudeau, P. (2020). NMR-based metabolomics and fluxomics: developments and future prospects. *Analyst* 145, 2457–2472. doi: 10.1039/d0an00142b

Gonzalez, P., and Pierron, F. (2015). "Omics in aquatic ecotoxicology: the ultimate response to biological questions?" in *Aquatic Ecotoxicology: Advancing Tools for Dealing With Emerging Risks*, eds C. Amiard-Triquet, J.-C. Amiard and C. Mouneyrac (London: Elsevier Inc.), 183–203. doi: 10.1016/B978-0-12-800949-9.00008-5

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9

Gross, T., Mapstone, M., Miramontes, R., Padilla, R., Cheema, A. K., Macciardi, F., et al. (2018). Toward reproducible results from targeted metabolomic studies: perspectives for data pre-processing and a basis for analytic pipeline development. *Curr. Top. Med. Chem.* 18, 883–895. doi: 10.2174/1568026618666180711144323

Guo, X., Fillmore, T. L., Gao, Y., and Tang, K. (2016). Capillary electrophoresis-nanoelectrospray ionization-selected reaction monitoring mass spectrometry via a true sheathless metal-coated emitter interface for robust and high-sensitivity sample quantification. *Anal. Chem.* 88, 4418–4425. doi: 10.1021/acs.analchem.5b04912

Guo, Y., Sheng, Q., Li, J., Ye, F., Samuels, D. C., and Shyr, Y. (2013). Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS ONE* 8:e71462. doi: 10.1371/journal.pone.0071462

Hagen, A., Poust, S., De Rond, T., Fortman, J. L., Katz, L., Petzold, C. J., et al. (2016). Engineering a polyketide synthase for *in vitro* production of adipic acid. *ACS Synth. Biol.* 5, 21–27. doi: 10.1021/acssynbio.5b00153

Hamann, E., Gruber-Vodicka, H., Kleiner, M., Tegetmeyer, H. E., Riedel, D., Littmann, S., et al. (2016). Environmental *Breviatea harbour* mutualistic *Arcobacter epibionts*. *Nature* 534, 254–258. doi: 10.1038/nature18297

Hansen, B. T., Jones, J. A., Mason, D. E., and Liebler, D. C. (2001). Salsa: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal. Chem.* 73, 1676–1683. doi: 10.1021/ac001172h

Hayoun, K., Gouveia, D., Grenga, L., Pible, O., Armengaud, J., and Alpha-Bazin, B. (2019). Evaluation of sample preparation methods for fast proteotyping of microorganisms by Tandem mass spectrometry. *Front. Microbiol.* 10:1985. doi: 10.3389/fmicb.2019.01985

Hinzke, T., Kouris, A., Hughes, R. A., Strous, M., and Kleiner, M. (2019). More is not always better: evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. *Front. Microbiol.* 10:238. doi: 10.3389/fmicb.2019.00238

Horinouchi, T., Maeda, T., and Furusawa, C. (2018). Understanding and engineering alcohol-tolerant bacteria using OMICS technology. *World J. Microbiol. Biotechnol.* 34, 1–9. doi: 10.1007/s11274-018-2542-4

Iwamoto, N., and Shimada, T. (2018). Recent advances in mass spectrometry-based approaches for proteomics and biologics: great contribution for developing therapeutic antibodies. *Pharmacol. Ther.* 185, 147–154. doi: 10.1016/j.pharmthera.2017.12.007

Jackson, L. A., Anderson, E. J., Rouphael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., et al. (2020). An mRNA vaccine against SARS-CoV-2—preliminary report. *N. Engl. J. Med.* 383, 1920–1931. doi: 10.1056/nejmoa2022483

Jakočiunas, T., Klitgaard, A. K., Kontou, E. E., Nielsen, J. B., Thomsen, E., Romero-Suarez, D., et al. (2020). Programmable polyketide biosynthesis platform for production of aromatic compounds in yeast. *Synth. Syst. Biotechnol.* 5, 11–18. doi: 10.1016/j.synbio.2020.01.004

Jeong, K., Kim, S., and Pevzner, P. A. (2013). UniNovo: a universal tool for *de novo* peptide sequencing. *Bioinformatics* 29, 1953–1962. doi: 10.1093/bioinformatics/btt338

Johansen, E. (2018). Use of natural selection and evolution to develop new starter cultures for fermented foods. *Annu. Rev. Food Sci. Technol.* 9, 411–428. doi: 10.1146/annurev-food-030117-012450

Kambale, R., Rahman, H., and Omics, S. (2019). *Fluorescence Activated Cell Sorting Single-Cell Omics Approaches in Plants*. London: Elsevier.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., and Subramaniam (2002). The KEGG database. *Novartis Found. Symp.* 247, 91–103. doi: 10.1002/0470857897.ch8

Kang, A., George, K. W., Wang, G., Baidoo, E., Keasling, J. D., and Lee, T. S. (2016). Isopentenyl diphosphate (IPP)-bypass mevalonate pathways for isopentenol production. *Metab. Eng.* 34, 25–35. doi: 10.1016/j.ymben.2015.12.002

Kang, A., Mendez-Perez, D., Goh, E. B., Baidoo, E. E. K., Benites, V. T., Beller, H. R., et al. (2019). Optimization of the IPP-bypass mevalonate pathway and fed-batch fermentation for the production of isoprenol in *Escherichia coli*. *Metab. Eng.* 56, 85–96. doi: 10.1016/j.ymben.2019.09.003

Karpievitch, Y. V., Polpitiya, A. D., Anderson, G. A., Smith, R. D., and Dabney, A. R. (2010). Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. *Ann. Appl. Stat.* 4, 1797–1823. doi: 10.1214/10-AOAS341

Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* 7, 296–307. doi: 10.1016/j.mib.2004.04.012

Kirpich, A. S., Ibarra, M., Moskalenko, O., Fear, J. M., Gerken, J., Mi, X., et al. (2018). SECIMTools: a suite of metabolomics data analysis tools. *BMC Bioinformatics* 19:151. doi: 10.1186/s12859-018-2134-1

Kleiner, M. (2019). Metaproteomics: much more than measuring gene expression in microbial communities. *mSystems* 4, 1–6. doi: 10.1128/msystems.00115-19

Kleiner, M., Dong, X., Hinzke, T., Wippler, J., Thorson, E., Mayer, B., et al. (2018). Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5576–E5584. doi: 10.1073/pnas.1722325115

Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., et al. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* 8:1558. doi: 10.1038/s41467-017-01544-x

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., et al. (2012). Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1173–1182. doi: 10.1073/pnas.1121198109

Koberstein, J. N. (2020). A sort-seq approach to the development of single fluorescent protein biosensors. *bioRxiv*. doi: 10.1101/2020.08.21.261578

Koh, H., Fermin, D., Choi, K. P., Ewing, R., and Choi, H. (2018). iOmicsPASS: a novel method for integration of multi-omics data over biological networks and discovery of predictive subnetworks. *bioRxiv* 374520. doi: 10.1101/374520

Kono, N., and Arakawa, K. (2019). Nanopore sequencing: review of potential applications in functional genomics. *Dev. Growth Differ.* 61, 316–326. doi: 10.1111/dgd.12608

Krisp, C., Yang, H., Van Soest, R., and Molloy, M. P. (2015). Online peptide fractionation using a multiphasic microfluidic liquid chromatography chip improves reproducibility and detection limits for quantitation in discovery and targeted proteomics. *Mol. Cell. Proteomics* 14, 1708–1719. doi: 10.1074/mcp.M114.046425

Kumar Awasthi, M., Ravindran, B., Sarsaiya, S., Chen, H., Wainaina, S., Singh, E., et al. (2020). Metagenomics for taxonomy profiling: tools and approaches. *Bioengineered* 11, 356–374. doi: 10.1080/21655979.2020.1736238

Kuo, W. P., Jenssen, T., Butte, A. J., Ohno-machado, L., and Kohane, I. S. (2002). Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 18, 405–412. doi: 10.1093/bioinformatics/18.3.405

Langley, R. J., Tipper, J. L., Bruse, S., Baron, R. M., Tsalik, E. L., Huntley, J., et al. (2014). Integrative "Omic" analysis of experimental bacteremia identi fi es a metabolic signature that distinguishes human sepsis from systemic inflammatory response *Syndromes*. 190, 445–455. doi: 10.1164/rccm.201404-0624OC

Langley, R. J., Tsalik, E. L., Velkinburgh, J. C. Van, Glickman, S. W., Rice, B. J., Wang, C., et al. (2013). An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci. Transl. Med.* 5:195ra95. doi: 10.1126/scitranslmed.3005893

Lee, S., and Kim, P. (2020). Current status and applications of adaptive laboratory evolution in industrial microorganisms. *J. Microbiol. Biotechnol.* 30, 793–803. doi: 10.4014/jmb.2003.03072

Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* 10, 291–305. doi: 10.1038/nrmicro2737

Li, C., Li, K., Li, K., Xie, X., and Lin, F. (2019). SWPepNovo: an efficient *de novo* peptide sequencing tool for large-scale MS/MS spectra analysis. *Int. J. Biol. Sci.* 15, 1787–1801. doi: 10.7150/ijbs.32142

Li, Y., Zhang, Z., Liu, X., Li, A., Hou, Z., Wang, Y., et al. (2015). A novel approach to the simultaneous extraction and non-targeted analysis of the small molecules metabolome and lipidome using 96-well solid phase extraction plates with column-switching technology. *J. Chromatogr. A* 1409, 277–281. doi: 10.1016/j.chroma.2015.07.048

Link, H., Fuhrer, T., Gerosa, L., Zamboni, N., and Sauer, U. (2015). Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nat. Methods* 12, 1091–1097. doi: 10.1038/nmeth.3584

Liu, H., Wang, Z., Xu, W., Zeng, J., Li, L., Li, S., et al. (2020). *Bacillus pumilus* LZP02 promotes rice root growth by improving carbohydrate metabolism and phenylpropanoid biosynthesis. *Mol. Plant Microbe Interact.* 33, 1222–1231. doi: 10.1094/MPMI-04-20-0106-R

Liu, X., Zhou, L., Shi, X., and Xu, G. (2019). New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *Trends Anal. Chem.* 121:115665. doi: 10.1016/j.trac.2019.115665

Liu, Y., and Nielsen, J. (2019). Recent trends in metabolic engineering of microbial chemical factories. *Curr. Opin. Biotechnol.* 60, 188–197. doi: 10.1016/j.copbio.2019.05.010

Long, C. P., and Antoniewicz, M. R. (2018). How adaptive evolution reshapes metabolism to improve fitness: recent advances and future outlook applications of adaptive laboratory evolution (ALE). *Curr. Opin. Chem. Eng.* 22, 209–215. doi: 10.1016/j.coche.2018.11.001

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567, 123–126. doi: 10.1038/s41586-019-0978-9

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. doi: 10.1093/bioinformatics/btq054

MacPherson, K. A., Starr, B., Wong, E. D., Dalusag, K. S., Hellerstedt, S. T., Lang, O. W., et al. (2017). Outreach and online training services at the saccharomyces genome database. *Database* 2017, 1–9. doi: 10.1093/database/bax002

Mahajan, K. R., and Ontaneda, D. (2017). The role of advanced magnetic resonance imaging techniques in multiple sclerosis clinical trials. *Neurotherapeutics* 14, 905–923. doi: 10.1007/s13311-017-0561-8

Manier, S. K., Keller, A., Schäper, J., and Meyer, M. R. (2019). Untargeted metabolomics by high resolution mass spectrometry coupled to normal and reversed phase liquid chromatography as a tool to study the *in vitro* biotransformation of new psychoactive substances. *Sci. Rep.* 9:2741. doi: 10.1038/s41598-019-39235-w

Marx, V. (2013). Targeted proteomics. *Nat. Methods* 10, 19–22. doi: 10.1038/nmeth.2285

Migaud, M., Gandotra, S., Chand, H. S., Gillespie, M. N., Thannickal, V. J., and Langley, R. J. (2020). Metabolomics to predict antiviral drug efficacy in COVID-19. *Am. J. Respir. Cell Mol. Biol.* 63, 396–398. doi: 10.1165/rcmb.2020-0206LE

Misra, B. B., Bassey, E., Bishop, A. C., Kusel, D. T., Cox, L. A., and Olivier, M. (2018). High-resolution gas chromatography/mass spectrometry metabolomics of non-human primate serum. *Rapid Commun. Mass Spectrom.* 32, 1497–1506. doi: 10.1002/rcm.8197

Misra, N., Panda, P. K., and Parida, B. K. (2013). Agrigenomics for microalgal biofuel production: an overview of various bioinformatics resources and recent studies to link omics to bioenergy and bioeconomy. *Omics* 17, 537–549. doi: 10.1089/omi.2013.0025

Mohanram, S., and Kumar, P. (2019). Rhizosphere microbiome: revisiting the synergy of plant-microbe interactions. *Ann. Microbiol.* 69, 307–320. doi: 10.1007/s13213-019-01448-9

Mousavi, F., Bojko, B., and Pawliszyn, J. (2015). Development of high throughput 96-blade solid phase microextraction-liquid chromatrography-mass spectrometry protocol for metabolomics. *Anal. Chim. Acta* 892, 95–104. doi: 10.1016/j.aca.2015.08.016

Mulligan, M. J., Lyke, K. E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* 586, 589–593. doi: 10.1038/s41586-020-2639-4

Nakamura, Y. (2018). Rice starch biotechnology: rice endosperm as a model of cereal endosperms. *Starch* 70, 1–20. doi: 10.1002/star.201600375

Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–1440. doi: 10.1074/mcp.R500012-MCP200

Nguyen, T. T. M., An, Y. J., Cha, J. W., Cha, J. W., Ko, Y. J., Lee, H., et al. (2020). Real-time in-organism NMR metabolomics reveals different roles of AMP-activated protein kinase catalytic subunits. *Anal. Chem.* 92, 7382–7387. doi: 10.1021/acs.analchem.9b05670

Nielsen, J., and Jewett, M. C. (2007). The role of metabolomics in systems biology. *Top. Curr. Genet.* 18, 1–10. doi: 10.1007/4735_2007_0228

Oksman-Caldentey, K.-M., and Saito, K. (2005). Integrating genomics and metabolomics for engineering plant metabolic pathways. *Curr. Opin. Biotechnol.* 16, 174–179. doi: 10.1016/j.copbio.2005.02.007

Opgenorth, P., Costello, Z., Okada, T., Goyal, G., Chen, Y., Gin, J., et al. (2019). Lessons from two design–build–test–learn cycles of dodecanol production in *Escherichia coli* aided by machine learning. *ACS Synth. Biol.* 8, 1337–1351. doi: 10.1021/acssynbio.9b00020

Oyetunde, T., Bao, F. S., Chen, J. W., Martin, H. G., and Tang, Y. J. (2018). Leveraging knowledge engineering and machine learning for microbial bio-manufacturing. *Biotechnol. Adv.* 36, 1308–1315. doi: 10.1016/j.biotechadv.2018.04.008

Pagani, I., Liolios, K., Jansson, J., Chen, I. M. A., Smirnova, T., Nosrat, B., et al. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 40, 571–579. doi: 10.1093/nar/gkr1100

Pan, M., and Barrangou, R. (2020). Combining omics technologies with CRISPR-based genome editing to study food microbes. *Curr. Opin. Biotechnol.* 61, 198–208. doi: 10.1016/j.copbio.2019.12.027

Pappireddi, N., Martin, L., and Wühr, M. (2019). A review on quantitative multiplexed proteomics. *Chembiochem* 20, 1210–1224. doi: 10.1002/cbic.201800650

Perfetto, L., Briganti, L., Calderone, A., Perpetuini, A. C., Iannuccelli, M., Langone, F., et al. (2016). SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* 44, D548–D554. doi: 10.1093/nar/gkv1048

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Phaneuf, P. V, Yurkovich, J. T., Heckmann, D., Wu, M., Sandberg, T. E., King, Z. A., et al. (2020). Causal mutations from adaptive laboratory evolution are outlined by multiple scales of genome annotations and condition-specificity. *BMC Genomics.* 21:514. doi: 10.1186/s12864-020-06920-4

Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J., et al. (2019). Systems biology and multi-omics integration: viewpoints from the metabolomics research community. *Metabolites* 9, 1–31. doi: 10.3390/metabo9040076

Ramirez-Gaona, M., Marcu, A., Pon, A., Guo, A. C., Sajed, T., Wishart, N. A., et al. (2017). YMDB 2.0: a significantly expanded version of the yeast metabolome database. *Nucleic Acids Res.* 45, D440–D445. doi: 10.1093/nar/gkw1058

Rathahao-Paris, E., Alves, S., Boussaid, N., Picard-Hagen, N., Gayrard, V., Toutain, P. L., et al. (2019). Evaluation and validation of an analytical approach for high-throughput metabolomic fingerprinting using direct introduction–high-resolution mass spectrometry: applicability to classification of urine of scrapie-infected ewes. *Eur. J. Mass Spectrom.* 25, 251–258. doi: 10.1177/1469066718806450

Ren, J. L., Zhang, A. H., Kong, L., and Wang, X. J. (2018). Advances in mass spectrometry-based metabolomics for investigation of metabolites. *RSC Adv.* 8, 22335–22350. doi: 10.1039/c8ra01574k

Rexroad, C., Vallet, J., Matukumalli, L. K., Reecy, J., Bickhart, D., Blackburn, H., et al. (2019). Genome to phenome: improving animal health, production, and well-being–a new USDA blueprint for animal genome research 2018–2027. *Front. Genet.* 10:327. doi: 10.3389/fgene.2019.00327

Roberts, L. D., Souza, A. L., Gerszten, R. E., and Clish, C. B. (2012). Targeted metabolomics. *Curr. Protoc. Mol. Biol.* 1, 1–24. doi: 10.1002/0471142727.mb3002s98

Robinson, C. J., Carbonell, P., Jervis, A. J., Yan, C., Hollywood, K. A., Dunstan, M. S., et al. (2020). Rapid prototyping of microbial production strains for the biomanufacture of potential materials monomers. *Metab. Eng.* 60, 168–182. doi: 10.1016/j.ymben.2020.04.008

Robinson, D. G., Chen, W., Storey, J. D., and Gresham, D. (2014). Design and analysis of bar-seq experiments. *G3 Genes Genomes Genet.* 4, 11–18. doi: 10.1534/g3.113.008565

Roh, K., Nerem, R. M., and Roy, K. (2016). Biomanufacturing of therapeutic cells: state of the art, current challenges, and future perspectives. *Annu. Rev. Chem. Biomol. Eng.* 7, 455–478. doi: 10.1146/annurev-chembioeng-080615-033559

Rohlhill, J., Sandoval, N. R., and Papoutsakis, E. T. (2017). Sort-seq approach to engineering a formaldehyde-inducible promoter for dynamically regulated *Escherichia coli* growth on methanol. *ACS Synth. Biol.* 6, 1584–1595. doi: 10.1021/acssynbio.7b00114

Roukos, D. H. (2012). Biotechnological, genomics and systems-synthetic biology revolution: Redesigning genetic code for a pragmatic systems medicine. *Expert Rev. Med. Devices* 9, 97–101. doi: 10.1586/erd.11.68

Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., and Watson, M. (2017). A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Front. Genet.* 8:23. doi: 10.3389/fgene.2017.00023

Rühl, M., Kühn, B., Roos, J., Maier, T. J., Steinhilber, D., and Karas, M. (2019). Elucidation of chemical modifier reactivity towards peptides and proteins and the analysis of specific fragmentation by matrix-assisted laser desorption/ionization collision-induced dissociation tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 33, 40–49. doi: 10.1002/rcm.8223

Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev.arplant.043008.092035

Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., et al. (2016). ECMDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Res.* 44, D495–D501. doi: 10.1093/nar/gkv1060

Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O., and Feist, A. M. (2019). The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* 56, 1–16. doi: 10.1016/j.ymben.2019.08.004

Savitski, M. M., Nielsen, M. L., and Zubarev, R. A. (2006). ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* 5, 935–948. doi: 10.1074/mcp.T500034-MCP200

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., van Ommen, B., et al. (2009). Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 5, 435–458. doi: 10.1007/s11306-009-0168-0

Schulz, S., Becker, M., Groseclose, M. R., Schadt, S., and Hopf, C. (2019). Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development. *Curr. Opin. Biotechnol.* 55, 51–59. doi: 10.1016/j.copbio.2018.08.003

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72.e15. doi: 10.1016/j.cell.2020.05.032

Simirgiotis, M. J., Quispe, C., Mocan, A., Villatoro, J. M., Areche, C., Bórquez, J., et al. (2017). UHPLC high resolution orbitrap metabolomic fingerprinting of the unique species *Ophryosporus triangularis* meyen from the atacama desert, Northern Chile. *Rev. Bras. Farmacogn.* 27, 179–187. doi: 10.1016/j.bjp.2016.10.002

Sirén, K., Mak, S. S. T., Fischer, U., Hansen, L. H., and Gilbert, M. T. P. (2019). Multi-omics and potential applications in wine production. *Curr. Opin. Biotechnol.* 56, 172–178. doi: 10.1016/j.copbio.2018.11.014

Smith, A. M., Heisler, L. E., Mellor, J., Kaper, F., Thompson, M. J., Chee, M., et al. (2009). Quantitative phenotyping via deep barcode sequencing. *Genome Res.* 19, 1836–1842. doi: 10.1101/gr.093955.109

Smith, A. M., Heisler, L. E., St.Onge, R. P., Farias-Hesson, E., Wallace, I. M., Bodeau, J., et al. (2010). Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* 38, 1–7. doi: 10.1093/nar/gkq368

Smith, J. C., and Figeys, D. (2006). Proteomics technology in systems biology. *Mol. Biosyst.* 2:364. doi: 10.1039/b606798k

Solovchenko, A. E., Ismagulova, T. T., Lukyanov, A. A., Vasilieva, S. G., Konyukhov, I. V., Pogosyan, S. I., et al. (2019). Luxury phosphorus uptake in microalgae. *J. Appl. Phycol.* 31, 2755–2770. doi: 10.1007/s10811-019-01831-8

Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmler, S., et al. (2018). Long read-based *de novo* assembly of low complex metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *bioRxiv* 1–18. doi: 10.1101/476747

Song, J. W., Lam, S. M., Fan, X., Cao, W. J., Wang, S. Y., Tian, H., et al. (2020). Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. *Cell Metab.* 32, 188–202.e5. doi: 10.1016/j.cmet.2020.06.016

Southam, A. D., Weber, R. J. M., Engel, J., Jones, M. R., and Viant, M. R. (2017). A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. *Nat. Protoc.* 12, 255–273. doi: 10.1038/nprot.2016.156

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2

Stettin, D., Poulin, R. X., and Pohnert, G. (2020). Metabolomics benefits from orbitrap GC-MS-comparison of low-and high-resolution GC-MS. *Metabolites* 10, 1–16. doi: 10.3390/metabo10040143

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14, 7–9. doi: 10.1177/1177932219899051

Sun, X., and Weckwerth, W. (2012). COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* 8, 81–93. doi: 10.1007/s11306-012-0399-3

Sun, X. M., Ren, L. J., Zhao, Q. Y., Ji, X. J., and Huang, H. (2018). Microalgae for the production of lipid and carotenoids: a review with focus on stress regulation and adaptation. *Biotechnol. Biofuels* 11:272. doi: 10.1186/s13068-018-1275-9

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tang, J. (2011). Microbial metabolomics. *Curr. Genomics* 12, 391–403. doi: 10.2174/138920211797248619

Tilocca, B., Britti, D., Urbani, A., and Roncada, P. (2020). Computational immune proteomics approach to target COVID-19. *J. Proteome Res.* 19, 4233–4241. doi: 10.1021/acs.jproteome.0c00553

Uranukul, B., Woolston, B. M., Fink, G. R., and Stephanopoulos, G. (2019). Biosynthesis of monoethylene glycol in *Saccharomyces cerevisiae* utilizing native glycolytic enzymes. *Metab. Eng.* 51, 20–31. doi: 10.1016/j.ymben.2018.09.012

Urgen Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80, 273–299. doi: 10.1146/annurev-biochem-061308-093216

van Doremalen, N., Lambe, T., Spencer, A., Belij-Rammerstorfer, S., Purushotham, J. N., Port, J. R., et al. (2020). ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature* 586, 578–582. doi: 10.1038/s41586-020-2608-y

Van Emon, J. M. (2016). The omics revolution in agricultural research. *J. Agric. Food Chem.* 64, 36–44. doi: 10.1021/acs.jafc.5b04515

Van Emon, J. M. (2019). *Omics in Fruit Nutrition: Concepts and Application*. Amsterdam: Elsevier Inc.

Vavricka, C. J., Hasunuma, T., and Kondo, A. (2020). Dynamic metabolomics for engineering biology: accelerating learning cycles for bioproduction. *Trends Biotechnol.* 38, 68–82. doi: 10.1016/j.tibtech.2019.07.009

VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., and Banfield, J. F. (2009). Systems biology: functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7, 196–205. doi: 10.1038/nrmicro2080

Volke, D. C., Rohwer, J., Fischer, R., and Jennewein, S. (2019). Investigation of the methylerythritol 4–phosphate pathway for microbial terpenoid production through metabolic control analysis. *Microb. Cell Fact.* 18, 1–15. doi: 10.1186/s12934-019-1235-5

Waegele, B., Dunger-Kaltenbach, I., Fobo, G., Montrone, C., Mewes, H. W., and Ruepp, A. (2009). CRONOS: the cross-reference navigation server. *Bioinformatics* 25, 141–143. doi: 10.1093/bioinformatics/btn590

Walker, C., Ryu, S., and Trinh, C. T. (2019). Exceptional solvent tolerance in *Yarrowia lipolytica* is enhanced by sterols. *Metab. Eng.* 54, 83–95. doi: 10.1016/j.ymben.2019.03.003

Wang, C., Cui, Y., and Qu, X. (2018). Mechanisms and improvement of acid resistance in lactic acid bacteria. *Arch. Microbiol.* 200, 195–201. doi: 10.1007/s00203-017-1446-2

Wang, J. H., Byun, J., and Pennathur, S. (2010). Analytical approaches to metabolomics and applications to systems biology. *Semin. Nephrol.* 30, 500–511. doi: 10.1016/j.semnephrol.2010.07.007

Wehrs, M., Thompson, M. G., Banerjee, D., Prahl, J. P., Morella, N. M., Barcelos, C. A., et al. (2020). Investigation of Bar-seq as a method to study population dynamics of *Saccharomyces cerevisiae* deletion library during bioreactor cultivation. *Microb. Cell Fact.* 19, 1–15. doi: 10.1186/s12934-020-01423-z

Wilmes, P., and Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920. doi: 10.1111/j.1462-2920.2004.00687.x

Wilmes, P., Heintz-Buschart, A., and Bond, P. L. (2015). A decade of metaproteomics: where we stand and what the future holds. *Proteomics* 15, 3409–3417. doi: 10.1002/pmic.201500183

Wilson, S. R., Vehus, T., Berg, H. S., and Lundanes, E. (2015). Nano-LC in proteomics: recent advances and approaches. *Bioanalysis* 7, 1799–1815. doi: 10.4155/bio.15.92

Wolters, D. A., Washburn, M. P., and Yates, J. R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73, 5683–5690. doi: 10.1021/ac010617e

Wu, T., Liu, J., Li, M., Zhang, G., Liu, L., Li, X., et al. (2020). Improvement of sabinene tolerance of *Escherichia coli* using adaptive

laboratory evolution and omics technologies. *Biotechnol. Biofuels* 13:79. doi: 10.1186/s13068-020-01715-x

Yan, Q., and Fong, S. S. (2018). Increasing carbon source uptake rates to improve chemical productivity in metabolic engineering. *Curr. Opin. Biotechnol.* 53, 254–263. doi: 10.1016/j.copbio.2018.06.001

Yang, F., Xiang, W., Li, T., and Long, L. (2018). Transcriptome analysis for phosphorus starvation-induced lipid accumulation in *Scenedesmus* sp. *Sci. Rep.* 8:16420. doi: 10.1038/s41598-018-34650-x

Yang, S., Synovec, R. E., Kalyuzhnaya, M. G., and Lidstrom, M. E. (2011). Development of a solid phase extraction protocol coupled with liquid chromatography mass spectrometry to analyze central carbon metabolites in lake sediment microcosms. *J. Sep. Sci.* 34, 3597–3605. doi: 10.1002/jssc.201100533

Yates, J. R. (2019). Recent technical advances in proteomics. *F1000Research* 8, 1–8. doi: 10.12688/f1000research.16987.1

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., and Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* 26, 255–260. doi: 10.1093/bioinformatics/btq183

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., Nunes, U., Shi, S., et al. (2018). substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3, 470–480. doi: 10.1038/s41564-018-0129-3

Zhang, Y. H. P., Sun, J., and Ma, Y. (2017). Biomanufacturing: history and perspective. *J. Ind. Microbiol. Biotechnol.* 44, 773–784. doi: 10.1007/s10295-016-1863-2

Zhao, J., Wang, G., Chu, J., and Zhuang, Y. (2020). Harnessing microbial metabolomics for industrial applications. *World J. Microbiol. Biotechnol.* 36:1. doi: 10.1007/s11274-019-2775-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 08 March 2021
doi: 10.3389/fbioe.2021.644216

Re-routing of Sugar Catabolism Provides a Better Insight Into Fungal Flexibility in Using Plant Biomass-Derived Monomers as Substrates

*Tania Chroumpi[1], Mao Peng[1], Lye Meng Markillie[2], Hugh D. Mitchell[2], Carrie D. Nicora[2], Chelsea M. Hutchinson[2], Vanessa Paurus[2], Nikola Tolic[2], Chaevien S. Clendinen[2], Galya Orr[2], Scott E. Baker[2], Miia R. Mäkelä[1,3] and Ronald P. de Vries[1]**

[1] Fungal Physiology, Westerdijk Fungal Biodiversity Institute & Fungal Molecular Physiology, Utrecht University, Utrecht, Netherlands, [2] Environmental Molecular Science Laboratory, Pacific Northwest National Laboratory, Richland, WA, United States, [3] Department of Microbiology, University of Helsinki, Helsinki, Finland

OPEN ACCESS

Edited by:
Maurycy Daroch,
Peking University, China

Reviewed by:
Limin Cao,
Capital Normal University, China
Min Jiang,
Nanjing Tech University, China

***Correspondence:**
Ronald P. de Vries
r.devries@wi.knaw.nl

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 20 December 2020
Accepted: 16 February 2021
Published: 08 March 2021

Citation:
Chroumpi T, Peng M, Markillie LM, Mitchell HD, Nicora CD, Hutchinson CM, Paurus V, Tolic N, Clendinen CS, Orr G, Baker SE, Mäkelä MR and de Vries RP (2021) Re-routing of Sugar Catabolism Provides a Better Insight Into Fungal Flexibility in Using Plant Biomass-Derived Monomers as Substrates. Front. Bioeng. Biotechnol. 9:644216. doi: 10.3389/fbioe.2021.644216

The filamentous ascomycete *Aspergillus niger* has received increasing interest as a cell factory, being able to efficiently degrade plant cell wall polysaccharides as well as having an extensive metabolism to convert the released monosaccharides into value added compounds. The pentoses D-xylose and L-arabinose are the most abundant monosaccharides in plant biomass after the hexose D-glucose, being major constituents of xylan, pectin and xyloglucan. In this study, the influence of selected pentose catabolic pathway (PCP) deletion strains on growth on plant biomass and re-routing of sugar catabolism was addressed to gain a better understanding of the flexibility of this fungus in using plant biomass-derived monomers. The transcriptome, metabolome and proteome response of three PCP mutant strains, Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA*, grown on wheat bran (WB) and sugar beet pulp (SBP), was evaluated. Our results showed that despite the absolute impact of these PCP mutations on pure pentose sugars, they are not as critical for growth of *A. niger* on more complex biomass substrates, such as WB and SBP. However, significant phenotypic variation was observed between the two biomass substrates, but also between the different PCP mutants. This shows that the high sugar heterogeneity of these substrates in combination with the high complexity and adaptability of the fungal sugar metabolism allow for activation of alternative strategies to support growth.

Keywords: lignocellulosic substrates, pentose catabolic pathway, D-galacturonic acid catabolic pathway, L-rhamnose catabolic pathway, wheat bran, sugar beet pulp, CAZymes, *Aspergillus niger*

INTRODUCTION

The majority of industrial processes for the production of chemicals, materials, and energy are still based on fossil fuels, especially coal and crude oil. However, to gain independence from these raw materials, more consideration has been given in the last decades to the use of renewable materials and agricultural residues as promising low-cost feedstocks for obtaining high added-value products.

The filamentous fungus *Aspergillus niger* is one of the most prominent fungal cell factories used in biotechnology. It is known for its ability to naturally degrade complex plant biomass polysaccharides, including both cell wall (cellulose, hemicellulose and pectin) and storage (inulin and starch) components, into simple sugars using a rich arsenal of Carbohydrate-Active Enzymes (CAZymes) (de Vries and Visser, 2001; Lombard et al., 2014; Benoit et al., 2015). Despite the complexity of the polysaccharides forming the cell wall, their backbone is mainly formed by simple sugars, such as D-glucose, D-xylose, L-arabinose, D-galactose, D-galacturonic acid, D-fructose and L-rhamnose (Somerville, 2006; Mohnen, 2008; Scheller and Ulvskov, 2010; Ochoa-Villarreal et al., 2012). In nature, fungi need to first recognize the plant biomass components to produce the right set of CAZymes that can break down the complex structures into these simple molecules. The resulting sugars are subsequently transported into the cell and converted into energy and intermediate metabolites through a wide range of metabolic pathways (Khosravi et al., 2015). An in-depth understanding of the *A. niger* metabolic network will provide a detailed blueprint for the metabolic engineering of this fungus to improve productivity of a broad range of proteins and metabolites.

The pentoses L-arabinose and D-xylose are the most abundant monosaccharides in nature after D-glucose, being major components of the hemicelluloses xylan and xyloglucan, and of pectin (Seiboth and Metz, 2011). In most fungi, L-arabinose and D-xylose are metabolized through the pentose catabolic pathway (PCP) (Witteveen et al., 1989), through oxidation, reduction and phosphorylation reactions to finally form D-xylulose-5-phosphate, which enters the pentose phosphate pathway (PPP) (Seiboth and Metz, 2011; **Figure 1A**). Although pentose catabolism is among the best studied pathways of *A. niger* primary carbon metabolism, the simplistic view of this pathway has recently been challenged (Chroumpi et al., 2021). Due to the residual growth of the PCP single deletion mutants, identification of additional genes involved in pentose catabolism was achieved: a second D-xylose reductase (XyrB), a second L-xylulose reductase (LxrB) and the role of sorbitol dehydrogenase (SdhA) in compensating for the loss of L-arabitol dehydrogenase (LadA) and xylitol dehydrogenase (XdhA) (**Figure 1A**). Additionally, all enzymatic steps of the PCP in *A. niger*, apart from the last one, were shown to be catalyzed by multiple enzymes, which together ensure efficient conversion of pentose sugars.

In this study, the influence of selected PCP deletion strains on growth on plant biomass and re-routing of sugar catabolism was analyzed to gain a better understanding of the flexibility of this fungus in using plant biomass-derived monomers as substrates. The transcriptome, metabolome and proteome responses of three pentose catabolic mutant strains, Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA*, grown on two plant biomass substrates, wheat bran (WB) and sugar beet pulp (SBP), were evaluated. These mutants have been previously shown to block pentose catabolism at different steps of the pathway and as a result accumulate different intermediates that could act as inducers (Chroumpi et al., 2021). While both substrates contain cellulose, WB is rich in arabinoxylan and SBP contains xyloglucan and pectin (**Figure 1C**). This means that both substrates contain considerable amounts of L-arabinose and D-xylose, making them highly suitable for analysis of this pathway.

MATERIALS AND METHODS

Strains, Media, and Growth Conditions

The *A. niger* strains used in this study are listed in **Table 1**. The strains were grown at 30°C using Minimal Medium (MM, pH 6) or Complete Medium (CM, pH 6) with the appropriate carbon source (de Vries et al., 2004). For solid cultivation, 1.5% (w/v) agar was added in the medium and, unless stated otherwise, all agar plates contained 1% D-glucose as carbon source. When required, media of auxotrophic strains were supplemented with 1.22 g/L uridine.

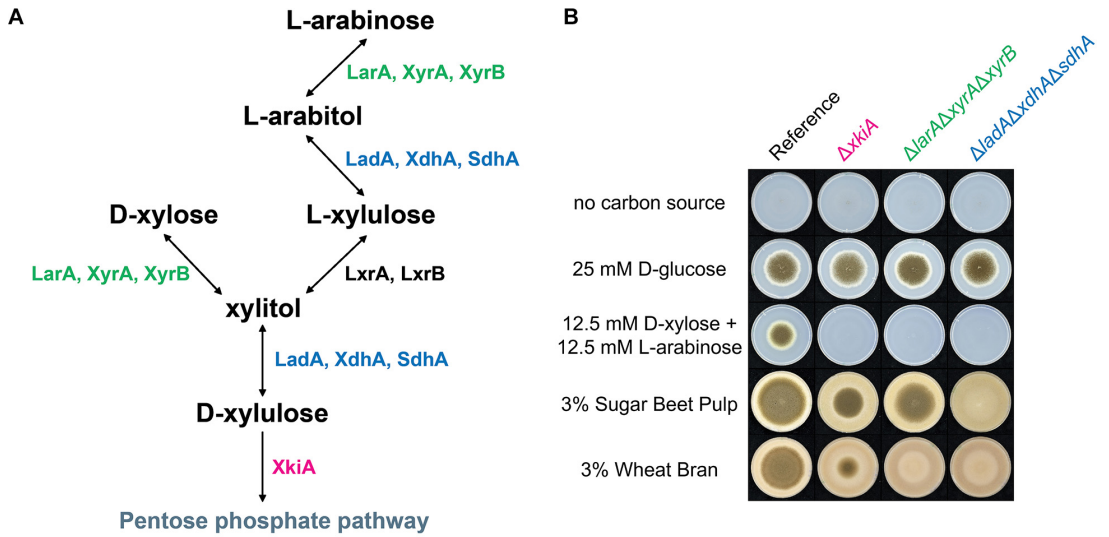
For growth profiling, 6 cm petri dishes with vents containing MM agar supplemented with 25 mM D-glucose (Sigma, G8270), a mixture of 12.5 mM D-xylose (Sigma, 95729) and 12.5 mM L-arabinose (Sigma, A3256), 3% wheat bran (WB) or 3% sugar beet pulp (SBP) were used. The monosaccharide composition analysis of WB and SBP is presented in **Figure 1C**. Spores were harvested from CM agar plates in ACES buffer, after five days of growth, and counted using a hemocytometer. Growth profiling plates were inoculated with 1,000 spores in 2 μl, and incubated at 30°C for 5 days.

All liquid cultures were incubated in an orbital shaker at 250 rpm and 30°C. For transfer experiments, the pre-cultures containing 250 mL CM with 2% D-fructose in 1 L Erlenmeyer flasks were inoculated with 10^6 spores/mL and incubated for 16 h. Subsequently, the mycelia were harvested by filtration on sterile cheesecloth, washed with MM and ∼0.5 g (dry weight) was transferred to 250 mL Erlenmeyer flasks containing 50 mL MM supplemented with 1% WB or 1% SBP. All cultures were performed in biological triplicate as were all the subsequent analyses. After 2, 8, and 24 h of incubation, the mycelia were harvested by vacuum filtration, dried between tissue paper and frozen in liquid nitrogen. Culture filtrates were also harvested for extracellular metabolomics and proteomics analysis. All samples were stored at −80°C until being processed.

Transcriptome Sequencing and Analysis

The transcriptomic response of the reference strain and the PCP deletion mutants induced after 2, 8, and 24 h on 1% WB or 1% SBP was analyzed using RNA-seq analysis. Total RNA was extracted from ground mycelial samples using TRIzol® reagent (Invitrogen) and purified with the NucleoSpin® RNA Clean-up Kit (Macherey-Nagel), while contaminant gDNA was removed by rDNase treatment directly on the silica membrane. The RNA quality and quantity were analyzed with a RNA6000 Nano Assay using the Agilent 2100 Bioanalyzer (Agilent Technologies). Purification of mRNA, synthesis of cDNA library and sequencing were conducted at the Environmental Molecular Sciences Laboratory (EMSL).

RNA samples were assessed using the Agilent 2100 Bioanalyzer. The TruSeq stranded mRNA (cat#20020594)



C

	Rha	Ara	Xyl	Man	Gal	Glc	GalUA	Total	polysaccharides
Wheat Bran (WB)	0.0	**16.5**	**34.6**	1.4	1.7	42.5	3.3	53.7	cellulose, arabinoxylan
Sugar Beet Pulp (SBP)	1.5	**29.0**	2.4	2.1	6.5	32.0	**27.0**	56.0	cellulose, pectin, xyloglucan

FIGURE 1 | (A) Pentose catabolic pathway (PCP) in *Aspergillus niger*. LarA = L-arabinose reductase, LadA = L-arabitol dehydrogenase, LxrA and LxrB = L-xylulose reductases, SdhA = sorbitol dehydrogenase, XyrA and XyrB = D-xylose reductases, XdhA = xylitol dehydrogenase, XkiA = D-xylulose kinase. **(B)** Growth profile of the *A. niger* reference strain (N593 ΔkusA) and the PCP deletion mutants on solid MM with or without addition of carbon source. Strains were grown for 5 days at 30°C and **(C)** composition analysis of polymeric carbon sources, wheat bran (WB) and sugar beet pulp (SBP), used for growth profile and multi-omics analysis of PCP mutants. Rha = rhamnose, Ara = arabinose, Xyl = xylose, Man = mannose, Gal = galactose, Glc = glucose, GalUA = galacturonic acid. Concentration in mol%.

was used to generate cDNA library for illumina NextSeq550 platform according to the manufacture protocol. Single-read sequencing of the cDNA libraries with a read length of 150 was performed with NextSeq 500 Sequencing System using NextSeq 500/550 High Output v2 kit 150 cycles (cat#20024907). Data quality was assessed using FastQC[1]. Reads were aligned to the *A. niger* NRRL 3 genome (Aguilar-Pontes et al., 2018) using bowtie2[2], with parameters -local, –sensitive-local. The RNAseq data set was deposited at the GEO (Barrett et al., 2012) database under the accession number GSE162901. Reads were aligned to genes using HTSeq−count (Anders et al., 2014) with parameters -a = 1, –mode = "union". The analysis was performed on three independent biological replicates. Differential gene expression was assessed using the R package DESeq2 (Love et al., 2014), with all subsequent analysis performed in R unless otherwise stated. Transcripts were considered differentially expressed if the DESeq2 fold change was >2 or <0.5 and Padj < 0.01. Transcripts with FPKM ≤ 50 were considered lowly (i.e., not substantially) expressed. MDS plots were also generated using DESeq2.

The Gene Ontology (GO) annotation was retrieved from JGI MycoCosm database[3] and the Gene Ontology (GO) annotation database from R Bioconductor was used to map their ancestor nodes in the GO hierarchy. The GO Slim terms defined in AspGD[4] were selected for enrichment analysis. The GO biological process terms enriched within the significant differentially expressed gene lists compared to the genome background were detected by a hypergeometric distribution model calculated with in-house script. The *P*-values for multiple tests were corrected with Benjamini and Hochberg's method, and significantly enriched GO terms were selected with *P*-values <0.01.

Proteomics Data Generation and Analysis

Equivalent volumes of culture supernatant were extracted using the MPLEx protocol (Nakayasu et al., 2016). The protein interlayer from the extraction was then resuspended in an 8 M urea solution, reduced with DTT, digested with Trypsin, put through C18 SPE for clean-up, and diluted to 0.1 $\mu g\ \mu L^{-1}$ for LC-MS/MS.

MS analysis was performed using a Q−Exactive Plus mass spectrometer (Thermo Scientific) outfitted with a homemade nano−electrospray ionization interface. Electrospray emitters were homemade using 150 µm o.d. × 20 µm i.d. chemically etched fused silica (Kelly et al., 2006). The ion transfer tube temperature and spray voltage were 250°C and 2.2 kV, respectively. Data were collected for 120 min following a 10 min delay after completion of sample trapping and start of gradient. FT−MS spectra were acquired from 300 to 1,800 m/z at a

[1]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[2]http://bowtiebio.sourceforge.net/bowtie2/index.shtml

[3]https://genome.jgi.doe.gov/Aspni_NRRL3_1/Aspni_NRRL3_1.home.html

[4]http://www.aspgd.org/

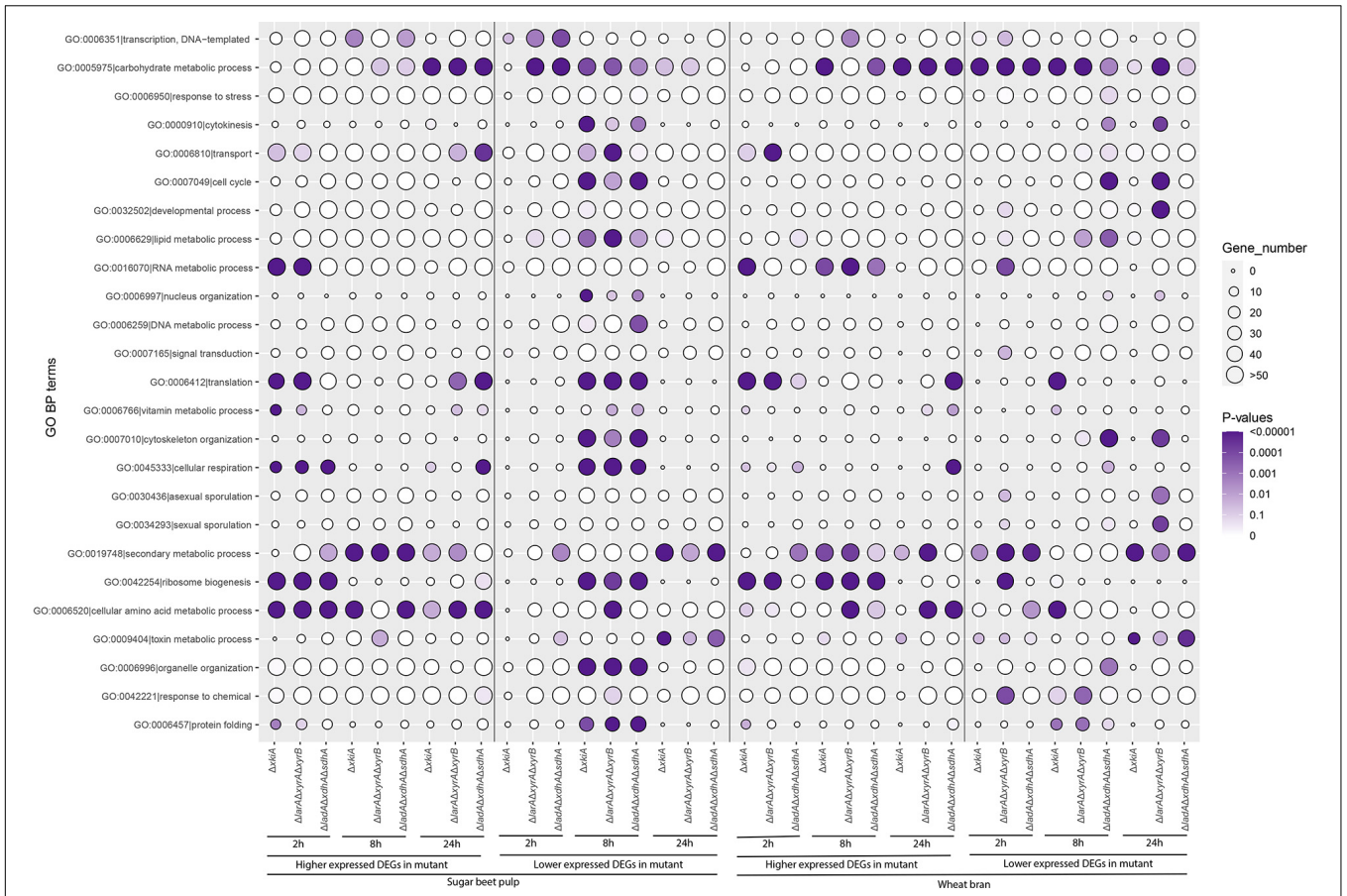


FIGURE 2 | Gene Ontology (GO) terms associated with the function of genes upregulated or downregulated during growth on sugar beet pulp (SBP) and wheat bran (WB) in the *A. niger* reference strain (N593 Δ*kusA*) and the PCP deletion mutants. The size and color of the circles represent the number of genes and statistical significance of enriched GO terms, respectively.

TABLE 1 | *A. niger* strains used in this study.

Strains	Gene ID	Enzyme activity	CBS number	Genotype	References
Reference strain (N593 Δ*kusA*)	–	–	CBS 138852	*A. niger* N593, *cspA1*, *kusA*::*amdS*, *pyrG*⁻	Meyer et al., 2007
Δ*larA* Δ*xyrA* Δ*xyrB*	NRRL3_10050 (*larA*) NRRL3_01952 (*xyrA*) NRRL3_10868 (*xyrB*)	L-arabinose/ D-xylose reductase	CBS 144530	*A. niger* N593, *cspA1*, *kusA*::*amdS*, *pyrA*⁻, ***larA*⁻, *xyrA*⁻, *xyrB*⁻**	Chroumpi et al., 2021
Δ*ladA* Δ*xdhA* Δ*sdhA*	NRRL3_02523 (*ladA*) NRRL3_09204 (*xdhA*) NRRL3_04328 (*sdhA*)	L-arabitol/xylitol dehydrogenase	CBS 144672	*A. niger* N593, *cspA1*, *kusA*::*amdS*, *pyrA*⁻, ***ladA*⁻, *xdhA*⁻, *sdhA*⁻**	Chroumpi et al., 2021
Δ*xkiA*	NRRL3_04471 (*xkiA*)	xylulokinase	CBS 144042	*A. niger* N593, *cspA1*, *kusA*::*amdS*, *pyrA*⁻, ***xkiA*⁻**	Chroumpi et al., 2021

resolution of 30 k (AGC target 3e6) and while the top 12 FT−HCD−MS/MS spectra were acquired in data−dependent mode with an isolation window of 1.5 m/z and at a resolution of 17.5 k (AGC target 1e5) using a normalized collision energy of 30 s exclusion time.

Generated MS/MS spectra were searched using the mass spectral generating function (MSGF +) algorithm (Kim et al., 2008; Kim and Pevzner, 2014) against the *A. niger* translated genome sequence available from Aspni_NRRL3_1 (Aguilar-Pontes et al., 2018). MSGF + was used in target/decoy mode with 20 ppm parent ion tolerance, partial tryptic rule and methionine oxidation (+ 15.9949) as dynamic modification. Best matches from the MSGF + searches were filtered at 1% FDR and only protein specific peptides were used in consequent aggregation and quantitative analysis. Relative peptide abundances can be determined by calculating the

area under the curve of the peptide ion peak in the MS measurement. This was accomplished using MASIC software (Monroe et al., 2008)[5] and results were aggregated using MS SQL (Microsoft) database. InfernoRDN software (Polpitiya et al., 2008)[6] was used to transform peptides abundances (log2) and perform mean central tendency normalization. Protein grouped normalized peptide abundances were de-logged, summed, transformed (log2) and normalized again in InfernoRDN to produce normalized abundances for the protein level roll-up. For an identified protein to be considered differentially produced, the requirements were a fold change of the mean intensity values of >2 or <0.5 and Padj < 0.05 from a two−tailed t −test of the log2 transformed intensity values. Note that where an intensity value was not detected for a protein in a sample, a zero value was used. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the data set identifier (PXD023205).

Metabolomics Data Generation and Analysis

Dried metabolite extracts from samples were derivatized using a modified version of the protocol used to create FiehnLib (Fiehn, 2016). Samples underwent methoximation to protect carbonyl groups and reduce tautomeric isomers, followed by silylation with *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide and 1% trimethylchlorosilane (MSTFA) to derivatize hydroxy and amine groups to trimethylsilated (TMS) forms. GC/MS data were collected over a mass range of 50–550 m/z using an Agilent GC 7890A coupled with a single quadrupole MSD 5975C (Agilent Technologies). A standard mixture of fatty acid methyl esters (FAMEs) (C8-C28) was analyzed with samples for RI alignment. The GC oven was held at 60°C for 1 min after injection, followed by a temperature increase of 10°C min^{-1} to a maximum of 325°C at which point it was held for 5 min.

Agilent.D files were converted to netCDF format using Agilent Chemstation. GC-MS raw data files were converted to binary files and processed using MetaboliteDetector software (version 2.5 beta) (Hiller et al., 2009). Retention indices (RIs) of detected metabolites were calculated based on analysis of the Fatty acid Methyl Esters standard mixture followed by chromatographic deconvolution and alignment. Metabolites were initially identified by matching experimental spectra to an augmented version of FiehnLib (Kind et al., 2009). All metabolite identifications were manually validated with the NIST 14 GC–MS library. The summed abundances of the three most abundant fragment ions of each identified metabolite were integrated across the GC elution profile (automatically determined by MetaboliteDetector). Fragment ions due to trimethylsilylation (that is, m/z 73 and 147) were excluded from the determination of metabolite abundance. Features resulting from GC column bleeding were removed from the data before further data processing and analysis.

[5]https://github.com/PNNL-Comp-Mass-Spec/MASIC/releases

[6]https://github.com/PNNL-Comp-Mass-Spec/InfernoRDN/releases

RESULTS

The Different *A. niger* PCP Deletion Mutants Cause Significant Phenotypic Variation on Lignocellulosic Biomass Substrates

Selected PCP gene deletion mutants that block conversion of both pentoses at different pathway steps (**Figure 1A**), and thus result in accumulation of different PCP intermediates, were grown on a mixture of the monosaccharides L-arabinose and D-xylose, and on the biomass substrates WB and SBP (**Figure 1B**). As expected, all three Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA* mutants were unable to grow on the pentose mixture, while these deletions resulted in reduced growth on WB and SBP, compared to the reference strain (**Figure 1B**). The extent of the growth reduction depended on the mutant strain and the substrate. The Δ*larA*Δ*xyrA*Δ*xyrB* mutant, which blocks the first step of pentose conversion, was practically unable to grow on WB, but showed only a small growth reduction on SBP. In contrast, growth of Δ*ladA*Δ*xdhA*Δ*sdhA* was similarly affected as that of the triple reductase mutant on WB, but was nearly abolished on SBP. Finally, the growth of Δ*xkiA* mutant, was reduced compared to the reference strain, but not abolished on both tested biomass substrates.

The rescued growth of Δ*xkiA* mutant on both biomass substates and of Δ*larA*Δ*xyrA*Δ*xyrB* mutant on SBP (**Figure 1B**), could suggest the expression of genes encoding alternative kinases and reductases, respectively, with sufficient specificity for the accumulated PCP intermediates to support growth. Following these observations, the re-routing of sugar metabolism in order to support growth of these PCP mutants on WB and SBP was further analyzed by multi-omics analysis. Mycelia of the reference strain and the PCP mutants were transferred to WB and SBP, and both mycelial and supernatant samples were harvested after 2, 8 and 24 h.

The PCP Deletions Affect the Transcriptome Abundance of Metabolic and CAZy Genes on WB and SBP

GO enrichment analysis of the expression data of the PCP mutants revealed a significant effect of these mutations on both primary and secondary metabolic responses of the fungus (**Figure 2**). During growth on both substrates, the expression of genes particularly involved in carbohydrate metabolic processes (GO:0005975), but also in cellular amino acid metabolic processes (GO:0006520) and in ribosome biogenesis (GO:0042254) was elevated compared to the reference strain. However, transport processes (GO:0006810) and metabolic processes do not seem to be synchronized. This is in line with the results of a previous study (Mäkelä et al., 2018), where sugar transport and metabolism were shown not to be co-regulated during growth of *A. niger* in liquid cultures. Interestingly, after 8 h of growth on SBP, an overall repression of genes involved in most of the studied GO terms was observed for all three mutants.

In this study, in order to investigate the adaptation strategy of each PCP mutant to these biomass substrates, we mainly focused on the genes encoding carbon catabolic enzymes and CAZymes involved in plant biomass degradation.

The Upregulation of the D-Galacturonic Acid and L-Rhamnose Catabolic Pathway Genes Could Partly Explain the Rescued Growth of the Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*xkiA* Mutants on SBP

Deletion of the PCP genes in the mutant strains led to altered expression of the remaining PCP genes on WB and SBP. On both biomass substrates, the expression of the remaining PCP genes increased after 8 h in all strains (**Figures 3A,B** and **Supplementary Table 1**). This is probably due to the accumulation of pentoses and polyols that have been previously indicated as potential inducers of the AraR and XlnR transcriptional activators of the PCP genes (de Vries, 2003; de Groot et al., 2007; Battaglia et al., 2011a,b). However, their expression in the reference and the Δ*xkiA* strains strongly reduced after 24 h on both biomass substrates, while they remained at significantly high levels in Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*ladA*Δ*xdhA*Δ*sdhA* (**Figures 3A,B**). Apart from the reference strain, the Δ*xkiA* mutant was the only strain that was still able to grow on both WB and SBP (**Figure 1B**). This indicates that the depletion of the pentose sugars and pathway intermediates under these conditions might be the reason of the observed reduction in expression of the PCP genes in these strains (**Figure 4**). The depletion of the pentose sugars in the Δ*xkiA* mutant could be explained by the presence of enzymes with some kinase activity on D-xylulose, of which the corresponding genes are induced under these conditions.

Similar to the PCP genes, the expression of genes involved in other carbon metabolic pathways was also affected. In particular, the expression of the genes involved in the catabolism of D-galacturonic acid was also significantly upregulated in the PCP deletion mutants compared to the reference strain on SBP (**Figure 3D** and **Supplementary Table 1**). On WB, which does not contain substantial amounts of D-galacturonic acid (**Figure 1C**), expression of the D-galacturonic acid catabolic pathway (GACP) genes was not induced (**Figure 3C** and **Supplementary Table 1**). Only the expression of *larA/gaaD*, encoding the enzyme involved in the last step of the GACP, was strongly induced on both substrates, since it is the same enzyme involved in the first step of L-arabinose metabolism in *A. niger* (Martens-Uzunova and Schaap, 2008; Mojzita et al., 2010). The expression of this gene was higher in the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant on both substrates (**Figures 3C,D** and **Supplementary Table 1**) showing that its induction is mainly a result of L-arabitol accumulation in this strain.

Finally, the expression of the genes involved in L-rhamnose catabolism was also upregulated on SBP (**Figure 3E,F** and **Supplementary Table 1**), while the absence of L-rhamnose in WB (**Figure 1C**) resulted in no expression of the L-rhamnose catabolic pathway (RCP) genes on this substrate (**Figure 3C**). In contrast to the other pathway genes, significant upregulation of *lkaA*, which was previously shown to be involved in the last step of L-rhamnose catabolism (Chroumpi et al., 2021), was observed on both substrates after 8 h of incubation.

Interestingly, albeit both biomass substrates are rich in D-glucose, increased expression of the glycolytic genes was not observed for any of the PCP mutants (data not shown).

The Rescued Growth of the Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*xkiA* Mutants on SBP Relies in Activation of Different Carbon Catabolic Re-routing Strategies

Similar to recently results on pure pentose sugars (Chroumpi et al., 2021), no significant accumulation of arabinose, xylose or other PCP intermediates were observed during growth on WB and SBP in the reference strain (**Figure 4**). Since the PCP remains intact in this strain, the released pentose sugars can efficiently be catabolized and used to support growth.

In the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, accumulation of arabinose and xylose occurred after 8 and 24 h of growth on SBP and WB (**Figure 4**). Although the Δ*larA*Δ*xyrA*Δ*xyrB* mutant could grow on SBP, the amount of accumulated pentose sugars increased with time. This observation supports our previous conclusion that this mutant cannot utilize the pentose sugars for growth. The absence of polyol accumulation also suggests that under these conditions no alternative enzymes are induced which are able to convert arabinose and xylose into their respective polyols. As expected, accumulation of D-galacturonic acid was also observed in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant at the later time points on SBP (**Figure 4**), due to the fact that the last step of the pathway is also disrupted after deletion of *larA*/*gaaD*. However, the growth of this mutant on SBP indicates the activation of other catabolic pathways that allow its adaptation under these conditions. Interestingly, glucose was shown to be significantly reduced after 8 and 24 h on SBP (**Figure 4**), indicating that it might be used as an alternative carbon source to support growth of the Δ*larA*Δ*xyrA*Δ*xyrB* mutant.

In the Δ*xkiA* mutant, which was able to grow on both biomass substrates (**Figure 1B**), the accumulated arabinose and xylose observed after 8 h of growth on SBP and WB were depleted after 24 h (**Figure 4**). In this mutant, a similar consumption pattern was also observed for the accumulated arabitol and xylitol. These observations again support our previous hypothesis for the presence of alternative enzymes induced under these conditions, which may facilitate the conversion of pentose sugars and of PCP intermediates in *A. niger*. The limited presence of the PCP inducers can also justify the reduced expression of the PCP genes after 24 h on both substrates (**Figures 3A,B**).

Finally, in the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant, the accumulated arabinose and xylose measured after 8 h of growth on SBP and WB were also depleted after 24 h (**Figure 4**). However, in contrast to the Δ*xkiA* mutant, significant accumulation of their respective polyols followed the depletion of the pentose sugars, showing that this was not a result of their use to support growth but their conversion into further downstream PCP

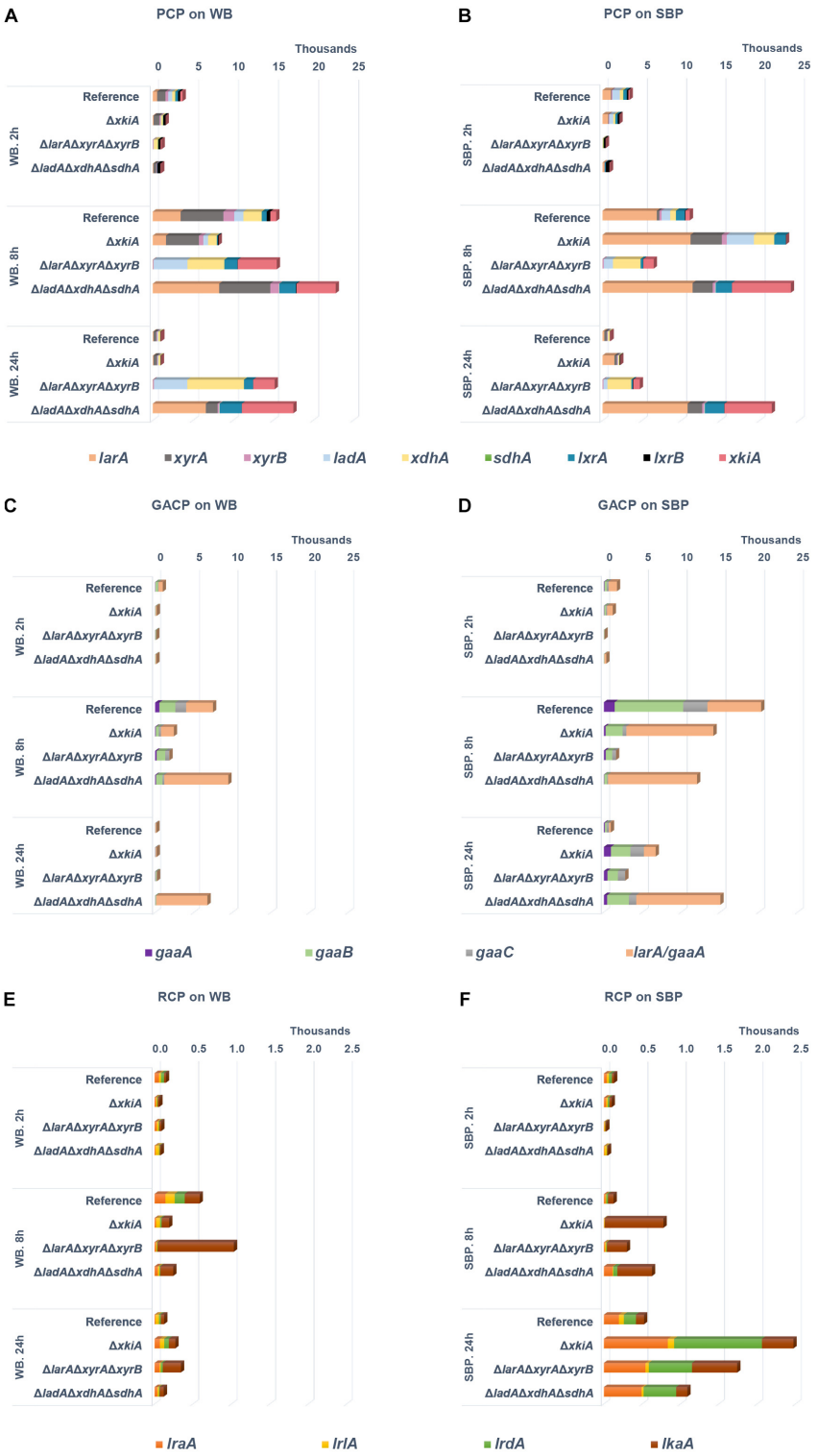


FIGURE 3 | Expression levels (FPKM) of catabolic genes during growth on **(A,C,E)** wheat bran (WB) and **(B,D,F)** sugar beet pulp (SBP) in the *A. niger* reference strain (N593 Δ*kusA*) and the PCP deletion mutants. PCP, pentose catabolic pathway; GACP, D-galacturonic acid catabolic pathway; RCP, L-rhamnose catabolic pathway. Genes encoding L-arabinose reductase (*larA*), L-arabitol dehydrogenase (*ladA*), L-xylulose reductases (*lxrA* and *lxrB*), sorbitol dehydrogenase (*sdhA*), D-xylose reductases (*xyrA* and *xyrB*), xylitol dehydrogenase (*xdhA*), D-xylulose kinase (*xkiA*), D-galacturonic acid reductase (*gaaA*), L-galactonate dehydratase (*gaaB*), 2-keto-3-deoxy- L-galactonate aldolase (*gaaC*), L-glyceraldehyde reductase (*larA/gaaD*), L-rhamnose-1-dehydrogenase (*lraA*), L-rhamnono-γ-lactonase (*lrlA*), L-rhamnonate dehydratase (*lrdA*), L-2-keto-3-deoxyrhamnonate aldolase (*lkaA*).

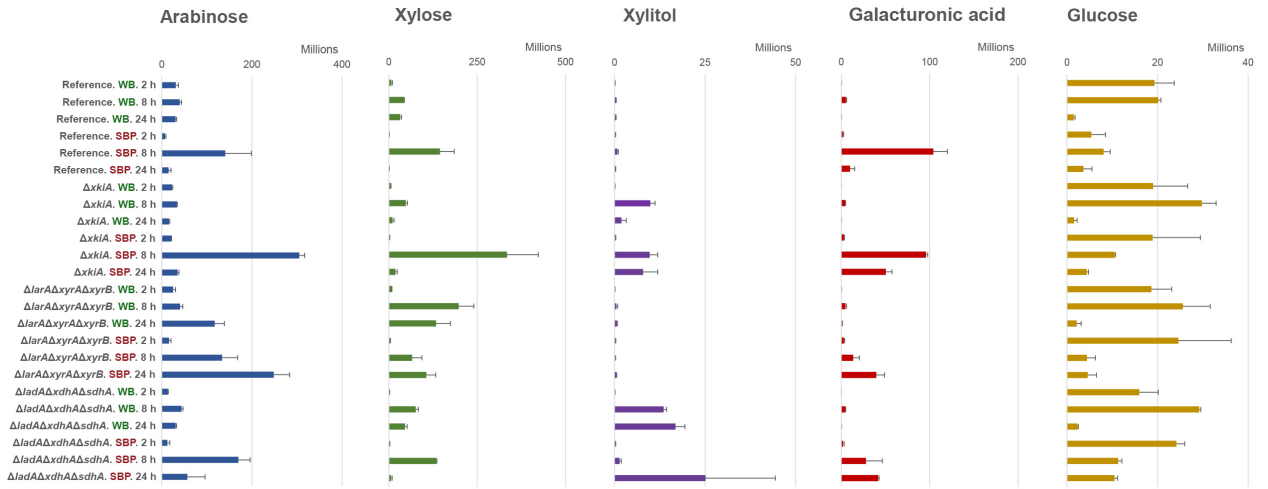


FIGURE 4 | Metabolite abundance in culture supernatant of arabinose, xylose, xylitol, galacturonic acid and glucose, after 2, 8, and 24 h transfer of the *A. niger* reference strain (N593 Δ*kusA*) and the PCP deletion mutants to 1% (w/v) wheat bran (WB) and 1% (w/v) sugar beet pulp (SBP). The error bars represent the standard deviation between three biological replicates.

intermediates. In this mutant, accumulation of D-galacturonic acid and glucose was also observed at the later time points on SBP, suggesting that neither of these sugars could be used as alternative carbon sources.

Significant Variation in Expression of the CAZy Genes Involved in Utilization of Arabinoxylan, Cellulose and Xyloglucan Was Observed Between the PCP Mutants

Expression of Carbohydrate Active Enzymes (CAZymes) was also impacted in the PCP deletion mutants (**Figure 5** and **Supplementary Table 2**) on both biomass substrates. Based on their substrate specificity, these enzymes were divided into 11 different sub-groups (**Figure 5** and **Supplementary Table 2**). The substrate specific enzyme sub-groups that showed the highest differences in the mutants compared to the reference strains were the ones involved in degradation of the polysaccharides arabinoxylan, cellulose and xyloglucan, justified by the composition of WB and SBP (**Figure 1C**). The arabinoxylan-specific sub-group comprised of several genes encoding β-1,4 endoxylanases (XLN), β-1,4 xylosidases (BXL), arabinoxylan arabinofuranohydrolase (AXH), α-glucoronidase (AGU) and acetyl xylan esterases (AXE). The cellulose−specific sub−group included genes encoding β-1,4-glucosidases (BGL), β−1,4−endoglucanases (EGL), cellobiohydrolases (CBH) and cellobiose dehydrogenases (CDH), while the xyloglucan−specific

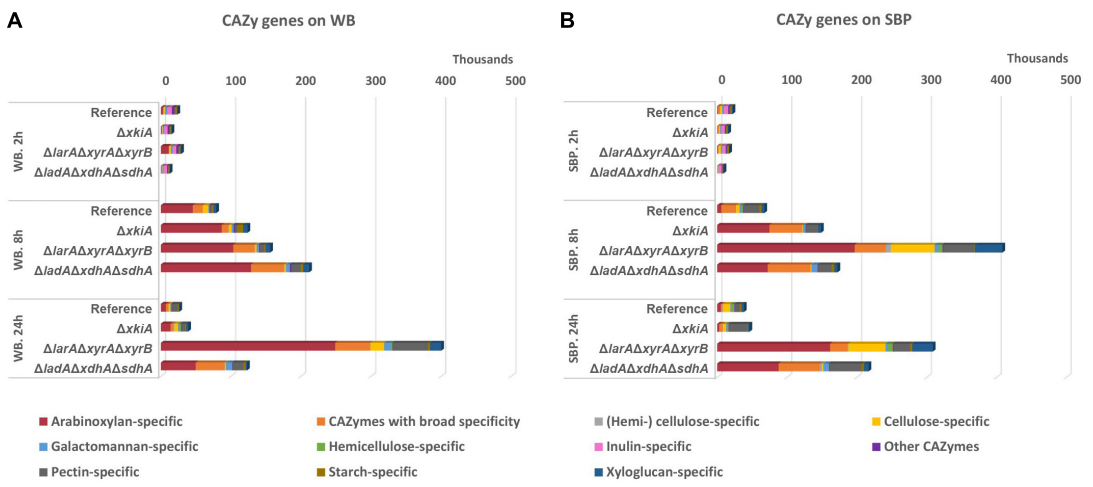


FIGURE 5 | Expression levels (FPKM) of genes encoding CAZymes during growth on **(A)** wheat bran (WB) and **(B)** sugar beet pulp (SBP) in the *A. niger* reference strain (N593 Δ*kusA*) and the PCP deletion mutants.

sub−group included xyloglucanases (XG-EGL), α−xylosidases (AXL) and α−fucosidases (AFC).

In *A. niger*, the transcriptional activator XlnR and AraR have been shown to regulate the expression of (hemi-)cellulolytic genes (van Peij et al., 1998a; de Vries, 2003), and arabinanolytic genes (Battaglia et al., 2011b), respectively. Genes under control of XlnR encode arabinoxylan−active enzymes, such as β-1,4 endoxylanases (*xlnB*, *xlnC*) (van Peij et al., 1998a,b), β-1,4 xylosidase (*xlnD*) (van Peij et al., 1998a,b), α−glucuronidase (*aguA*) (van Peij et al., 1998b; de Vries et al., 2002), acetyl xylan esterase (*axeA*) (van Peij et al., 1998b) and arabinoxylan arabinofuranohydrolase (*axhA*) (van Peij et al., 1998b), but also cellulose−active enzymes, such as β−endoglucanases (*eglA*, *eglB*, *eglC*) (van Peij et al., 1998b) and cellobiohydrolases (*cbhA*, *cbhB*) (Gielkens et al., 1999), and CAZymes with broad specificity such as feruloyl esterase (*faeA*) (van Peij et al., 1998b), α−galactosidase (*aglB*) (de Vries et al., 1999a) and β−galactosidase (*lacA*) (de Vries et al., 1999a). The transcriptional regulator AraR controls the expression of genes encoding α−arabinofuranosidases (*abfA*, *abfB*) in *A. niger* (Battaglia et al., 2011b).

The expression of the arabinoxylan-specific gene sub-group on both substrates remained low in the reference strain, where the accumulation of the XlnR and AraR inducers has been shown to remain at significantly lower levels compared to the PCP mutants (Chroumpi et al., 2021). In general, growth on WB, which is a particularly rich in arabinoxylan, led to higher expression of the genes encoding arabinoxylan-specific enzymes compared to the other groups of enzymes (**Figure 5A** and **Supplementary Table 2**). After 24 h of growth on WB, an even stronger upregulation of these genes was observed, coinciding with higher accumulation of pentoses in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant (**Figure 4**). On SBP, the expression of the arabinoxylan-specific genes was also strongly upregulated in all three Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA* mutants after 8 h compared to the reference strain (**Figure 5B** and **Supplementary Table 2**). However, after 24 h of incubation on both substrates, the expression of the arabinoxylan-specific genes in the Δ*xkiA* mutant was reduced at similar levels to the reference strain (**Figure 5** and **Supplementary Table 2**), as also earlier observed for the PCP genes (**Figures 3A,B**).

Although cellulose is a very abundant component of both WB and SBP, the expression of the cellulose-specific genes was not significantly upregulated on WB in the mutants compared to the reference strain (**Figure 5A** and **Supplementary Table 2**). On SBP, significantly higher expression of the cellulose-specific sub-group was only observed in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant (**Figure 5B** and **Supplementary Table 2**). In this mutant, increased expression of the xyloglucan-specific sub-group genes was also observed on SBP (**Figure 5B**). This upregulation of the cellulose-specific and xyloglucan−specific genes in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant could contribute in its ability to grow on SBP (**Figure 1B**).

Finally, the sub-group of CAZymes with broad specificity, consisting of enzymes that act on various polysaccharides such as α−arabinofuranosidase (ABF), feruloyl esterase (FAE), β−1,4−galactosidase (LAC), β−1,4−glucosidase (BGL) and lytic polysaccharide monooxygenase (LPMO), was also significantly affected. These activities are necessary for complete depolymerization of cellulose, arabinoxylan and xyloglucan which are present in WB and the expression of some of them has been shown to be under the control of XlnR. In the conditions where the expression of the arabinoxylan-specific enzymes was increased compared to the reference strain, the expression of CAZymes with broad specificity seems to follow the same pattern (**Figure 5** and **Supplementary Table 2**).

The Exo-Proteome Confirmed the Large Impact of the PCP Mutants on Lignocellulolytic Enzyme Production

The exo-proteome of the *A. niger* reference strain and the studied PCP mutants grown on WB and SBP was also analyzed to explain their phenotypic differences on these biomass materials among the studied strains. A complete list of CAZymes secreted by the reference strain, and the Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA* mutants is presented in **Supplementary Table 3**. All experimentally identified CAZymes were qualified (number of detected CAZymes) and quantified (protein abundance). In general, a delayed response of the extracellular proteome (**Figure 6**) compared to the transcriptome (**Figure 5**) was observed. On both substrates, the strongest representation of detected CAZymes in relation to the expressed CAZy genes, as well as the highest total protein abundance were measured in the culture supernatant after 24 h of incubation (**Supplementary Table 3**). Additionally, differences between the transcriptomic and proteomic responses were also observed, possibly related to factors such as temporal differences, membrane binding and/or stability of the produced proteins.

On WB, the total CAZyme abundance was higher in the reference strain compared to the mutants (**Figure 6**), which could partly explain the reduced phenotype of the mutants on this substrate compared to the reference strain (**Figure 1B**). Although no significant differences were observed in the arabinoxylan-specific sub-group among the different studied stains, high variability was detected in the secreted CAZymes involved in degradation of the polysaccharides cellulose, pectin and starch (**Figure 6A**). On WB, the cellulose-specific CAZymes were significantly reduced in all mutants compared to the reference strain. The least affected mutant regarding the abundance of this CAZy sub-group was Δ*xkiA*, although it was still significantly reduced compared to the reference strain. This could partly explain the fact that Δ*xkiA* was the only mutant that could still grow well on WB (**Figure 1B**). In the other two mutants, Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*ladA*Δ*xdhA*Δ*sdhA*, which were both nearly unable to utilize WB for growth, the levels of the cellulose-specific CAZymes in the secretome was dramatically reduced compared to the reference (**Figure 6A** and **Supplementary Table 3**), while the abundance of the pectin-specific CAZymes was significantly increased compared to Δ*xkiA* and the reference strain. However, the higher abundance of the pectin-specific subgroup in both mutants does not seem to be able to compensate for their reduced ability to grow on WB (**Figure 1B**). Finally, the secretion of the starch-specific CAZymes was significantly reduced in all three mutants compared to the

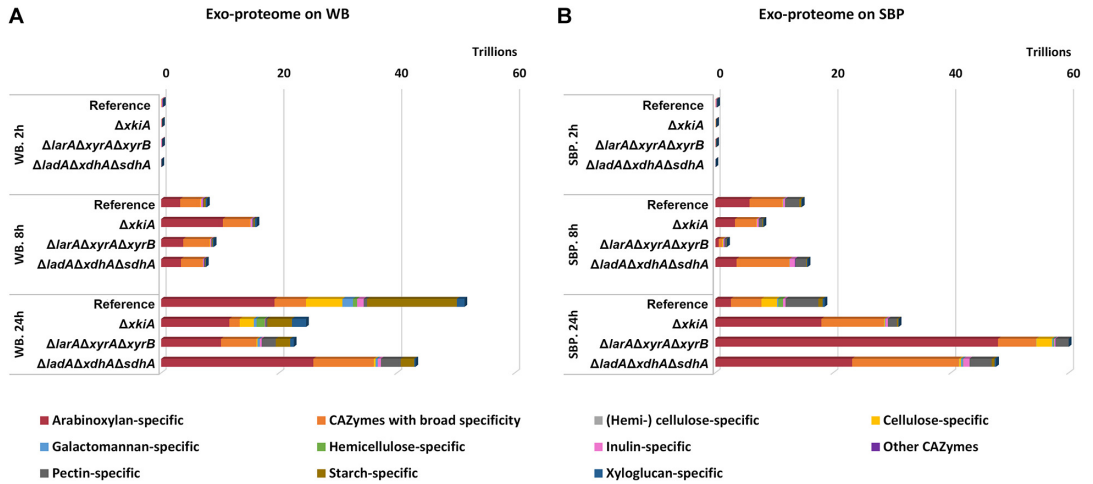


FIGURE 6 | CAZy protein abundance (protein intensity values) in the secretome of the *A. niger* reference strain (N593 Δ*kusA*) and the PCP deletion mutants grown on **(A)** wheat bran (WB) and **(B)** sugar beet pulp (SBP) for 2, 8, and 24 h.

reference strain. The most affected strains were the severely growth impaired Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*ladA*Δ*xdhA*Δ*sdhA*, highlighting the importance of this polysaccharide during growth of *A. niger* on WB.

On SBP, the abundance of the arabinoxylan-specific CAZymes significantly varied between the reference strain and the mutants (**Figure 6B**). The most pronounced difference was detected in Δ*larA*Δ*xyrA*Δ*xyrB*, showing the highest production of this CAZy subgroup compared to the reference strain and Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA*. Similar to the transcriptome response (**Figure 5B**), the cellulose-specific CAZy sub-group in Δ*larA*Δ*xyrA*Δ*xyrB* was secreted at similar levels to the reference strain (**Figure 6** and **Supplementary Table 3**), both showing the same growth pattern on this biomass substrate (**Figure 1B**). In Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA*, the abundance of the cellulose degrading CAZymes was significantly reduced compared to the reference strain. Finally, the xyloglucan-specific CAZy sub-group, which was also strongly upregulated in Δ*larA*Δ*xyrA*Δ*xyrB* grown on SBP (**Figure 5B**), was also secreted in significantly higher levels in Δ*larA*Δ*xyrA*Δ*xyrB*, compared to the other strains (**Figure 6B**).

DISCUSSION

In this study, the transcriptome, metabolome and proteome response of three pentose catabolic mutant strains of *A. niger*, Δ*larA*Δ*xyrA*Δ*xyrB*, Δ*ladA*Δ*xdhA*Δ*sdhA* and Δ*xkiA*, grown on two plant biomass substrates, WB and SBP, was evaluated. Both substrates contain considerable amounts of L-arabinose and D-xylose (**Figure 1C**), making them highly suitable for the analysis of this pathway. All three combinatorial mutations have been previously shown to block pentose catabolism at different steps of the pathway, resulting in abolishment of growth of all three mutants on L-arabinose and D-xylose, as well as in accumulation of different pathway intermediates (Chroumpi et al., 2021). However, our results reveal that despite the absolute impact of these PCP mutations during growth on pure pentose sugars, they are not as critical for growth of *A. niger* on more complex biomass substrates, such as WB and SBP.

In the Δ*xkiA* mutant, which was able to grow on both WB and SBP (**Figure 1B**), a strong reduction in expression of the PCP and the arabinoxylan-active CAZy genes was observed after 24 h on both substrates (**Figures 3A,B**, **5**). Both the PCP and the arabinoxylan-specific CAZy genes have been previously shown to be under the control of the transcriptional regulators AraR and XlnR, which are specifically induced in the presence of the pentose sugars and their polyols (van Peij et al., 1998a; Battaglia et al., 2011b). Our hypothesis is that their reduced expression is due to the reduction of the arabinose, xylose and xylitol concentrations observed on both WB and SBP after 24 h in this strain, in combination with the presence of high glucose levels (**Figure 4**). Similar suggestions have been also made in a previous study (de Vries et al., 1999b), where the expression of the arabinoxylan-specific genes *xlnB*, *xlnD*, *aguA* and *faeA* was drastically decreased at lower concentrations of D-xylose in the presence of D-glucose, due to activation of the carbon catabolite repressor CreA. Considering that Δ*xkiA* was unable to grown on L-arabinose and D-xylose, this suggest that during growth on plant biomass a possible by-pass mechanism is induced, resulting in efficient re-routing of the metabolism of Δ*xkiA* and rescue of its growth.

The depletion of the pentose sugars in the Δ*xkiA* mutant, especially on WB, could be explained by the presence of enzymes with some kinase activity on D-xylulose, of which the corresponding genes are induced under these conditions. Blocking the pathway at earlier steps, as in the Δ*larA*Δ*xyrA*Δ*xyrB* and Δ*ladA*Δ*xdhA*Δ*sdhA* mutants does not result in similar rescued growth on WB (**Figure 1B**), suggesting that it is D-xylulose for which an alternative enzyme or pathway is induced. However, no putative enzymes, classified in the same Pfam family or showing relatively close homology to

A. niger D-xylulokinase, were identified as likely candidates in our analysis. Alternatively, the conversion of D-xylulose into another compound that can be further metabolized through alternative carbon catabolic pathways, and not be phosphorylated, should be also considered. The better growth of the Δ*xkiA* mutant on SBP (**Figure 1B**), compared to WB, could be explained by the additional utilization of alternative carbon sources available for growth on this substrate (**Figure 1C**). SBP contains a significant amount of pectin, that next to L-arabinose also contains significant amounts of D-galacturonic acid and L-rhamnose. These sugars are converted through the GACP (Martens-Uzunova and Schaap, 2008; Alazi et al., 2017) and the RCP (Khosravi et al., 2017; Chroumpi et al., 2020; **Figure 7**). In Δ*xkiA*, high expression of the GACP and RCP genes were observed after 24 h (**Figures 3D,F**), while the accumulated D-galacturonic acid was significantly reduced after 24 h of growth on SBP (**Figure 4**). Higher expression of the pectin-specific CAZymes was also observed in this mutant on SBP, compared to the reference strain (**Figure 5**), supporting our conclusion that the presence of pectin can largely rescue growth on SBP.

The Δ*larA*Δ*xyrA*Δ*xyrB* mutant, which blocks the first step of pentose conversion (Chroumpi et al., 2021), was nearly unable to grow on WB, but only a small growth reduction was observed on SBP, compared to the reference strain (**Figure 1B**). As expected, accumulation of both pentoses were observed on both biomass substrates (**Figure 4**), while the expression of the PCP and arabinoxylan-active CAZy genes remained at high levels after 24 h (**Figures 3A,B**, **5**). The lower expression levels of the PCP genes on SBP, compared to WB, could be explained by the different composition of these biomass substrates in pentoses (**Figure 1C**). Additionally, the significantly higher expression of the arabinoxylan-active CAZy genes in this mutant (**Figure 5**), compared to the Δ*xkiA* and Δ*ladA*Δ*xdhA*Δ*sdhA* mutants, shows that the pentose sugars and no other PCP intermediates are the main inducers of this CAZy sub-group.

In the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, re-routing of sugar catabolism toward utilization of D-galacturonic acid as a substrate was also proposed, since the expression of the GACP genes (**Figure 3D**) was slightly increased after 24 h on SBP (**Figure 4**). However, low accumulation of D-galacturonic acid was observed after 24 h on SBP (**Figure 4**), showing reduced ability of the Δ*larA*Δ*xyrA*Δ*xyrB* mutant to metabolize D-galacturonic acid. Reduced growth and slower consumption rates of D-galacturonic acid were also previously reported for the *A. niger* Δ*larA* mutant grown on the pure sugar (Alazi et al., 2017). In their study, it was proposed that other partially redundant enzymes may also contribute in the conversion of 2-keto-3-deoxy- L-galactonate to pyruvate and L-glyceraldehyde. Deletion of the *lkaA* gene, involved in the last step of the L-rhamnose catabolism in *A. niger* (Chroumpi et al., 2020), in the Δ*gaaD* background, resulted in slightly reduced growth compared to the single Δ*gaaD* mutant (Chroumpi et al., 2021). However, growth of the double Δ*gaaD*Δ*lkaA* was not abolished on D-galacturonic acid, showing that also other enzymes are involved in this step.

The expression of the RCP genes was significantly upregulated in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, compared to the reference strain after 24 h of growth on SBP. The delayed response of the RCP genes on SBP, compared to the previously shown PCP and GACP genes, could be simply attributed to the sequential manner in which *A. niger* consumes sugars in liquid cultures (Mäkelä et al., 2018). L-rhamnose has been shown to be the least preferred carbon source between them, and as a result, the upregulation of the genes involved in its catabolism was significantly delayed compared to the other sugars (**Figure 3F**). Several other genes, encoding enzymes that are classified in the same Pfam family or show relatively close homology to *A. niger* pentose reductases and polyol dehydrogenases, were also found

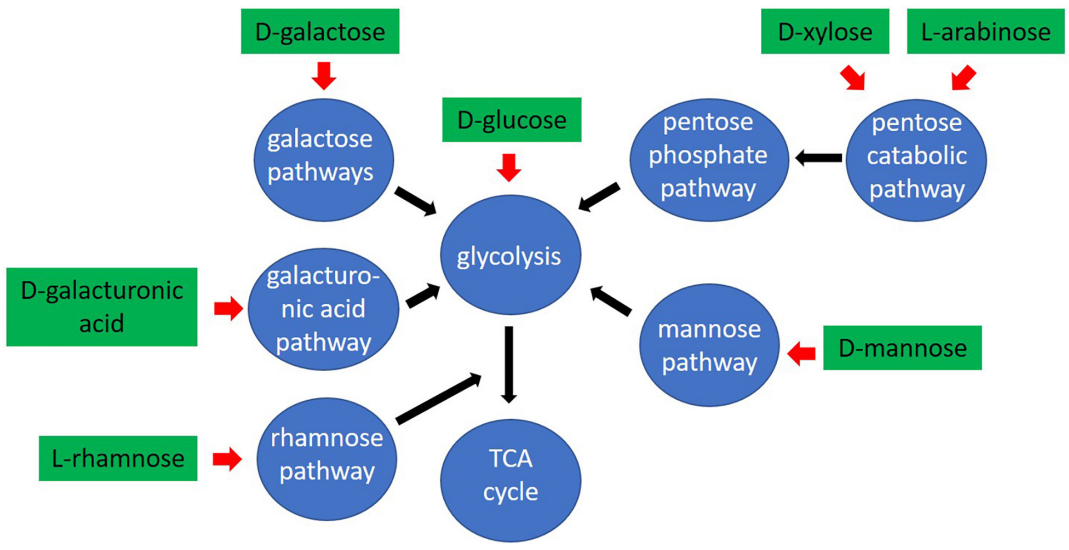


FIGURE 7 | Schematic overview of the pathways used by *Aspergillus niger* to convert plant biomass derived monosaccharides.

to be upregulated on SBP in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, compared to the reference strain (**Supplementary Table 4**). It is possible that apart from the utilization of other available sugars in SBP to support growth, these enzymes could be also involved in the rescue phenotype of this mutant on SBP, by re-routing the PCP metabolism of the Δ*larA*Δ*xyrA*Δ*xyrB* mutant. The expression of these genes was not particularly affected in the non-growing Δ*ladA*Δ*xdhA*Δ*sdhA* mutant on SBP, further supporting our theory.

In contrast to the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, the growth of the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant on SBP was severely impaired (**Figure 1B**). The Δ*ladA*Δ*xdhA*Δ*sdhA* mutant lacks *sdhA*, a gene that has been previously shown to be part of both the PCP (Chroumpi et al., 2021) and the oxido-reductive D-galactose pathway of *A. niger* (Koivistoinen et al., 2012). Thus, the growth impairment of Δ*ladA*Δ*xdhA*Δ*sdhA* on SBP could be attributed to the inability of this mutant in using both pentose sugars and D-galactose for growth. However, although this mutant was not able to grow on D-sorbitol, its growth was only slightly reduced on D-galactose compared to the reference strain (data not shown). This shows that on pure D-galactose, the D-galactose catabolism in the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant can follow an alternative route than the oxidoreductive D-galactose catabolic pathway. Although, the presence of all *A. niger* genes/enzymes involved in the Leloir and the oxido-reductive D-galactose catabolic pathways have been previously shown, the relative contribution of these two pathways during growth on D-galactose or D-galactose containing complex carbohydrates is not known. It could be that under the tested conditions the oxido-reductive pathway is mainly active and as a result catabolism of D-galactose in the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant is blocked, severely affecting its growth of on SBP. Finally, the potential involvement of the deleted dehydrogenase encoding genes also in other pathway steps of *A. niger* sugar catabolism or a possible severe intracellular redox imbalance effect, caused by the simultaneous deletion of these enzymes involved in NADH regeneration reactions (Witteveen et al., 1989; Koivistoinen et al., 2012), can also not be excluded. Because of the broad cellular and system functions of NAD^+-dependent enzymes, such an imbalance in the intracellular NAD^+/NADH ratio could alter cellular homeostasis by affecting enzymes that are involved in metabolism, regulation of gene expression, DNA repair, intracellular trafficking, aging, and cell death.

Interestingly, accumulation of D-galacturonic acid and glucose was also observed at the later time points on SBP in this mutant, suggesting that neither of these sugars could be used as alternative carbon sources. This was surprising as the Δ*ladA*Δ*xdhA*Δ*sdhA* mutant was not hypothesized to affect the D-galacturonic acid catabolism.

It was previously shown that blocking the direct entry of hexoses to the glycolytic pathway by deletion of the *hxkA* and *glkA* genes activates alternative metabolic conversion of these sugars in *Aspergillus nidulans* during growth on WB, but also upregulates conversion of other sugars, such as pentoses (Khosravi et al., 2018). However, in our study, although both substrates are rich in cellulose, this does not seem to significantly compensate for their reduced ability to use pentoses. In all the PCP mutants, the expression of the glycolytic genes on both substrates was unaffected, compared to the reference strain (data not shown). Additionally, neither the genes encoding cellulose-active enzymes were significantly upregulated in most of the PCP mutants (**Figure 5**) nor extracellular detection of these enzymes was observed in the proteomic data (**Figure 6**). Only in the Δ*larA*Δ*xyrA*Δ*xyrB* mutant, the cellulase-specific CAZy genes were upregulated on SBP (**Figure 5B**) and earlier reduction of the accumulated glucose levels was observed (**Figure 4**), which may have also contributed to its rescued growth on this substrate. Improved cellulase production by *A. niger* has been previously achieved by deleting the *noxR* gene, encoding the regulatory subunit of the NADPH oxidase complex (Patyshakuliyeva et al., 2016). A combination of the PCP and *noxR* gene deletions might help to improve cellulase utilization of these strains and as a result the growth of the PCP mutants on these biomass substrates.

Our results also suggests that an adaptation to other components of the substrates may cause a general growth arrest. This was in particular true for SBP as at 8 h an overall repression of genes involved in most of the studied GO terms was observed for all three mutants. This could be an indication that at that later time point the mutants' metabolism is paused, due to the fact that the most available carbon sources L-arabinose and D-xylose cannot be used for growth, before they redirect their metabolism and adapt in the new situation.

To conclude, our results demonstrate that despite the significant impact of these catabolic gene deletions during growth on pure pentose sugars, they are not as critical for growth of *A. niger* on more complex biomass substrates. It seems that the high sugar heterogeneity of these substrates in combination with the high complexity and adaptability of the fungal sugar metabolism allow for activation of alternative strategies to support growth. Production of additional enzymes that have side-activity on the PCP sugars, and therefore contribute to the conversion of D-xylose or L-arabinose, or re-routing of sugar catabolism toward utilization of other available plant derived monosaccharides apart from pentoses were shown to be involved. This advanced understanding of central metabolic pathways is critical when applying metabolic engineering strategies in biotechnology. The use of low-cost lignocellulosic biomass materials as feedstocks combined with metabolic engineering could facilitate efficient utilization of the raw materials, but high production rates and high growth rates are required to attain economically feasible biotechnological processes. The effects of the mutations on CAZy genes were less clear, which could be due to the fact that many of these genes are affected by multiple regulators in *A. niger* (Gruben et al., 2017).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://massive.ucsd.edu/, MSV000086614; https://www.ncbi.nlm.nih.gov/genbank/, GSE162901.

AUTHOR CONTRIBUTIONS

TC performed the experiments, analyzed the data, and wrote the manuscript. MP contributed to the bioinformatics analysis and data visualization. LM, HM, CN, CH, VP, NT, CC, GO, and SB performed the RNA sequencing and metabolomic and proteomic analyses. RV and TC designed the study. RV and MM critically revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

TC was supported by a grant of the NWO ALWOP.233 to RV. A portion of the research was performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program. The Academy of Finland grant no. 308284 to MM is also acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.644216/full#supplementary-material

Supplementary Table 1 | Expression levels (FPKM) of catabolic genes during growth on wheat bran (WB) and sugar beet pulp (SBP) in the *A. niger* reference strain (N593 ΔkusA) and the PCP deletion mutants.

Supplementary Table 2 | Expression levels (FPKM) of genes encoding CAZymes during growth on wheat bran (WB) and sugar beet pulp (SBP) in the *A. niger* reference strain (N593 ΔkusA) and the PCP deletion mutants.

Supplementary Table 3 | Comparison between the transcriptome and exo-proteome analysis of the *A. niger* reference strain (N593 ΔkusA) and the PCP deletion mutants grown on wheat bran (WB) and sugar beet pulp (SBP) for 2, 8, and 24 h. Transcripts were considered differentially expressed if the DESeq2 fold change was >2 (red font) or <0.5 (green font) and Padj <0.01. Similarly, for an identified protein to be considered differentially produced, the requirement was a fold change of the mean intensity values of >2 (highlighted in red) or <0.5 (highlighted in green) and Padj <0.05. Note that where an intensity value was not detected for a protein in a sample, a zero value was used. "Inf" values shown in the fold change columns are caused by zero values in the mutant. All values are averages of triplicates. The fold change is the difference between the reference strain and the deletion mutants.

Supplementary Table 4 | Expression levels (FPKM) of genes encoding enzymes that are classified in the same Pfam family or show relatively close homology to *A. niger* pentose reductases and polyol dehydrogenases, during growth on (A) wheat bran (WB) and (B) sugar beet pulp (SBP) in the *A. niger* reference strain (N593 ΔkusA) and the PCP deletion mutants. Transcripts were considered differentially expressed if the DESeq2 fold change was > 2 (highlighted in orange) or <0.5 (highlighted in green) and Padj <0.01. Transcripts with FPKM ≤50 were considered lowly (i.e., not substantially) expressed and highlighted in red font. The values are averages of triplicates. The fold change is the difference between the reference strain and the deletion mutants.

REFERENCES

Aguilar-Pontes, M. V., Brandl, J., McDonnell, E., Strasser, K., Nguyen, T. T. M., Riley, R., et al. (2018). The gold-standard genome of *Aspergillus niger* NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. *Stud. Mycol.* 91, 61–78. doi: 10.1016/j.simyco.2018.10.001

Alazi, E., Khosravi, C., Homan, T. G., du Pré, S., Arentshorst, M., Di Falco, M., et al. (2017). The pathway intermediate 2-keto-3-deoxy-L-galactonate mediates the induction of genes involved in D-galacturonic acid utilization in *Aspergillus niger*. *FEBS Lett.* 591, 1408–1418. doi: 10.1002/1873-3468.12654

Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). NCBI geo: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193

Battaglia, E., Hansen, S. F., Leendertse, A., Madrid, S., Mulder, H., Nikolaev, I., et al. (2011a). Regulation of pentose utilisation by AraR, but not XlnR, differs in *Aspergillus nidulans* and *Aspergillus niger*. *Appl. Microbiol. Biotechnol.* 91, 387–397. doi: 10.1007/s00253-011-3242-2

Battaglia, E., Visser, L., Nijssen, A., van Veluw, G. J., Wösten, H. A. B., and de Vries, R. P. (2011b). Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in *Eurotiales*. *Stud. Mycol.* 69, 31–38. doi: 10.3114/sim.2011.69.03

Benoit, I., Culleton, H., Zhou, M., DiFalco, M., Aguilar-Osorio, G., Battaglia, E., et al. (2015). Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. *Biotechnol. Biofuels* 8:107. doi: 10.1186/s13068-015-0285-0

Chroumpi, T., Aguilar-Pontes, M. V., Peng, M., Wang, M., Lipzen, A., Ng, V., et al. (2020). Identification of a gene encoding the last step of the L-rhamnose catabolic pathway in *Aspergillus niger* revealed the inducer of the pathway regulator. *Microbiol. Res.* 234:126426. doi: 10.1016/j.micres.2020.126426

Chroumpi, T., Peng, M., Aguilar-Pontes, M. V., Müller, A., Wang, M., Yan, J., et al. (2021). Revisiting a 'simple' fungal metabolic pathway reveals redundancy, complexity and diversity. *Micobial biotechnol.* (in press).

de Groot, M. J. L., van den Dool, C., Wösten, H. A. B., Levisson, M., Vankuyk, P. A., Ruijter, G. J. G., et al. (2007). Regulation of pentose catabolic pathway genes of *Aspergillus niger*. *Food Technol. Biotechnol.* 45, 134–138.

de Vries, R. P. (2003). Regulation of *Aspergillus genes* encoding plant cell wall polysaccharide-degrading enzymes; relevance for industrial production. *Appl. Microbiol. Biotechnol.* 61, 10–20. doi: 10.1007/s00253-002-1171-9

de Vries, R. P., Burgers, K., van de Vondervoort, P. J., Frisvad, J. C., Samson, R. A., and Visser, J. (2004). A new black *Aspergillus species*, a. vadensis, is a promising host for homologous and heterologous protein production. *Appl. Environ. Microbiol.* 70, 3954–3959. doi: 10.1128/AEM.70.7.3954-3959.2004

de Vries, R. P., van de Vondervoort, P., Hendriks, L., van de Belt, M., and Visser, J. (2002). Regulation of the α-glucuronidase-encoding gene (aguA) from *Aspergillus niger*. *Mol. Genet. Genom.* 268, 96–102. doi: 10.1007/s00438-002-0729-7

de Vries, R. P., and Visser, J. (2001). Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol. Mol. Biol. Rev.* 65, 497–522. doi: 10.1128/MMBR.65.4.497-522.2001

de Vries, R. P., van den Broeck, H. C., Dekkers, E., Manzanares, P., de Graaff, L. H., and Visser, J. (1999a). Differential expression of three α-galactosidase genes and a single β-galactosidase gene from *Aspergillus niger*. *Appl. Environ. Microbiol.* 65, 2453–2460. doi: 10.1128/AEM.65.6.2453-2460.1999

de Vries, R. P., Visser, J., and de Graaff, L. H. (1999b). CreA modulates the XlnR-induced expression on xylose of *Aspergillus niger* genes involved in xylan degradation. *Res. Microbiol.* 150, 281–285. doi: 10.1016/S0923-2508(99)80053-9

Fiehn, O. (2016). Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling. *Curr. Protoc. Mol. Biol.* 114, 30.4.1–30.4.32. doi: 10.1002/0471142727.mb3004s114

Gielkens, M. M., Dekkers, E., Visser, J., and de Graaff, L. H. (1999). Two cellobiohydrolase-encoding genes from *Aspergillus niger* require D-xylose and the xylanolytic transcriptional activator XlnR for their expression. *Appl. Environ. Microbiol.* 65, 4340–4345. doi: 10.1128/AEM.65.10.4340-4345.1999

Gruben, B. S., Mäkelä, M. R., Kowalczyk, J. E., Zhou, M., Benoit-Gelber, I., and de Vries, R. P. (2017). Expression-based clustering of CAZyme-encoding genes of *Aspergillus niger*. *BMC Genomics* 18:900. doi: 10.1186/s12864-017-4164-x

Hiller, K., Hangebrauk, J., Jäger, C., Spura, J., Schreiber, K., and Schomburg, D. (2009). MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal. Chem.* 81, 3429–3439. doi: 10.1021/ac802689c

Kelly, R. T., Page, J. S., Luo, Q., Moore, R. J., Orton, D. J., Tang, K., et al. (2006). Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem.* 78, 7796–7801. doi: 10.1021/ac061133r

Khosravi, C., Battaglia, E., Kun, R. S., Dalhuijsen, S., Visser, J., Aguilar-Pontes, M. V., et al. (2018). Blocking hexose entry into glycolysis activates alternative metabolic conversion of these sugars and upregulates pentose metabolism in *Aspergillus nidulans*. *BMC Genomics* 19:214. doi: 10.1186/s12864-018-4609-x

Khosravi, C., Benocci, T., Battaglia, E., Benoit, I., and de Vries, R. P. (2015). Sugar catabolism in *Aspergillus* and other fungi related to the utilization of plant biomass. *Adv. Appl. Microbiol.* 90, 1–28. doi: 10.1016/bs.aambs.2014.09.005

Khosravi, C., Kun, R. S., Visser, J., Aguilar-Pontes, M. V., de Vries, R. P., and Battaglia, E. (2017). In vivo functional analysis of L-rhamnose metabolic pathway in *Aspergillus niger*: a tool to identify the potential inducer of RhaR. *BMC Microbiol.* 17:214. doi: 10.1186/s12866-017-1118-z

Kim, S., Gupta, N., and Pevzner, P. A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* 7, 3354–3363. doi: 10.1021/pr8001244

Kim, S., and Pevzner, P. A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5:5277. doi: 10.1038/ncomms6277

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., et al. (2009). FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* 81, 10038–10048. doi: 10.1021/ac9019522

Koivistoinen, O. M., Richard, P., Penttilä, M., Ruohonen, L., and Mojzita, D. (2012). Sorbitol dehydrogenase of *Aspergillus niger*, SdhA, is part of the oxido-reductive D-galactose pathway and essential for D-sorbitol catabolism. *FEBS Lett.* 586, 378–383. doi: 10.1016/j.febslet.2012.01.004

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Mäkelä, M. R., Aguilar-Pontes, M. V., van Rossen-Uffink, D., Peng, M., and de Vries, R. P. (2018). The fungus *Aspergillus niger* consumes sugars in a sequential manner that is not mediated by the carbon catabolite repressor CreA. *Sci. Rep.* 8:6655. doi: 10.1038/s41598-018-25152-x

Martens-Uzunova, E. S., and Schaap, P. J. (2008). An evolutionary conserved D-galacturonic acid metabolic pathway operates across filamentous fungi capable of pectin degradation. *Fungal Genet. Biol.* 45, 1449–1457. doi: 10.1016/j.fgb.2008.08.002

Meyer, V., Arentshorst, M., El-Ghezal, A., Drews, A. C., Kooistra, R., van den Hondel, C. A., et al. (2007). Highly efficient gene targeting in the *Aspergillus niger* kusA mutant. *J. Biotechnol.* 128, 770–775. doi: 10.1016/j.jbiotec.2006.12.021

Mohnen, D. (2008). Pectin structure and biosynthesis. *Curr. Opin. Plant Biol.* 11, 266–277. doi: 10.1016/j.pbi.2008.03.006

Mojzita, D., Penttilä, M., and Richard, P. (2010). Identification of an L-arabinose reductase gene in *Aspergillus niger* and its role in L-arabinose catabolism. *J. Biol. Chem.* 285, 23622–23628. doi: 10.1074/jbc.M110.113399

Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N., and Smith, R. D. (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem* 32, 215–217. doi: 10.1016/j.compbiolchem.2008.02.006

Nakayasu, E. S., Nicora, C. D., Sims, A. C., Burnum-Johnson, K. E., Kim, Y.-M., Kyle, J. E., et al. (2016). MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *mSystems* 1:e00043-16. doi: 10.1128/mSystems.00043-16

Ochoa-Villarreal, M., Aispuro, E., Vargas-Arispuro, I., and Martínez-Téllez, M. (2012). ""Plant cell wall polymers: function, structure and biological activity of their derivatives"," in *Polymerization*, ed. A. De Souza Gomes (London: InTechOpen).

Patyshakuliyeva, A., Arentshorst, M., Allijn, I. E., Ram, A. F., de Vries, R. P., and Gelber, I. B. (2016). Improving cellulase production by *Aspergillus niger* using adaptive evolution. *Biotechnol. Lett.* 38, 969–974. doi: 10.1007/s10529-016-2060-0

Polpitiya, A. D., Qian, W.-J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G., et al. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556–1558. doi: 10.1093/bioinformatics/btn217

Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289. doi: 10.1146/annurev-arplant-042809-112315

Seiboth, B., and Metz, B. (2011). Fungal arabinan and L-arabinose metabolism. *Appl. Microbiol. Biotechnol.* 89, 1665–1673. doi: 10.1007/s00253-010-3071-8

Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78. doi: 10.1146/annurev.cellbio.22.022206.160206

van Peij, N. N., Visser, J., and de Graaff, L. H. (1998a). Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. *Mol. Microbiol.* 27, 131–142. doi: 10.1046/j.1365-2958.1998.00666.x

van Peij, N. N. M. E., Gielkens, M. M. C., de Vries, R. P., Visser, J., and de Graaff, L. H. (1998b). The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in *Aspergillus niger*. *Appl. Environ. Microbiol.* 64, 3615–3619. doi: 10.1128/aem.64.10.3615-3619.1998

Witteveen, C. F. B., Busink, R., van de Vondervoort, P., Dijkema, C., Swart, K., and Visser, J. (1989). L-arabinose and D-xylose catabolism in *Aspergillus niger*. *Microbiology* 135, 2163–2171. doi: 10.1099/00221287-135-8-2163

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 07 April 2021
doi: 10.3389/fbioe.2021.603832

Integration of Proteomics and Metabolomics Into the Design, Build, Test, Learn Cycle to Improve 3-Hydroxypropionic Acid Production in *Aspergillus pseudoterreus*

Kyle R. Pomraning[1], Ziyu Dai[1], Nathalie Munoz[1], Young-Mo Kim[1], Yuqian Gao[1], Shuang Deng[1], Joonhoon Kim[1,2], Beth A. Hofstad[1], Marie S. Swita[1], Teresa Lemmon[1], James R. Collett[1], Ellen A. Panisko[1], Bobbie-Jo M. Webb-Robertson[1], Jeremy D. Zucker[1], Carrie D. Nicora[1], Henrique De Paoli[3], Scott E. Baker[1], Kristin E. Burnum-Johnson[1], Nathan J. Hillson[3] and Jon K. Magnuson[1,2]* for the Agile BioFoundry*

[1] Pacific Northwest National Laboratory, Richland, WA, United States, [2] Joint BioEnergy Institute, Emeryville, CA, United States, [3] Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Biological engineering of microorganisms to produce value-added chemicals is a promising route to sustainable manufacturing. However, overproduction of metabolic intermediates at high titer, rate, and yield from inexpensive substrates is challenging in non-model systems where limited information is available regarding metabolic flux and its control in production conditions. Integrated multi-omic analyses of engineered strains offers an in-depth look at metabolites and proteins directly involved in growth and production of target and non-target bioproducts. Here we applied multi-omic analyses to overproduction of the polymer precursor 3-hydroxypropionic acid (3HP) in the filamentous fungus *Aspergillus pseudoterreus*. A synthetic pathway consisting of aspartate decarboxylase, beta-alanine pyruvate transaminase, and 3HP dehydrogenase was designed and built for *A. pseudoterreus*. Strains with single- and multi-copy integration events were isolated and multi-omics analysis consisting of intracellular and extracellular metabolomics and targeted and global proteomics was used to interrogate the strains in shake-flask and bioreactor conditions. Production of a variety of co-products (organic acids and glycerol) and oxidative degradation of 3HP were identified as metabolic pathways competing with 3HP production. Intracellular accumulation of nitrogen as 2,4-diaminobutanoate was identified as an off-target nitrogen sink that may also limit flux through the engineered 3HP pathway. Elimination of the high-expression oxidative 3HP degradation pathway by deletion of a putative malonate semialdehyde dehydrogenase improved the yield of 3HP by 3.4 × after 10 days in shake-flask culture. This is the first report of 3HP production in a filamentous fungus amenable to industrial scale biomanufacturing of organic acids at high titer and low pH.

Keywords: 3-hydroxypropionic acid (3-HP), *Aspergillus pseudoterreus*, beta-alanine pathway, Agile BioFoundry, 3HP

OPEN ACCESS

Edited by:
Jose Ruben Morones-Ramirez, Autonomous University of Nuevo León, Mexico

Reviewed by:
Jingyu Wang, Westlake Institute for Advanced Study (WIAS), China
Francisco Balderas-Cisneros, Autonomous University of Nuevo León, Mexico

****Correspondence:***
Kyle R. Pomraning
Kyle.Pomraning@pnnl.gov
Jon K. Magnuson
Jon.Magnuson@pnnl.gov

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *08 September 2020*
Accepted: *16 March 2021*
Published: *07 April 2021*

Citation:
Pomraning KR, Dai Z, Munoz N, Kim Y-M, Gao Y, Deng S, Kim J, Hofstad BA, Swita MS, Lemmon T, Collett JR, Panisko EA, Webb-Robertson B-JM, Zucker JD, Nicora CD, De Paoli H, Baker SE, Burnum-Johnson KE, Hillson NJ and Magnuson JK (2021) Integration of Proteomics and Metabolomics Into the Design, Build, Test, Learn Cycle to Improve 3-Hydroxypropionic Acid Production in Aspergillus pseudoterreus. Front. Bioeng. Biotechnol. 9:603832. doi: 10.3389/fbioe.2021.603832

INTRODUCTION

Since the industrial revolution, petroleum-based feedstocks have been the primary source for production of fuels and chemicals. However, their non-renewable nature and the detrimental effects of extractive practices have fueled a movement to produce drop-in or alternative fuels and chemicals from renewable feedstocks. A bioeconomy, based around the production of fuels and chemicals from renewable biological feedstocks has been proposed and supported as a way to improve the sustainability of fuel and commodity chemical production. Building new and modifying existing industries to produce the myriad of chemicals currently derived from petrochemical feedstocks will require extensive metabolic engineering of a wide variety of bacteria, plants, and fungi and integration with catalytic and thermochemical conversion processes to attain near-term economic viability. To support this, a public Agile BioFoundry has been established to efficiently engineer microorganisms for the production of fuels and chemicals from renewable feedstocks[1] (Hillson et al., 2019). The breadth of capabilities available to Design, Build, Test, and Learn (DBTL) from engineered organisms at a dedicated biofoundry enables a system-wide examination of engineered pathways and a deeper understanding of metabolic capability in non-model bioconversion hosts. An emphasis on maximizing the efficiency of multi-omic analyses during the Test and Learn portions of the DBTL cycle will expedite the arrival of a functioning bioeconomy by allowing strain and bioprocess development to overcome challenges associated with the complexity of metabolic systems.

Of interest to supporting a sustainable bioeconomy are biological production of chemicals that can be polymerized to produce biobased and biodegradable plastics. One such chemical, 3-hydroxypropionic acid (3HP) can be used to produce polymers directedly, such as poly-3HP, or more complex structures that also incorporate 3-hydroxybutyric acid, 4-hydroxybutyric acid, and/or lactic acid to produce polymers with altered physical properties (Tripathi et al., 2013; Meng et al., 2015; Ren et al., 2017). For drop-in to existing industrial processes, 3HP can be dehydrated to produce acrylic acid (Decoster et al., 2013) which has received commercial interest for biobased production of acrylates from glycerol (BASF, 2014). 3HP is also a precursor for carbon fiber via conversion to acrylonitrile and polymerization to polyacrylonitrile, which can be used in standard industrial processes to make carbon fibers (Davey, 2018).

3HP is produced naturally by phototrophic bacteria in the family *Chloroflexaceae* capable of CO_2 fixation via the 3HP cycle (Strauss and Fuchs, 1993; Klatt et al., 2007) as well as chemotrophic *Crenarchaeota* via the 3HP/4HB cycle (Berg et al., 2007). Enzymes from naturally occurring 3HP biosynthetic pathways have been used to design 3HP production pathways that route metabolic flux through beta-alanine, malonyl-CoA, or glycerol (Valdehuesa et al., 2013; Ji et al., 2018). Iterations of these pathways have been engineered into a variety of yeast and bacteria with notable successes. In the yeast *Saccharomyces cerevisiae* both the malonyl-CoA (Kildegaard et al., 2016) and beta-alanine (Borodina et al., 2015) pathways have been established with yields of 0.13 and 0.14 g/g 3HP from glucose. In bacteria production of 3HP via the glycerol pathway has been particularly successful with yields as high as 0.457 g/g 3HP directly from glycerol with engineered *Escherichia coli* (Kim et al., 2020) and 0.51 g/g 3HP from glucose with engineered *Corynebacterium glutamicum* which is also capable of converting xylose to 3HP (Chen et al., 2017). These examples demonstrate that microorganisms can be engineered to produce 3HP at reasonably high levels in laboratory conditions and further research will continue to improve production efficiency toward levels needed for economic viability. However, apart from *C. glutamicum*, the organisms being engineered for 3HP production may not be ideal hosts for bioconversion of lignocellulosic and low-cost renewable feedstocks that are often composed of a spectrum of organic compounds that model organisms may struggle to metabolize. Further, production of 3HP as a free acid, rather than its conjugate base may aid online or down-stream separation processes. Filamentous fungi from the genus *Aspergillus* are notable for their ability to produce organic acids such as citric acid, malic acid, and itaconic acid at industrial scale and at a pH that is well below the pKa = 4.5 of 3HP (Liu et al., 2017; Cairns et al., 2018; Kuenz and Krull, 2018). To that end, we examined utilization of *Aspergillus pseudoterreus* as a host for bioconversion of sugars to 3HP through the beta-alanine pathway (Borodina et al., 2015) because of its acid tolerance, high glycolytic flux, and ability to utilize a broad range of carbon and nitrogen sources (Kuenz and Krull, 2018; Saha et al., 2019; Deng et al., 2020). Engineered strains were evaluated using multi-omics analyses to understand the impacts of the introduced metabolic pathway and identify targets for improvement of 3HP production.

MATERIALS AND METHODS

Strain Maintenance and Cultivation

Aspergillus pseudoterreus strain ATCC® 32359™ was obtained from American Type Culture Collection (Manassas, VA, United States). All strains were maintained on complete medium agar (CM) (Deng et al., 2020) and spore suspensions stored in 15% glycerol at −80°C. Spore inoculum was grown on CM agar plates for 5 days to produce conidia, which were harvested by washing the plates with sterile 0.4% Tween 80 solution. For shake-flask experiments, 2E6 *A. pseudoterreus* spores were inoculated in 50 mL minimal medium [MM; 100 g/L glucose, 2.36 g/L $(NH_4)_2SO_4$, 0.11 g/L KH_2PO_4, 2.08 g/L $MgSO_4.7H_2O$, 0.13 g/L $CaCl_2.2H_2O$, 74 mg/L NaCl, 1.3 mg/L $ZnSO_4.7H_2O$, 0.7 mg/L $MnCL_2.4H_2O$, 5.5 mg/L $FeSO_4.7H_2O$, 0.2 mg/L $CuSO_4.5H_2O$, adjusted to pH 3.4 with 5M KOH] adapted from Riscaldati et al. (2000) in 250 mL shake-flasks at 30°C and 200 rpm in an orbital shaker. For submerged stirred tank 20 L bioreactor cultivation, 2E6 *A. pseudoterreus* spores were inoculated within a 30-L Sartorius Biostat-C bioreactor filled to 20 L with MM at 30°C. Impeller stir-rate was set at a constant rate of 100 rpm until the pH dropped to 2.8 after which stir-rate was increased to 400 rpm and pH maintained at 2.8 by online addition of 5M KOH. For submerged stirred tank 0.5 L bioreactor cultivation,

[1]https://agilebiofoundry.org/

2E6 *A. pseudoterreus* spores were inoculated within 0.5 L Sixfors bioreactors (Infors HT, Basel, Switzerland) filled to 0.45 L with MM at 30°C. Impeller stir-rate was set at a constant rate of 200 rpm until the pH dropped to 2.8 after which stir-rate was increased to 600 rpm and pH maintained at 2.8 by online addition of 5M KOH. For both reactor scales air was sparged at a constant rate of 0.4 vessel volumes/min. Sterile broth samples were collected by filtration through a 0.2 μm filter and then frozen at −20°C for analysis of sugar, nutrient, and metabolite concentrations. Cell pellets were collected by vacuum filtration on 0.45 μm nylon Whatman filters and washed with phosphate buffered saline. Cells were transferred to Eppendorf tubes, flash frozen in liquid nitrogen, and stored at −80°C. Mycelial dry cell weight at each time point was determined by harvesting the mycelia on a pre-weighed filter by vacuum filtration and washing with 20 mL distilled water. Subsequently, the dry weight was determined after freeze-drying in a lyophilizer overnight.

Strain Construction

Coding sequences for *Tc*PAND, *Bc*BAPAT, and *Ec*HPDH were codon optimized for *Aspergillus* species and synthesized. Individual DNA fragments were isolated with the following primer pairs; 5′-*cad1* of *A. pseudoterreus* 5′-(ccctcgaggtcgacggtatcgatagatatcggttgtagcagcgtaaacac-3′/5′-tctttcatagtagccttggtgaacatcttgagg-3′); *gpdA* promoter of *A. niger* (5′-atgttcaccaaggctactatgaaagaccgcgatg-3′/5′-cgccggtggcgggcattgtttagatgtgtctatgtg-3′); *Tc*PAND coding sequence (5′-catctaaacaatgcccgccaccggcgaggacca-3′/5′-atccaacccatcagaggtcggagcccaggcgttcg-3′); bi-direction transcriptional terminator of *A. niger elf3* gene (5′-gggctccgacctctgatgggttggatgacgatg-3′/5′-tctggcccagctctgagtcctagatgggtggtg-3′); *Bc*BAPAT coding sequence (5′-catctaggactcagagctgggccagacattccttc-3′/5′-gtccatcaacatggaactgatgatcgtccaggtcac-3′); and *A. niger eno1* promoter (5′-cgatcatcagttccatgttgatggactggaggg-3′/5′-gaactagtggatcccccgggctgcgttaactcgagcttacaagaagtagcc-3′) by high-fidelity PCR and assembled by Gibson Assembly (Gibson et al., 2008) into the pBlueScript SK(−) vector linearized with restriction enzymes *Hin*dIII and *Pst*I to form pZD4025 with an addition of *Hpa*I restriction enzyme site downstream of the *enol* promoter.

Individual DNA fragments were then isolated with the following primer pairs; *gpdA* (or *tdh*) promoter of *A. nidulans* (5′-acaggctacttcttgtaagctcgagtttctgtacagtgaccggtgac-3′/5′-tgaccagcacgatcatggtgatgtctgctcaag-3′); EcHPDH coding sequence (5′-agacatcaccatgatcgtgctggtcacgggcgc-3′/5′-gccatcggtcctattggcggtggacgttcaggc); *trpC* transcriptional terminator (5′-cgtccaccgccaataggaccgatggctgtgtag-3′/5′-cccgtctgtcagagagcggattcctcagtctcg-3′); pyrithiamine resistance gene (ptrA) of *A. oryzae* (5′-gaggaatccgctctctgacagacgggcaattgattac-3′/5′-gaatgttgctgaggagccgctcttgcatctttg-3′); 3′-*cad1* gene (5′-gcaagagcggctcctcagcaacattcgccatgttc-3′/5′-actaaagggaacaaaagctggagctcagctccactgctcatagtctttg-3′) by PCR and assembled along with *Hpa*I linearized pZD4025 to form pZD4028. pZD4028 was linearized with *Eco*RV for chemical mediated protoplast transformation into *A. pseudoterreus* (ATCC® 32359™) as previously described (Deng et al., 2020) to construct strains ABF_004528_2 (*cad*:3HP+) and ABF_004528_6 (*cad*:3HP+, 3HP+).

Upstream and downstream regions of the candidate malonate semialdehyde dehydrogenase (Ap*ald6*) were PCR amplified from *A. pseudoterreus* genomic DNA using primer pairs (5′-gcctataagcaaccgcctgtataaggaagagccttggtgctaacggc-3′/5′-gctgcgcaactgttgggaagggcgatgctgccagactgcaacaagaacc-3′) and (5′-gcagcccagtagtaggttgaggccgtcgacgaagctgatggccttgatgg-3′/5′-tgttgactccactagctccagccaaggcttctcgcactctcaccaccgc-3′) respectively. Hygromycin phosphotransferase (*hph*) was amplified with oligo pair (5′-cttggctggagctagtggagg-3′/5′-tccttatacaggcggttgc-3′) and the backbone of vector pRF-HU2 (Frandsen et al., 2008) with oligo pair (5′-atcgcccttcccaacagttgcg-3′/5′-acggcctcaacctactactgggc-3′). The four fragments were assembled using NEBuilder® HiFi DNA Assembly Master Mix (NEB, Ipswitch, MA, United States) according to the manufacturer's instructions and transformed into strain ABF_004528_6 (*cad*:3HP+, 3HP+) using agrobacterium mediated transformation as previously described (Michielse et al., 2008) to make strain ABF_008339 (*cad*:3HP+, 3HP+; ΔAp*ald6*).

Sample Preparation

TissueLyser II system (Qiagen, Valencia, CA, United States) 2 mL trays were frozen at −20°C overnight. Two 3 mm stainless steel beads were added to 2 mL snap-cap centrifuge tubes (Eppendorf, Hamburg, Germany) along with 0.5 mL H_2O and each fungus piece. The tubes were placed inside the trays and they were then set on the TissueLyser II system. The frozen samples were ground for 2 min at 30 Hz until it reached a completely homogenized, cold solution. To separate the protein and metabolites, 1 mL cold (−20°C) chloroform:methanol mix [prepared 2:1 (v/v)] was pipetted into a chloroform compatible 2 mL Sorenson MulTI™ SafeSeal™ microcentrifuge tubes (Sorenson bioscience, Salt Lake City, UT, United States) inside an ice-block. Then 200 μl of the sample homogenate in water was added to the Sorenson tube at a ratio of 1:5 sample:chloroform mix [2:1 (v/v)] and vigorously vortexed. The sample was then placed in the ice block for 5 min and then vortexed for 10 s followed by centrifugation at 10,000 × *g* for 10 min at 4°C. The upper water-soluble metabolite phase and the lower lipid soluble phase were also collected into the same vial for metabolomics analysis. The metabolite samples were dried to complete dryness in a speed vac and then capped and stored dry at −20°C until analysis. The remaining protein interphase had 1 mL of cold (−20°C) pure methanol added, vortexed and centrifuged at 12,000 × *g* for 5 min to remove the chloroform and pellet the protein. The methanol supernatant was decanted into waste and the pellet lightly dried in a fume hood and then stored at −80°C until protein digestion.

Protein Digestion

The protein interlayer pellet was digested by adding 200 μl of an 8 nM urea solution to the protein pellets and vortexed into solution. A bicinchoninic acid (BCA) assay (Thermo Scientific, Waltham, MA, United States) was performed to determine protein concentration. Following the assay, 10 mM dithiothreitol (DTT) was added to the samples and incubated at 60°C for 30 min with constant shaking at 800 rpm. Reduced cysteine residues were alkylated by adding 400 mM iodoacetamide (Sigma-Aldrich) to a final concentration of 40 mM and

incubating in the dark at room temperature for 1 h. Samples were then diluted eightfold for preparation for digestion with 100 mM NH_4HCO_3, 1 mM $CaCl_2$ and sequencing-grade modified porcine trypsin (Promega, Madison, WI, United States) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio for 3 h at 37°C. Digested samples were desalted using a 4-probe positive pressure Gilson GX-274 ASPEC^{TM} system (Gilson Inc., Middleton, WI, United States) with Discovery C18 100 mg/1 mL solid phase extraction tubes (Supelco, St. Louis, MO, United States), using the following protocol: 3 mL of methanol was added for conditioning followed by 2 mL of 0.1% triflouroacetic acid (TFA) in H_2O. The samples were then loaded onto each column followed by 4 mL of 95:5: H_2O:ACN, 0.1% TFA. Samples were eluted with 1 mL 80:20 ACN:H_2O, 0.1% TFA. The samples were concentrated down to ∼100 μL using a Speed Vac and a final BCA was performed to determine the peptide concentration, and then stored at –80°C until usage.

Global Proteomics Analysis

Peptides digests were diluted to 0.1 μg/μL with nanopure water for LC-MS/MS analysis. Five μL of samples were loaded onto in-house packed reversed-phase capillary columns (70 cm × 75 μm i.d.) with 3 μm Jupiter C18 particles. The separation was carried out using a nanoAcquity HPLC system (Waters Corporation, Milford, MA, United States) at room temperature. The mobile phase A is 0.1% formic acid in water while the mobile phase B is 0.1% formic acid in acetonitrile. The elution was carried out at 300 nL/min with the following gradient: 0–2 min 1% B; 2–20 min 8% B; 20–75 min 12%B; 75–97 min 30%B; 97–100 min 45%; 100–105 95%; 105–110 min 95%; 110–140 min 1%. The eluting peptides were directly analyzed using a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) in data-dependent acquisition mode. Mass spectrometer settings were as following: full MS (AGC, 3×10^6; resolution, 60,000; m/z range, 300–1800; maximum ion time, 20 ms); MS/MS (AGC, 1×10^5; resolution, 15000; m/z range, 200–2000; maximum ion time, 100 ms; TopN, 12; isolation width, 2 Da; dynamic exclusion, 30.0 s; collision energy, NCE 30).

All mass spectrometry data were searched using MS-GF+ (Kim and Pevzner, 2014) and MASIC (Monroe et al., 2008) software. MS-GF + software was used to identify peptides by scoring MS/MS spectra against peptides derived from the whole protein sequence database. MASIC software was used to generate the selected ion chromatographs (SICs) of all the precursors in MSMS datasets and calculate their peak areas as abundance. MASICResultsMerger[2] was used to append the relevant MASIC stats for each peptide hit result in MS-GF+. The MS-GF+ data were then filtered based on 1% false discovery rate (FDR) and less than 5-ppm mass accuracy to generate a list of qualified peptide-hit results. The abundance of peptides was determined as the highest peak area identified for the peptide within a sample.

Sample level quality was ensured by a robust Principal Component Analysis (PCA) to compute a robust Mahalanobis distance (rMd) based on sample-level parameters (Matzke et al., 2011). Peptides were also filtered to remove those with inadequate data for statistics, either three samples with measured values in one group or two values measured for subjects within two distinct groups. The default for normalization is standard global median centering to account for total abundance differences between samples. A test was performed to assure that these factors are not biases (Webb-Robertson et al., 2010). For this dataset there was no bias detected and we utilized global median centering (Callister et al., 2006). Protein quantification was performed with standard reference-based median averages (Polpitiya et al., 2008; Matzke et al., 2013). Statistics were performed with established standard methods (Webb-Robertson et al., 2017). For time we utilized an ANOVA with a Dunnett's test to compare all time points back to the first time point within each strain. We also utilized a g-test with a Bonferonni correction to identify qualitative markers both to compare strains or time (Webb-Robertson et al., 2010).

Targeted Proteomics Analysis

Targeted Proteomics was performed via Liquid Chromatography (LC) – Selected Reaction Monitoring (SRM) approach. 3–5 peptides per protein were initially selected based on either their identification in global proteomics or their SRM suitability scores predicated by CONSeQuence (Eyers et al., 2011) and Prego (Searle et al., 2015) software tools. All the peptides were further blasted to ensure their uniqueness to target proteins in the organism. Crude synthetic heavy isotope-labeled (e.g., 13C/15N on C-terminal lysine and arginine) peptides were purchased from New England Peptide (Gardner, MA, United States). Upon receiving, the crude synthetic heavy peptides were mixed together and diluted with 0.1% formic acid in 15% acetonitrile in water to obtain a nominal concentration of 1.5 pmol/μL for each individual peptide. The heavy peptide mixture stock solution was aliquoted and stored at –80°C until further use.

To develop targeted proteomics assay, all the SRM precursor-fragment ion pairs (i.e., transitions) were first analyzed using LC-SRM by spiking heavy peptides in test samples. 2–3 transitions per peptide and 1–5 peptides per protein were selected in a final assay based on their LC performance, MS response, transition interferences, and endogenous peptide detectability. Collision energies of transitions were obtained using empirical equations provided in Skyline software (MacLean et al., 2010).

For peptide samples, crude heavy peptide mixture stock solution was spiked in the 0.25 μg/μL peptide samples at a nominal concentration of 35 fmol/μL for each peptide. LC-SRM analysis utilized a nanoACQUITY UPLC® system (Waters Corporation, Milford, MA, United States) coupled online to a TSQ Altis^{TM} triple quadrupole mass spectrometer (Thermo Fisher Scientific). The UPLC® system was equipped with an ACQUITY UPLC BEH 1.7 μm C18 column (100 μm i.d. × 10 cm) and the mobile phases were (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. Two μL of sample (i.e., 0.4 μg of peptides) were loaded onto the column and separated using a 110-min gradient profile as follows (min:flow-rate-μL/min:%B): 0:0.4:1, 6:0.6:1, 7:0.4:1, 9:0.4:6, 40:0.4:13, 70:0.4:22, 80:0.4:40, 85:0.4:95, 91:0.5:95, 92:0.5:95, 93:0.5:50, 94:0.5:95, 95:0.6:1, 98:0.4:1. The LC column is operated with a temperature of 42°C. The TSQ Altis^{TM} triple quadrupole

[2]https://omics.pnl.gov/software/masic-results-merger

mass spectrometer was operated with ion spray voltages of 2100 ± 100 V and a capillary inlet temperature of 350°C. Tube lens voltages were obtained from automatic tuning and calibration without further optimization. Both Q1 and Q3 were set at unit resolution of 0.7 FWHM and Q2 gas pressure was optimized at 1.5 mTorr. The transitions were scanned with a 7 min retention time window and a cycle time of 0.8 s.

All the LC-SRM data were imported into Skyline software and the peak boundaries were manually inspected to ensure correct peak assignment and peak boundaries. Peak detection and integration were determined based on two criteria: (1) the same LC retention time and (2) approximately the same relative peak intensity ratios across multiple transitions between the light peptides and heavy peptide standards. The total peak area ratios of endogenous light peptides and their corresponding heavy isotope-labeled internal standards were then exported from Skyline software as Ratio-to-Standard. For each peptide, the total peak area ratios of individual samples were normalized to the average total peak area ratio of all the samples.

Sample level quality was ensured by a robust Principal Component Analysis (PCA) to compute a robust Mahalanobis distance (rMd) based on sample-level parameters (Matzke et al., 2011). The default for normalization is standard global median centering to account for total abundance differences between samples. A test was performed to assure that these factors are not biases (Webb-Robertson et al., 2010). For this dataset, we utilized global median centering (Callister et al., 2006). Protein quantification was performed with standard reference-based median averages (Polpitiya et al., 2008; Matzke et al., 2013). Statistics included standard methods (Webb-Robertson et al., 2017). For this specific dataset, we utilized an analysis of variance (ANOVA) with Tukey's adjustment to compare all strains to on another within each time point. In the time-course study, we utilized an ANOVA with a Dunnett's test to compare all time points back to the first time point within each strain.

Global Intracellular Metabolomics

The stored metabolite extracts were completely dried under speed-vacuum to remove moisture and chemically derivatized as previously reported (Kim et al., 2015). Briefly, the extracted metabolites were derivatized by methoxyamination and trimethylsilyation (TMS), then the samples were analyzed by GC-MS. Samples were run in an Agilent GC 7890A using a HP-5MS column (30 m × 0.25 mm × 0.25 μm; Agilent Technologies, Santa Clara, CA, United States) coupled with a single quadrupole MSD 5975C (Agilent Technologies). One microliter of sample was injected into a splitless port at constant temperature of 250°C. The GC temperature gradient started at 60°C and hold for 1 min after injection, followed by increase to 325°C at a rate of 10°C/min and a 5-min hold at this temperature. Fatty acid methyl ester standard mix (C8-28) (Sigma-Aldrich) was analyzed in parallel as standard for retention time calibration. GC-MS raw data files were processed using the Metabolite Detector software (Hiller et al., 2009). Retention indices (RI) of detected metabolites were calculated based on the analysis of a FAMEs mixture, followed by their chromatographic alignment across all analyses after deconvolution. Metabolites were initially identified by matching experimental spectra to a PNNL augmented version of Agilent GC-MS metabolomics Library, containing spectra and validated retention indices for over 850 metabolites. Then, the unknown peaks were additionally matched with the NIST17/Wiley11 GC-MS library. All metabolite identifications and quantification ions were validated and confirmed to reduce deconvolution errors during automated data-processing and to eliminate false identifications.

Quantification of Extracellular Metabolites

For quantification of metabolites in the spent medium, samples were diluted 1:20 and 20 μL of these diluted samples were dried in a glass vial. Solutions ranging from 0 to 300 μg/mL containing standards of organic acids of interest, including 3HP, were prepared and 20 μL of these in triplicate were dried to build an external calibration curve. Derivatization protocol, analysis by GC-MS and processing of the data was done as described above, but the analysis was targeted to only the metabolites present in the standard mixes. Calibration curves were established for each of the standards in the mixture of organic acids. After linear regression fitting, concentration of the metabolites in the spent medium samples was determined upon consideration of the dilution factor. For HPLC, samples were analyzed using high performance liquid chromatography (HPLC) equipped with a Waters 2414 refractive index detector. A Bio-Rad Aminex HPX-87H ion exclusion column (300 mm × 7.8 mm), heated to 65°C was used for analyte separation. Sulfuric acid (0.0045 M) was used as eluent at a flow rate of 0.55 mL/min. Extracellular metabolite quantities were used as boundary constraints for parsimonious flux balance analysis (pFBA) using OptFlux (Vilaca et al., 2018) with *Aspergillus niger* genome scale metabolic model iJB1325 (Brandl et al., 2018).

RESULTS

Design and Build to Establish 3HP Production in *Aspergillus pseudoterreus*

Production of 3HP in fungi has been explored in a number of species including *Pichia pastoris* (Zhou et al., 2011), *Schizosaccharomyces pombe* (Takayama et al., 2018), and via a variety of pathways in *S. cerevisiae* (Borodina et al., 2015; Kildegaard et al., 2016). However, to our knowledge 3HP production has not been explored in filamentous fungi. *Aspergillus* species, in particular, are notable for their ability to produce commodity chemicals at low pH which is valuable for the production of free acids. We chose *A. pseudoterreus* as a filamentous fungal host in which to establish 3HP production due to its demonstrated ability to produce and tolerate high quantities of itaconic acid at low pH (Okabe et al., 2009). The beta-alanine production pathway (**Figure 1A**), originally demonstrated in *S. cerevisiae* (Borodina et al., 2015), was recapitulated for *A. pseudoterreus* by codon optimizing the aspartate-1-decarboxylase (*pand*), beta-alanine pyruvate transaminase (*bapat*) and 3-hydroxypropionate dehydrogenase

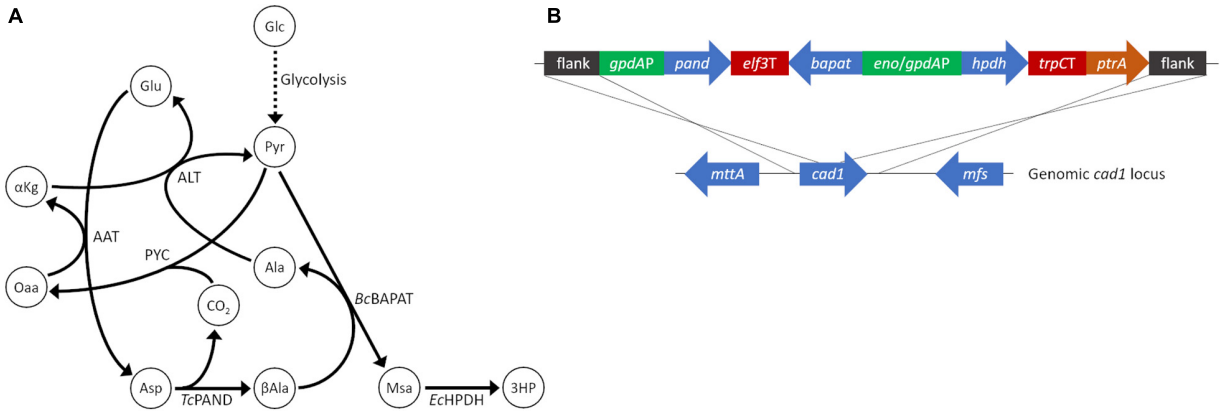


FIGURE 1 | Establishment of 3HP production in *A. pseudoterreus*. **(A)** Ideal carbon flux in a high theoretical yield (2 mol/mol) metabolic pathway for production of 3HP from sugar monomers. **(B)** The beta-alanine metabolic pathway to enable 3HP production was codon optimized for *A. pseudoterreus* and placed under the control of high expression promoters. The engineered pathway was targeted to the *cis*-aconitic acid decarboxylase (*cad*1) locus to disrupt the gene. 3HP, 3-hydroxypropionate; AAT, aspartate aminotransferase, Ala, alanine; ALT, alanine transaminase; αKg, alpha-ketoglutarate; Asp, Aspartate; βAla, beta-alanine; *Bc*BAPAT, beta-alanine pyruvate aminotransferase; *Ec*HPDH, 3-hydroxypropionate dehydrogenase; Glc, glucose; Glu, glutamate; Msa, malonate semialdehyde; Oaa, oxaloacetate; PYC, pyruvate carboxylase; Pyr, pyruvate; *Tc*PAND, aspartate 1-decarboxylase.

(*hpdh*) genes from *Tribolium castaneum*, *Bacillus cereus*, and *E. coli* respectively; and expressing them under the control of high-expression *gpdA* and *enoA* promoters (**Figure 1B**). The major product secreted by *A. pseudoterreus* during industrial fermentations is itaconic acid (Lockwood and Nelson, 1946). To eliminate production of itaconic acid, the beta-alanine pathway overexpression construct was targeted to the *cis*-aconitic acid decarboxylase (*cad1*) locus (Deng et al., 2020) to allow redistribution of metabolic flux elsewhere (**Figure 1B**). Screening of transformants identified a transgenic strain with a single copy of the synthetic pathway integrated at the *cad1* locus (strain ABF_004528_2; *cad*:3HP+) and a strain with one copy of the synthetic pathway integrated at the *cad1* locus and a second copy of the pathway that integrated elsewhere in the genome (strain ABF_004528_6; *cad*:3HP+, 3HP+). Deletion of *cad1* (strain ABF_002234) eliminated itaconic acid production as previously described (Deng et al., 2020). Overexpression of the beta-alanine pathway established 3HP production in *A. pseudoterreus* while a second copy of the biosynthetic pathway enabled nearly twice the yield of extracellular 3HP (**Figure 2B**).

Test/Learn; Multi-Omics Analysis of 3HP

3HP producing strains with single and multi-copy integration of the beta-alanine 3HP pathway were selected and compared to a *cad1* parental strain during an 8-day production period in shake flasks. Flasks were harvested at 2-day intervals and samples collected for intracellular and extracellular metabolomics as well as global and targeted proteomics. In total, global profiling detected 4,674 proteins, while targeted proteomics quantified 79 proteins involved in central carbon metabolism. Nine extracellular metabolites were absolutely quantified while 320 intracellular metabolites were detected, 122 of which were structurally identified (**Figure 2**).

Individual metabolites and proteins were assessed to identify metabolic nodes and pathways that are perturbed in *Aspergillus* strains engineered for 3HP production. The most strongly perturbed proteins and metabolites identified are directly involved in 3HP production as intermediates to the biosynthetic pathway. In the strain with the synthetic 3HP pathway integrated at multiple locations in the genome, we detected 1.66 ± 0.13× as much of the three heterologous proteins (*Tc*PAND, *Bc*BAPAT, and *Ec*HPDH) (**Figure 2B**). Conversely, the intracellular levels of beta-alanine and 3HP are not significantly different in the engineered strains suggesting that the production of extracellular 3HP is not transport limited, nor is the flux toward 3HP limited by either of these reactions. The fact that aspartate is present at a somewhat higher level in the single-copy strain suggests that flux through Aspartate 1-decarboxylase may be the limiting step in the pathway.

The engineered 3HP biosynthetic route from pyruvate involves a series of balanced transamination reactions (alanine transaminase, aspartate aminotransferase, and the introduced beta-alanine pyruvate transaminase; **Figure 1A**) that under ideal conditions cycle nitrogen between alanine, glutamate, aspartate, and beta-alanine. Establishment and increase in expression level of the 3HP biosynthetic pathway did not significantly alter the alanine, glutamate, or aspartate pool sizes though they did tend to decrease in concentration. The beta-alanine pool size substantially increased (148 × – 249 × higher from day 2 to 8) with establishment of the pathway but was not significantly altered by higher expression level. In addition to beta-alanine, we found that the pool size of 2,4-diaminobutanoate is substantially increased (41 × –131 × higher from day 2 to 8) with establishment of the pathway. Metabolic pathways that produce 2,4-diaminobutanoate are not annotated in *Aspergillus* species. We therefore examined the proteome for up-regulated transaminases capable of producing 2,4-diaminobutanoate from substrates that are enriched or suspected to be high-flux in the engineered strains. The most significantly up-regulated enzyme in the engineered

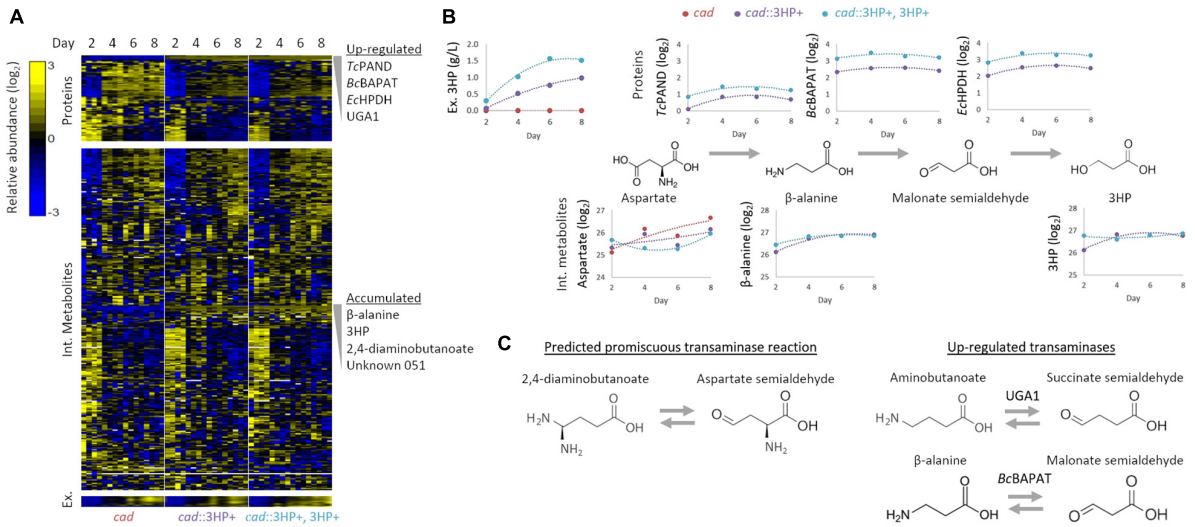


FIGURE 2 | Multi-omics analysis of 3HP producing *A. pseudoterreus* strains. **(A)** Profiling of intracellular and extracellular metabolites and targeted proteomics of central metabolic pathways. **(B)** Effect of establishment and increase in production level on enzymes and metabolites in the engineered 3HP metabolic pathway. **(C)** The metabolite 2,4-diaminobutanoate is accumulated in engineered strains. Upregulated transaminases that may promiscuously use aspartate semialdehyde as a substrate are shown.

strains is the 4-aminobutyrate aminotransferase Ap*uga1*. If Ap*uga1*also functions promiscuously with aspartate-semialdehyde as a substrate (**Figure 2C**), 2,4-diaminobutanoate would be produced rather than 4-aminobutanoate (predicted reaction: aspartate-semialdehyde + glutamate ↔ alpha-ketoglutarate + 2,4-diaminobutanoate). It is also possible that *Bc*BAPAT, overexpressed as part of the engineered 3HP pathway, is promiscuous and produces 2,4-diaminobutanoate with aspartate-semialdehyde as a substrate (predicted reaction: aspartate-semialdehyde + alanine ↔ pyruvate + 2,4-diaminobutanoate). We did not detect 2,4-diaminobutanoate extracellularly suggesting that this by-product of the engineered pathway accumulates as an intracellular nitrogen sink and may compete with nitrogen recycling necessary for high-yield 3HP production.

Test/Learn; 3HP Scale-Up

To assess performance changes during scale-up, the highest producing 3HP strain from shake-flask experiments (*cad*:3HP+, 3HP+) was cultured in a 30 L bioreactor. Off-gas, biomass, and targeted extracellular metabolite concentrations were monitored to provide boundary constraints for parsimonious flux balance analysis (pFBA) (Vilaca et al., 2018). During bioreactor cultivation, biomass and CO_2 were the main products during the first 2 days of cultivation with CO_2 production peaking at hour 44. Production of 3HP and glycerol began early during the exponential growth phase of the culture and prior to peak CO_2 production. In contrast, production of organic acids (aconitic acid, isocitric acid, citric acid, alpha-ketoglutaric acid, and malic acid) began later when biomass production had slowed and become a near-linear process (**Figure 3**). This suggests that production of 3HP via the beta-alanine pathway is primarily a growth linked process, and as such, scale-up requirements may be very different than bioprocesses optimized for production of TCA-cycle derived organic acids in *Aspergillus* species.

Production of extracellular 3HP in bioreactor conditions reached a maximum titer of 0.69 g/L, considerably less than in shake-flasks, at hour 67 and then declined steadily to a final titer of 0.17 g/L. Follow-up experiments where exogenous 3HP was incubated in comparable conditions without the presence of a fungus or with an *A. pseudoterreus* strain not engineered for 3HP production demonstrated that 3HP is stable in acidic culture conditions and degraded biotically (**Figure 4B**). Production of 3HP began tapering off at the on-set of organic acid production suggesting the pathway involved in 3HP degradation is activated during the transition from growth to organic acid production stage.

We examined these results further within the context of metabolic model iJB1325, constructed for the related fungus *Aspergillus niger* (Brandl et al., 2018) and modified by the addition of reactions encoding beta-alanine pyruvate transaminase (beta-alanine + pyruvate ↔ malonate semialdehyde + alanine), 3-hydroxypropionic acid dehydrogenase (malonate semialdehyde + NADPH ↔ 3-hydroxypropionic acid + NADP+), and associated transport reactions and drains. Consumption and production rates for glucose, CO_2, 3HP, aconitic acid, isocitric acid, citric acid, alpha-ketoglutaric acid, malic acid, and glycerol were used to constrain the model for pFBA. We found that the model is able to recapitulate the slow growth of *A. pseudoterreus* observed in the bioreactor but that all consumed 3HP is secreted as beta-alanine regardless of objective function. Within the metabolic network, 3HP can be reversibly metabolized to beta-alanine, but due to irreversibility of aspartate 1-decarboxylase (aspartate → beta-alanine + CO_2) and the energy requirement to metabolize excess beta-alanine through the pantothenate biosynthetic pathway,

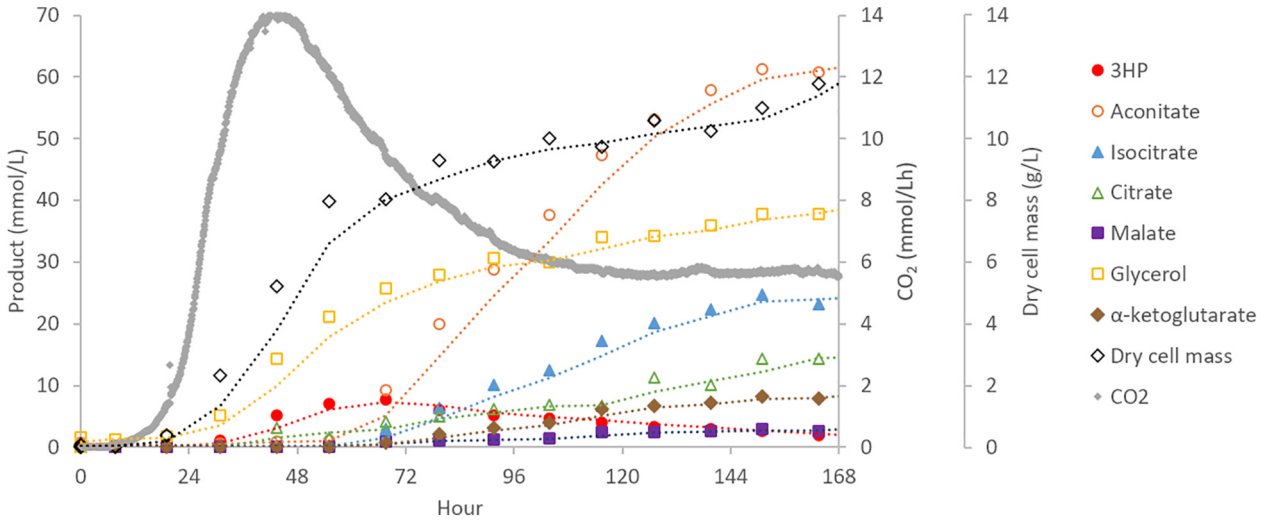


FIGURE 3 | Bioreactor cultivation of *A. pseudoterreus* for 3HP production. Time-course production of 3HP and other bioproducts detected by extracellular metabolomics.

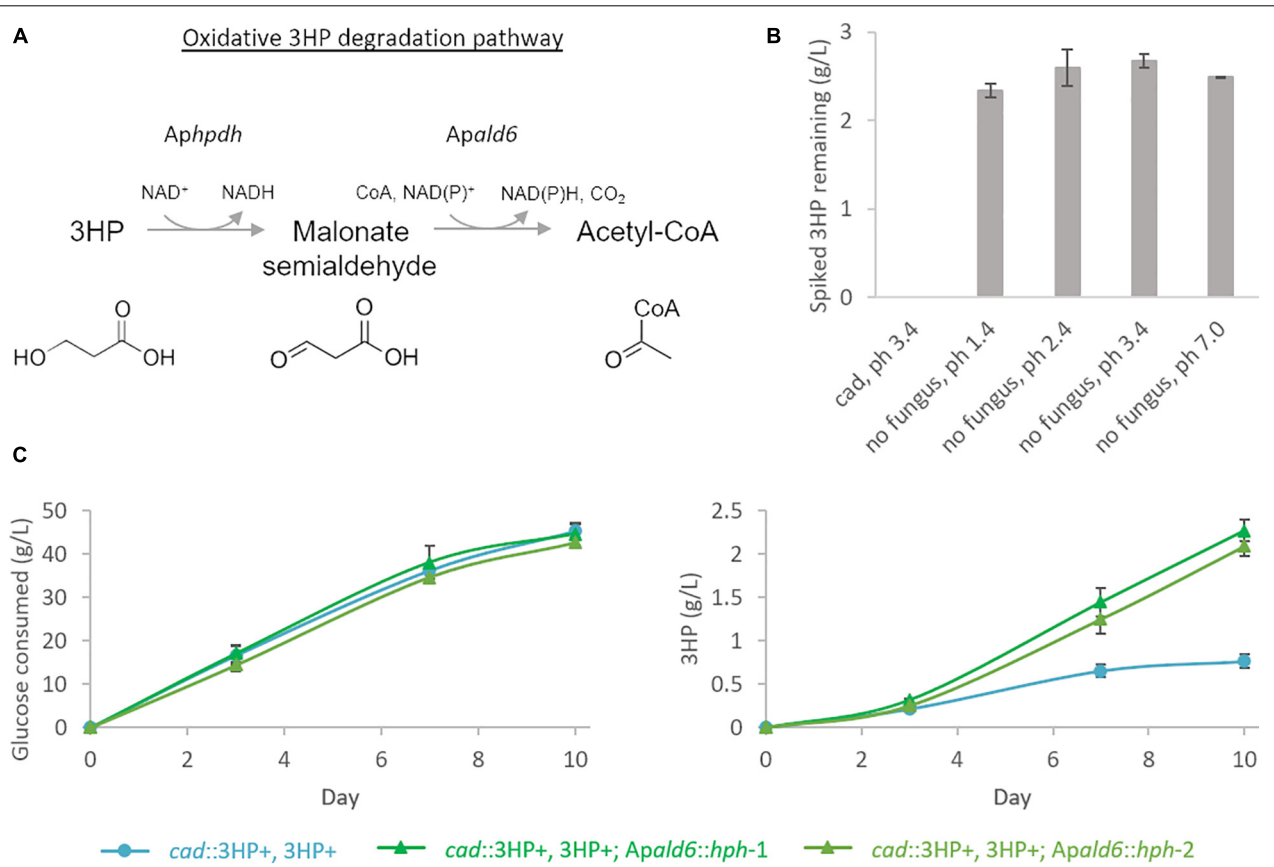


FIGURE 4 | 3HP is degraded biotically through Acetyl-CoA by *A. pseudoterreus*. **(A)** Model for 3HP oxidative degradation pathway present in *A. pseudoterreus*. **(B)** Shake-flasks were spiked with 3HP and incubated for 5 days at 30°C with or without the presence of a strain not engineered for 3HP production (*cad*). Uninoculated flasks were examined across a pH range expected to be experienced during *A. pseudoterreus* growth. The expected abiotic degradation product, acrylic acid was not detected in any flask. 3HP was not detected after 5 days incubation with *A. pseudoterreus* (*cad*). **(C)** Deletion of the putative malonate semialdehyde dehydrogenase (Ap*ald6*) improves yield of 3HP from glucose. Two independent deletion mutants are shown.

it must be secreted under steady-state. However, extracellular beta-alanine was not observed in shake-flask or bioreactor cultures. This suggests that existing pathway(s) to metabolize 3HP are not represented within the metabolic model.

In bacteria pathways have been identified to directly metabolize 3HP via a reductive route to propionyl-CoA (Schneider et al., 2012) or an oxidative route to malonate semialdehyde, essentially the reverse of the final reaction step of the metabolic pathway deployed here (Zhou et al., 2014). Malonate semialdehyde is then degraded to acetyl-CoA by malonate semialdehyde dehydrogenase (Talfournier et al., 2011). In the fungus *Candida albicans*, 3HP is degraded by the oxidative pathway described above by the enzymes Hpdhp and Ald6p (Otzen et al., 2014). Genes encoding enzymes homologous to the oxidative 3HP degradation pathway were identified by BlastP in *A. pseudoterreus*. The best BlastP hit for Ald6p (*E*-value 2E-114) is among the most significantly up-regulated genes during production phase ($p < 1E\text{-}7$ from days 4 to 8) in shake-flasks in strains engineered for 3HP production while the best BlastP hit for Hpdhp (*E*-value 8E-34) was not detected by global proteomics.

It is expected that both the native 3HP dehydrogenase (Ap*hpd1*) and the overexpressed version from *E. coli* (*hpdh*) are reversible and it is therefore unlikely that we could eliminate 3HP degradation while retaining the engineered pathway. We therefore focused on elimination of the competing degradation pathway at the point of malonate semialdehyde by deleting the putative malonate semialdehyde dehydrogenase (Ap*ald6*). In shake-flasks the yield of 3HP is improved by 3.4× in the Ap*ald6* mutant (**Figure 4C**) confirming the presence and impact of this competing pathway in *A. pseudoterreus*.

The ΔAp*ald6* and parent strains were then scaled-up to 0.5 L in bioreactors to re-evaluate 3HP production with pH control and improved aeration (**Figure 5**). By day seven the parent strain had accumulated 0.27 ± 0.10 g/L 3HP while yield was improved by 3.3 × to 0.88 ± 0.11 g/L 3HP in the Ap*ald6* mutant. We did not observe actively decreasing 3HP in shake-flask or small-scale bioreactor experiments as we did in the 20 L bioreactor, but improved yield suggests that flux toward acetyl-CoA is a major competitor with flux toward 3HP from malonate semialdehyde. At 0.5 L scale the rate of CO_2 evolution was slower while higher levels of glycerol and erythritol were accumulated. The major product during cultivation in the 20L bioreactor, aconitic acid, was not observed in the 0.5 L bioreactors supporting the hypothesis that a shift to organic acid production coincides with loss of 3HP. The predicted reaction catalyzed by malonate semialdehyde dehydrogenase (EC 1.2.1.27; **Figure 4A**) was added to the metabolic model. This modification eliminates production of unobserved metabolites during pFBA and improves the consistency of the model with empirical observations from bioreactor and shake-flask cultivations.

DISCUSSION

Biological engineering of microorganisms is a promising route to sustainable manufacturing of fuels and chemicals that are currently derived from the petroleum industry. Here we focused on engineering an acidophilic filamentous fungal host, *A. pseudoterreus*, for conversion of sugars to the polymer precursor 3HP. Production of 3HP using *A. pseudoterreus* is possible in acidic conditions at a pH of 2–3 where 3HP is primarily in the protonated free-acid form, potentially reducing the cost of downstream separations. Overexpression of a synthetic pathway that has been shown to function in yeast (Borodina et al., 2015) demonstrated that production of 3HP is feasible using *A. pseudoterreus* but initial yield was poor. Understanding low productivity is challenging in non-model systems where limited information is available regarding metabolic flux and its control. Here we applied multi-omic analyses of engineered *A. pseudoterreus* strains to generate a systems-level snapshot of metabolites and proteins involved in 3HP production and metabolism as a whole. This approach quantified the effect of initial and increased expression level of the 3HP pathway on the bioconversion host and identified pathways competing with 3HP production.

Establishment of the beta-alanine pathway for 3HP biosynthesis had a greater overall effect on metabolism than increased expression level of the pathway and resulted in substantially increased intracellular pools of beta-alanine, 2,4-diaminobutanoate, and 3HP and up-regulation of malonate-semialdehyde dehydrogenase (Ap*ald6*), enzymes involved in 4-aminobutanoate (GABA) degradation to succinate (Ap*uga1*, Ap*uga2*), and genes that may be involved in metabolism of 3HP pathway intermediates through a CoA bound intermediate (Ap*suc5*, succinyl-CoA:3-ketoacid-CoA transferase; Ap*adh11*, hydroxyacid-oxoacid transhydrogenase; Ap*coaT*, acyl-CoA:carboxylate-CoA transferase; Ap*ahd2*, aldehyde dehydrogenase; Ap*acat1*, keto-acyl-CoA thiolase; and Ap*acat2*, acyl-CoA thiolase), similar to the 3HP/4HB CO_2 fixation cycle described in archaea (Berg et al., 2007).

The GABA degradation response to the 3HP pathway ties nitrogen and carbon metabolism together through succinate. In yeast, *uga1* and *uga2* are up-regulated in response to the presence of GABA (Ramos et al., 1985), however we did not detect a significant impact on the intracellular GABA pool size in response to the presence of the 3HP pathway. 4-hydroxybutanoate (GHB) was also accumulated in the engineered strains (**Supplementary File 1**). In yeast, there is evidence to support the presence of a succinate semialdehyde dehydrogenase to produce GHB (Bach et al., 2009) but not for the reverse reaction catalyzed by GHB dehydrogenase (Mekonnen and Ludewig, 2016), though *Aspergillus* species with a larger genome may catalyze this reaction. Elucidation of the genes responsible in fungi would aid in understanding the effect of 3HP on GABA and GHB metabolism. It is likely that accumulation of 2,4-diaminobutanoate is a direct result of increased Ap*uga1* expression (**Figure 2**), with the proposed precursor aspartate semialdehyde being supplied by an enhanced aspartate pool through the activity of aspartate kinase and aspartate semialdehyde dehydrogenase, though neither enzyme is significantly differentially expressed in response to the presence of the 3HP pathway (**Supplementary File 1**).

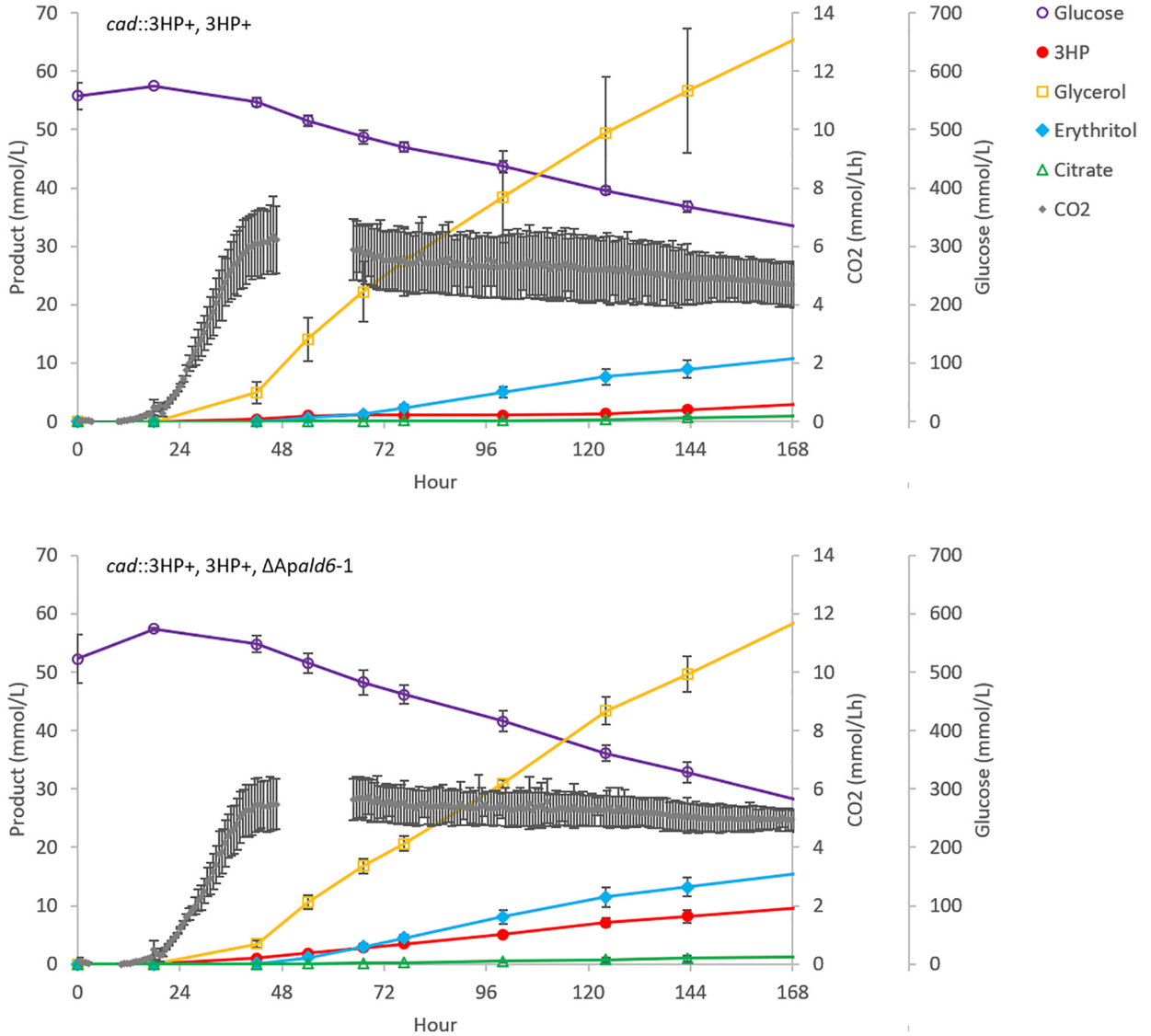


FIGURE 5 | Deletion of A*pald6* increases 3HP yield at 0.5 L bioreactor scale. Time-course production of 3HP and other detected bioproducts. Error bars represent the standard deviation from three replicates.

Extracellular metabolomics during bioreactor cultivation revealed that 3HP is produced and then metabolized at the onset of organic acid production. This led us to search for metabolic pathways involved in catabolism of 3HP, which are poorly characterized in fungi. During shake-flask cultivation the methylmalonate-semialdehyde dehydrogenase (g153; Ap*ald6*) is strongly up-regulated at all times in the engineered strains. This enzyme, normally involved in branched-chain amino acid degradation, has been implicated in 3HP metabolism in bacteria (Zhou et al., 2014; Arenas-Lopez et al., 2019) and the yeast *C. albicans* (Otzen et al., 2014) by also acting as a malonate-semialdehyde dehydrogenase and diverting flux toward acetyl-CoA rather than 3HP. Here we deleted a putative malonate-semialdehyde dehydrogenase Ap*ald6* and found that 3HP yield was drastically improved (**Figure 4**). Reductive degradation of 3HP via propanoyl-CoA synthase has also been demonstrated in bacteria (Schneider et al., 2012), and while homologs of the enzymes involved are not readily identifiable in fungi, we have identified a variety of significantly up-regulated enzymes that may be directly involved in alternative 3HP assimilation pathways. Further genetic characterization will reveal the metabolic repertoire that fungi express to metabolize 3HP.

Multi-omic technologies offer a promising route to systematic analysis of bioconversion processes and identification of targets for improvement of metabolic flux. The results in this study demonstrate that 3HP production can be effectively improved by application of multi-omic analyses to further our system level understanding of genetically engineered *A. pseudoterreus*. However, extracellular metabolomics in shake-flasks and bioreactor scale-up revealed a wide variety of co-products being

produced at levels higher than that of 3HP. These co-products compete metabolically with production of 3HP and contaminate the desired end-product, making separations impractical. Improvement of 3HP production in *A. pseudoterreus* will require minimization of these co-products either through growth optimization or further metabolic engineering.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

KP, ZD, NM, Y-MK, YG, SD, JK, BH, JC, EP, B-JW-R, JZ, HD, SB, KB-J, NH, and JM conceived and designed the experiments. KP, ZD, NM, Y-MK, YG, SD, JK, BH, MS, TL, JC, EP, B-JW-R, CN, and KB-J performed the experiments and analyzed the data. KP, ZD, NM, Y-MK, YG, and JM wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

The research was supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies Office (BETO), under Award No. DE-NL0030038. The multi-omic analysis in the current research was performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. Pacific Northwest National Laboratory is multi-program national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO1830. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.603832/full#supplementary-material

REFERENCES

Arenas-Lopez, C., Locker, J., Orol, D., Walter, F., Busche, T., Kal'nowski, J., et al. (2019). The genetic basis of 3-hydroxypropanoate metabolism in Cupriavidus necator H16. *Biotechnol. Biofuels* 12:150.

Bach, B., Meudec, E., Lepoutre, J. P., Rossignol, T., Blondin, B., Dequin, S., et al. (2009). New insights into {gamma}-aminobutyric acid catabolism: evidence for {gamma}-hydroxybutyric acid and polyhydroxybutyrate synthesis in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* 75, 4231–4239. doi: 10.1128/aem.00051-09

BASF (2014). *BASF, Cargill and Novozymes convert 3-HP to acrylic acid*. Germany: BASF.

Berg, I. A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318, 1782–1786. doi: 10.1126/science.1149976

Borodina, I., Kildegaard, K. R., Jensen, N. B., Blicher, T. H., Maury, J., Sherstyk, S., et al. (2015). Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via beta-alanine. *Metab. Eng.* 27, 57–64. doi: 10.1016/j.ymben.2014.10.003

Brandl, J., Aguilar-Pontes, M. V., Schape, P., Noerregaard, A., Arvas, M., Ram, A. F. J., et al. (2018). A community-driven reconstruction of the Aspergillus niger metabolic network. *Fungal Biol. Biotechnol.* 5:16.

Cairns, T. C., Nai, C., and Meyer, V. (2018). How a fungus shapes biotechnology: 100 years of Aspergillus niger research. *Fungal Biol. Biotechnol.* 5:13.

Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W. J., Webb-Robertson, B. J., et al. (2006). Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 5, 277–286. doi: 10.1021/pr050300l

Chen, Z., Huang, J., Wu, Y., Wu, W., Zhang, Y., and Liu, D. (2017). Metabolic engineering of Corynebacterium glutamicum for the production of 3-hydroxypropionic acid from glucose and xylose. *Metab. Eng.* 39, 151–158. doi: 10.1016/j.ymben.2016.11.009

Davey, S. G. (2018). Sweet new route to acrylonitrile. *Nat. Rev. Chem.* 2:0110.

Decoster, D., Hoyt, S., and Roach, S. (2013). *Dehydration of 3-Hydroxypropionic Acid to Acrylic Acid*. World Intellectual Property Organization No.WO2013192451A1. Geneva: International Bureau.

Deng, S., Dai, Z., Swita, M., Pomraning, K. R., Hofstad, B., Panisko, E., et al. (2020). Deletion analysis of the itaconic acid biosynthesis gene cluster components in Aspergillus pseudoterreus ATCC32359. *Appl. Microbiol. Biotechnol.* 104, 3981–3992. doi: 10.1007/s00253-020-10418-0

Eyers, C. E., Lawless, C., Wedge, D. C., Lau, K. W., Gaskell, S. J., and Hubbard, S. J. (2011). CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol. Cell. Proteomics* 10:M110003384.

Frandsen, R. J., Andersson, J. A., Kristensen, M. B., and Giese, H. (2008). Efficient four fragment cloning for the construction of vectors for targeted gene replacement in filamentous fungi. *BMC Mol. Biol.* 9:70. doi: 10.1186/1471-2199-9-70

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., et al. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* 319, 1215–1220. doi: 10.1126/science.1151721

Hiller, K., Hangebrauk, J., Jäger, C., Spura, J., Schreiber, K., and Schomburg, D. (2009). MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal. Chem.* 81, 3429–3439. 10.1021/ac802689c

Hillson, N., Caddick, M., Cai, Y., Carrasco, J. A., Chang, M. W., Curach, N. C., et al. (2019). Building a global alliance of biofoundries. *Nat. Commun.* 10:2040.

Ji, R. Y., Ding, Y., Shi, T. Q., Lin, L., Huang, H., Gao, Z., et al. (2018). Metabolic Engineering of Yeast for the Production of 3-Hydroxypropionic Acid. *Front. Microbiol.* 9:2185. doi: 10.3389/fmicb.2018.02185

Kildegaard, K. R., Jensen, N. B., Schneider, K., Czarnotta, E., Ozdemir, E., Klein, T., et al. (2016). Engineering and systems-level analysis of Saccharomyces cerevisiae for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb. Cell Fact.* 15:53.

Kim, J. W., Ko, Y. S., Chae, T. U., and Lee, S. Y. (2020). High-level production of 3-hydroxypropionic acid from glycerol as a sole carbon source using metabolically engineered *Escherichia coli*. *Biotechnol. Bioeng.* 117, 2139–2152. doi: 10.1002/bit.27344

Kim, S., and Pevzner, P. A., (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5:5277.

Kim, Y. M., Nowack, S., Olsen, M. T., Becraft, E. D., Wood, J. M., Thiel, V., et al. (2015). Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. *Front. Microbiol.* 6:209. doi: 10.3389/fmicb.2015.00209

Klatt, C. G., Bryant, D. A., and Ward, D. M. (2007). Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in

filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. *Environ. Microbiol.* 9, 2067–2078. doi: 10.1111/j.1462-2920.2007.01323.x

Kuenz, A., and Krull, S. (2018). Biotechnological production of itaconic acid-things you have to know. *Appl. Microbiol. Biotechnol.* 102, 3901–3914. doi: 10.1007/s00253-018-8895-7

Liu, J., Xie, Z., Shin, H. D., Li, J., Du, G., Chen, J., et al. (2017). Rewiring the reductive tricarboxylic acid pathway and L-malate transport pathway of Aspergillus oryzae for overproduction of L-malate. *J. Biotechnol.* 253, 1–9. doi: 10.1016/j.jbiotec.2017.05.011

Lockwood, L. B., and Nelson, G. E. (1946). Some factors affecting the production of itaconic acid by Aspergillus terreus in agitated cultures. *Arch. Biochem.* 10, 365–374.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. doi: 10.1093/bioinformatics/btq054

Matzke, M. M., Brown, J. N., Gritsenko, M. A., Metz, T. O., Pounds, J. G., Rodland, K. D., et al. (2013). A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics* 13, 493–503. doi: 10.1002/pmic.201200269

Matzke, M. M., Waters, K. M., Metz, T. O., Jacobs, J. M., Sims, A. C., Baric, R. S., et al. (2011). Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* 27, 2866–2872. doi: 10.1093/bioinformatics/btr479

Mekonnen, D. W., and Ludewig, F. (2016). Phenotypic and chemotypic studies using Arabidopsis and yeast reveal that GHB converts to SSA and induce toxicity. *Plant Mol. Biol.* 91, 429–440. doi: 10.1007/s11103-016-0475-6

Meng, D. C., Wang, Y., Wu, L. P., Shen, R., Chen, J. C., Wu, Q., et al. (2015). Production of poly(3-hydroxypropionate) and poly(3-hydroxybutyrate-co-3-hydroxypropionate) from glucose by engineering *Escherichia coli*. *Metab. Eng.* 29, 189–195. doi: 10.1016/j.ymben.2015.03.015

Michielse, C. B., Hooykaas, P. J., van den Hondel, C. A., and Ram, A. F. (2008). Agrobacterium-mediated transformation of the filamentous fungus Aspergillus awamori. *Nat. Protoc.* 3, 1671–1678. doi: 10.1038/nprot.2008.154

Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N., and Smith, R. D. (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* 32, 215–217. doi: 10.1016/j.compbiolchem.2008.02.006

Okabe, M., Lies, D., Kanamasa, S., and Park, E. Y. (2009). Biotechnological production of itaconic acid and its biosynthesis in Aspergillus terreus. *Appl. Microbiol. Biotechnol.* 84, 597–606. doi: 10.1007/s00253-009-2132-3

Otzen, C., Bardl, B., Jacobsen, I. D., Nett, M., and Brock, M. (2014). Candida albicans utilizes a modified beta-oxidation pathway for the degradation of toxic propionyl-CoA. *J. Biol. Chem.* 289, 8151–8169. doi: 10.1074/jbc.m113.517672

Polpitiya, A. D., Qian, W. J., Jaitly, N., Petyuk, V. A., Adkins, J. N., Camp, D. G. II, et al. (2008). DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556–1558. doi: 10.1093/bioinformatics/btn217

Ramos, F., el Guezzar, M., Grenson, M., and Wiame, J. M. (1985). Mutations affecting the enzymes involved in the utilization of 4-aminobutyric acid as nitrogen source by the yeast Saccharomyces cerevisiae. *Eur. J. Biochem.* 149, 401–404. doi: 10.1111/j.1432-1033.1985.tb08939.x

Ren, Y., Meng, D., Wu, L., Chen, J., Wu, Q., and Chen, G. Q. (2017). Microbial synthesis of a novel terpolyester P(LA-co-3HB-co-3HP) from low-cost substrates. *Microb. Biotechnol.* 10, 371–380. doi: 10.1111/1751-7915.12453

Riscaldati, E., Moresi, M., Federici, F., and Petruccioli, M. (2000). Effect of pH and stirring rate on itaconate production by Aspergillus terreus. *J. Biotechnol.* 83, 219–230. doi: 10.1016/s0168-1656(00)00322-9

Saha, B. C., Kennedy, G. J., Bowman, M. J., Qureshi, N., and Dunn, R. O. (2019). Factors Affecting Production of Itaconic Acid from Mixed Sugars by Aspergillus terreus. *Appl. Biochem. Biotechnol.* 187, 449–460. doi: 10.1007/s12010-018-2831-2

Schneider, K., Asao, M., Carter, M. S., and Alber, B. E. (2012). Rhodobacter sphaeroides uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. *J. Bacteriol.* 194, 225–232. doi: 10.1128/jb.05959-11

Searle, B. C., Egertson, J. D., Bollinger, J. G., Stergachis, A. B., and MacCoss, M. J. (2015). Using Data Independent Acquisition (DIA) to Model High-responding Peptides for Targeted Proteomics Experiments. *Mol. Cell. Proteomics* 14, 2331–2340. doi: 10.1074/mcp.m115.051300

Strauss, G., and Fuchs, G. (1993). Enzymes of a novel autotrophic CO2 fixation pathway in the phototrophic bacterium Chloroflexus aurantiacus, the 3-hydroxypropionate cycle. *Eur. J. Biochem.* 215, 633–643. doi: 10.1111/j.1432-1033.1993.tb18074.x

Takayama, S., Ozaki, A., Konishi, R., Otomo, C., Kishida, M., Hirata, Y., et al. (2018). Enhancing 3-hydroxypropionic acid production in combination with sugar supply engineering by cell surface-display and metabolic engineering of Schizosaccharomyces pombe. *Microb. Cell Fact.* 17:176.

Talfournier, F., Stines-Chaumeil, C., and Branlant, G. (2011). Methylmalonate-semialdehyde dehydrogenase from Bacillus subtilis: substrate specificity and coenzyme A binding. *J. Biol. Chem.* 286, 21971–21981. doi: 10.1074/jbc.m110.213280

Tripathi, L., Wu, L. P., Meng, D., Chen, J., and Chen, G. Q. (2013). Biosynthesis and characterization of diblock copolymer of p(3-hydroxypropionate)-block-p(4-hydroxybutyrate) from recombinant *Escherichia coli*. *Biomacromolecules* 14, 862–870. doi: 10.1021/bm3019517

Valdehuesa, K. N., Liu, H., Nisola, G. M., Chung, W. J., Lee, S. H., and Park, S. J. (2013). Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical. *Appl. Microbiol. Biotechnol.* 97, 3309–3321. doi: 10.1007/s00253-013-4802-4

Vilaca, P., Maia, P., Giesteira, H., Rocha, I., and Rocha, M. (2018). Analyzing and Designing Cell Factories with OptFlux. *Methods Mol. Biol.* 1716, 37–76. doi: 10.1007/978-1-4939-7528-0_2

Webb-Robertson, B. J., McCue, L. A., Waters, K. M., Matzke, M. M., Jacobs, J. M., Metz, T. O., et al. (2010). Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides from MS-based proteomics data. *J. Proteome Res.* 9, 5748–5756. doi: 10.1021/pr1005247

Webb-Robertson, B. M., Bramer, L. M., Jensen, J. L., Kobold, M. A., Stratton, K. G., White, A. M., et al. (2017). P-MartCancer-Interactive Online Software to Enable Analysis of Shotgun Cancer Proteomic Datasets. *Cancer Res.* 77, e47–e50.

Zhou, F. L., Zhang, Y. G., Zhang, R. B., Liu, W., and Xian, M. (2011). Expression and characterization of a novel propionyl-CoA dehydrogenase gene from Candida rugosa in Pichia pastoris. *Appl. Biochem. Biotechnol.* 165, 1770–1778. doi: 10.1007/s12010-011-9393-x

Zhou, S., Ashok, S., Ko, Y., Kim, D. M., and Park, S. (2014). Development of a deletion mutant of *Pseudomonas* denitrificans that does not degrade 3-hydroxypropionic acid. *Appl. Microbiol. Biotechnol.* 98, 4389–4398. doi: 10.1007/s00253-014-5562-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fbioe.2021.707749

Thermodynamic and Kinetic Modeling of Co-utilization of Glucose and Xylose for 2,3-BDO Production by *Zymomonas mobilis*

Chao Wu[1], Ryan Spiller[1], Nancy Dowe[1,2], Yannick J. Bomble[1] and Peter C. St. John[1]**

[1] Biosciences Center, National Renewable Energy Laboratory, Golden, CO, United States, [2] National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, United States

Prior engineering of the ethanologen *Zymomonas mobilis* has enabled it to metabolize xylose and to produce 2,3-butanediol (2,3-BDO) as a dominant fermentation product. When co-fermenting with xylose, glucose is preferentially utilized, even though xylose metabolism generates ATP more efficiently during 2,3-BDO production on a BDO-mol basis. To gain a deeper understanding of *Z. mobilis* metabolism, we first estimated the kinetic parameters of the glucose facilitator protein of *Z. mobilis* by fitting a kinetic uptake model, which shows that the maximum transport capacity of glucose is seven times higher than that of xylose, and glucose is six times more affinitive to the transporter than xylose. With these estimated kinetic parameters, we further compared the thermodynamic driving force and enzyme protein cost of glucose and xylose metabolism. It is found that, although 20% more ATP can be yielded stoichiometrically during xylose utilization, glucose metabolism is thermodynamically more favorable with 6% greater cumulative Gibbs free energy change, more economical with 37% less enzyme cost required at the initial stage and sustains the advantage of the thermodynamic driving force and protein cost through the fermentation process until glucose is exhausted. Glucose-6-phosphate dehydrogenase (g6pdh), glyceraldehyde-3-phosphate dehydrogenase (gapdh) and phosphoglycerate mutase (pgm) are identified as thermodynamic bottlenecks in glucose utilization pathway, as well as two more enzymes of xylose isomerase and ribulose-5-phosphate epimerase in xylose metabolism. Acetolactate synthase is found as potential engineering target for optimized protein cost supporting unit metabolic flux. Pathway analysis was then extended to the core stoichiometric matrix of *Z. mobilis* metabolism. Growth was simulated by dynamic flux balance analysis and the model was validated showing good agreement with experimental data. Dynamic FBA simulations suggest that a high agitation is preferable to increase 2,3-BDO productivity while a moderate agitation will benefit the 2,3-BDO titer. Taken together, this work provides thermodynamic and kinetic insights of *Z. mobilis* metabolism on dual substrates, and guidance of bioengineering efforts to increase hydrocarbon fuel production.

Keywords: dynamic flux balance analysis, thermodynamics analysis, enzyme protein cost analysis, kinetic models, *Zymomonas mobilis*, biofuel production, 23-butanediol

OPEN ACCESS

Edited by:
Chris Petzold, Lawrence Berkeley National Laboratory, United States

Reviewed by:
Uldis Kalnenieks, University of Latvia, Latvia
Christopher Rao, University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Yannick J. Bomble
yannick.bomble@nrel.gov
Peter C. St. John
peter.stjohn@nrel.gov

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *10 May 2021*
Accepted: *30 June 2021*
Published: *26 July 2021*

Citation:
Wu C, Spiller R, Dowe N, Bomble YJ and St. John PC (2021) Thermodynamic and Kinetic Modeling of Co-utilization of Glucose and Xylose for 2,3-BDO Production by Zymomonas mobilis. Front. Bioeng. Biotechnol. 9:707749. doi: 10.3389/fbioe.2021.707749

INTRODUCTION

Zymomonas mobilis is a facultative anaerobic Gram-negative microorganism, well known for its efficient production of bioethanol as replacement for fossil fuels (Swings and De Ley, 1977; Skotnicki et al., 1982; Sprenger, 1996; Gunasekaran and Raj, 1999; He et al., 2014; Yang et al., 2016a). The bacterium possesses a relatively simple central metabolic network, including a non-functional Embden-Meyerhof-Parnas (EMP) glycolytic route (Fuhrer et al., 2005; Seo et al., 2005; Felczak et al., 2019), incomplete pentose phosphate pathway (Feldmann et al., 1992; De Graaf et al., 1999) and truncated tricarboxylic acid (TCA) cycle (Lee et al., 2010; Jacobson et al., 2019). Instead, hexoses including glucose and fructose are metabolized *via* the Entner-Doudoroff (ED) pathway to form primarily ethanol, along with glycerol and succinic, lactic and acetic acid by-products (Zhang et al., 2019b). Because of its unusual and efficient metabolic machinery, *Z. mobilis* can achieve remarkably high bioprocess efficiency, with up to 98% of hexose carbon converted into ethanol (Kalnenieks et al., 2019). In recent decades, genetic engineering has been used to broaden the spectrum of fermentation substrates and products by *Z. mobilis*. Introduction of exogenous enzymes of xylose isomerase, xylulokinase, transketolase, and transaldolase from *Escherichia coli* endows the bacterium the capability of fermenting pentose sugars (Zhang et al., 1995), whereas *Enterobacter cloacae* derived acetolactate synthase, acetolactate decarboxylase and butanediol dehydrogenase help the redirection of carbon flux to 2,3-butanediol (2,3-BDO), a bio-derived precursor for gasoline and jet fuel (Syu, 2001; Celinska and Grajek, 2009; Ji et al., 2011; Yang et al., 2016b).

Metabolism begins with the uptake of sugar substrates from the media by the glucose facilitator protein, glf, in the cell membrane of *Z. mobilis*. Glf is a low-affinity, high-velocity carrier that also transports xylose but with the competitive inhibition by glucose (Dimarco and Romano, 1985). In fermentation, glucose is preferentially utilized by *Z. mobilis* when co-fermenting with xylose (Zhang et al., 1995). Stoichiometrically, 1 mol of glucose is consumed to form 1 mol of 2,3-BDO with 1 mol of ATP generated *via* the ED pathway as:

$$1\text{glucose} \rightarrow 2,3-\text{BDO} + 2CO_2 + \text{NADPH} + \text{ATP}$$

For xylose, 1.2 mol of ATP will be produced when xylose is metabolized through the completed pentose phosphate pathway and lower glycolysis to produce an equal amount of 2,3-BDO:

$$1.2\text{xylose} \rightarrow 2,3-\text{BDO} + 2CO_2 + 0.2\text{NADH} + 0.8\text{NADPH} + 1.2\text{ATP}$$

Yields of ATP from xylose are higher than for glucose on 2,3-BDO basis in large part because carbon from xylose is rearranged in the pentose phosphate pathway into glyceraldehyde-3-phosphate, whereas for glucose, only half of the carbon flows through glyceraldehyde-3-phosphate due to 2-dehydro-3-deoxy-phosphogluconate aldolase in the ED pathway (**Figure 1A**). Engineered xylose metabolism in *Z. mobilis,* therefore, more closely resembles xylose metabolism in organisms that utilize the EMP pathway. Accordingly, a bioenergetic question arises as to why an energy inefficient glucose metabolism pathway is preferred over xylose metabolism.

Max-min driving force (MDF) optimization and enzyme protein cost estimation have been proposed to evaluate to what extent a metabolic pathway's flux is constrained by thermodynamic driving force and determine the protein expense required to support a pathway's enzymes (Flamholz et al., 2013; Noor et al., 2014). Thermodynamically, the directionality and feasibility of a reaction is determined by the Gibbs free energy change, $\Delta G'$, in which a large negative value indicates a strong driving force for the reaction to proceed forward. Similarly, the driving force of a metabolic pathway can be assessed by solving a maximin problem, which seeks to maximize the negative $\Delta G'$ of the most thermodynamically unfavorable reaction by tuning the concentrations of involved metabolites (Flamholz et al., 2013; Noor et al., 2014; Wu et al., 2020). On the other side, the enzyme level required by a reaction to proceed is determined by both reaction thermodynamics and the kinetic properties of the enzyme, according to a common modular rate law (Liebermeister et al., 2010) and the Haldane relationship (Alberty, 2006). Accordingly, the total enzyme mass required to support a unit pathway flux can be estimated by solving a non-linear optimization with respect to intermediate concentrations (Flamholz et al., 2013; Wu et al., 2020). So far, the methodology has been successfully applied in the thermodynamic comparison between glycolytic pathways in *E. coli* (Flamholz et al., 2013; Noor et al., 2014), and photosynthesis pathways in cyanobacteria with thermodynamic bottlenecks and potential engineering targets identified to enhance CO_2 fixation (Janasch et al., 2019; Wu et al., 2020).

In this work, we first use MDF and enzyme cost to investigate the thermodynamic and kinetic properties of co-utilization of glucose and xylose by a 2,3-BDO producing *Z. mobilis* strain BC42C Δpdc which diverts carbon flux from ethanol biosynthesis through a pyruvate decarboxylase knockout (Zhang and Himmel, 2019; Zhang et al., 2019a). Next, we extend the mathematical modeling from pathway level to the core metabolic network of *Z. mobilis* using dynamic flux balance analysis (dynamic FBA). With a validated model, optimization is performed to identify the optimal initial sugar ratio and agitation conditions for 2,3-BDO fermentation. This work reveals a deeper understanding of mixed C5/C6 sugar metabolism by *Z. mobilis* and sheds light on rational design of bioengineering and fermentation for hydrocarbon fuel production.

RESULTS

Uptake Kinetics of Glucose and Xylose in *Zymomonas mobilis*

Following previous studies (Dimarco and Romano, 1985; Schoberth and de Graaf, 1993; Ren et al., 2009), we began by constructing a Michaelis–Menten kinetics model with competitive inhibition term for the dual-functional transporter in

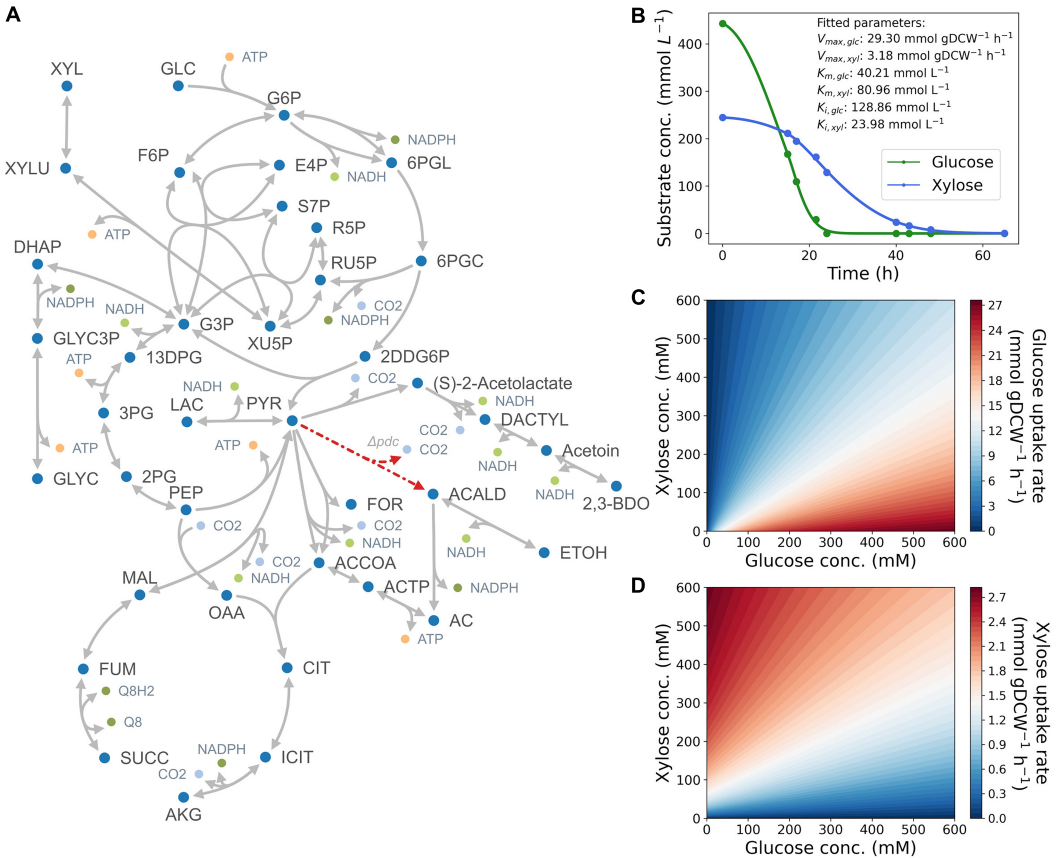


FIGURE 1 | (A) Map of core-carbon metabolism in the engineered *Z. mobilis* strain. A list of abbreviations used are shown in **Supplementary Table 1**. **(B)** Kinetic properties of glucose facilitator (glf) in *Z. mobilis* BC42C Δpdc. Kinetic parameters (V_{max}, K_m, and K_i) are estimated by fitting the time-course utilization of glucose (green) and xylose (blue) with coefficient of determination R^2 = 0.99. Dots denote experimental data points, and lines denote the best fit curve of substrates uptake kinetics. **(C,D)** Contour plots illustrating substrate uptake rates as a function of glucose **(C)** and xylose **(D)** concentrations.

Z. mobilis (Materials and Methods). Kinetics parameters (V_{max}, K_m, and k_i) for both glucose and xylose were estimated by fitting experimental results. As shown in **Figure 1B**, simulated kinetic curves match well with the data (R^2 = 0.99). The fitting results suggest that the maximum transport capacity of glucose is ninefold higher than xylose (29.30 mmol gDCW^{-1} h^{-1} vs. 3.18 gDCW^{-1} h^{-1} in V_{max}), and glucose is twofold more affinitive to the transporter than xylose (40.21 mM vs. 80.96 mM in K_m), which could explain the preferential utilization of glucose when co-fermenting with xylose from a kinetics perspective. It is also noteworthy that the inhibitor constant of glucose was estimated to be much greater than that of xylose, which suggests that competitive inhibition might not be the dominant factor in substrate preference of *Z. mobilis*.

With the estimated kinetic parameters, we simulated glucose and xylose uptake rates at different glucose and xylose concentrations (**Figures 1C,D**). The contour plots illustrate how glucose and xylose uptake rates change as a function of concentrations of both substrates. In general, glucose uptake rate correlates positively with glucose concentration, and negatively with xylose concentration. Xylose uptake rate mirrors the pattern but lowered by almost an order of magnitude, determined by the maximum uptake capacity of the two substrates. Despite the low uptake rate of xylose, transport of this substrate still occurs in the initial phase of fermentation, which leads to simultaneous and not sequential utilization of glucose and xylose.

Thermodynamic Driving Force and Enzyme Protein Cost of Glucose and Xylose Metabolism in *Zymomonas mobilis*

Thermodynamic analysis was performed to compare the feasibility of glucose and xylose metabolism pathways. Kinetic parameters for substrate uptake were estimated from experimental data, while parameters for other pathway enzymes were taken from the Brenda database (Jeske et al., 2019). Three enzyme reactions catalyzed by glucose-6-phosphate dehydrogenase (g6pdh), glyceraldehyde-3-phosphate dehydrogenase (gapdh) and phosphoglycerate mutase (pgm) in the glucose utilization pathway were identified as thermodynamic bottlenecks for positive Gibbs free energy changes assuming 1 mM concentration of substrate and product (**Figure 2**). In addition to the three shared reactions, two more bottlenecks,

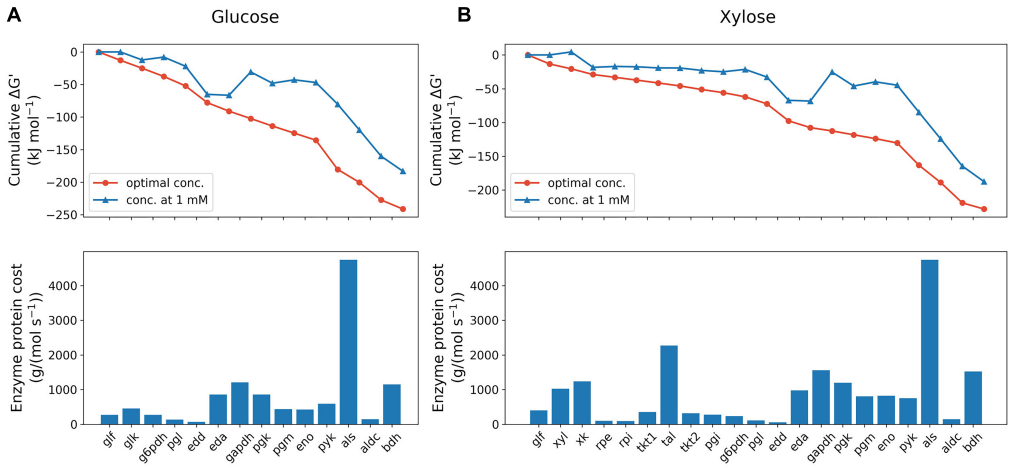


FIGURE 2 | Max-min driving force (MDF) optimization (upper panels) and enzyme protein cost estimation (lower panels) of glucose **(A)** and xylose **(B)** metabolism pathway of *Z. mobilis* BC42C Δpdc at the initial state. MDF optimization is presented as the cumulative sum of reaction Gibbs free energies, ΔG′. Blue line with triangle symbols denotes standard Gibbs free energies with all metabolite concentrations fixed at 1 mM, and red line with circle symbols denotes Gibbs free energies when the minimal ΔG′ is optimized with metabolite concentrations constrained ranging from 1 μM to 10 mM except that extracellular glucose/xylose concentration fixed at measured value. Enzyme protein costs were optimized for the minimal protein mass required to support unit pathway flux under the same concentration constraints with MDF optimization. eda, 2-dehydro-3-deoxy-phosphogluconate aldolase; edd, phosphogluconate dehydratase; eno, enolase; g6pdh, glucose-6-phosphate dehydrogenase; gapdh, glyceraldehyde-3-phosphate dehydrogenase; glf, glucose facilitator protein; glk, glucokinase; pgi, glucose-6-phosphate isomerase; pgl, 6-phosphogluconolactonase; pgm, phosphoglycerate mutase; pgk, phosphoglycerate kinase; pyk, pyruvate kinase; rpe, ribulose-5-phosphate epimerase; rpi, ribose-5-phosphate isomerase; tal, transaldolase; tkt, transketolase; xk, xylulokinase; xyl, xylose isomerase.

xylose isomerase (xyl) and ribulose-5-phosphate epimerase (rpe), were found in the xylose pathway. The MDF of the initial phase of sugar fermentation was optimized for both pathways by tuning concentrations of involved metabolites with extracellular substrates fixed at measured concentrations. The results show that glucose metabolism is initially slightly more favorable (−241.2 vs. −228.0 kJ mol^{-1}) than the xylose pathway thermodynamically.

Enzyme protein costs were estimated in both pathways, which calculate the total protein cost needed to support a unit pathway flux. As shown in **Figure 2**, the total protein requirement of xylose utilization pathway is 58% higher [1.9 × 10^4 g/(mol s^{-1}) vs. 1.2 × 10^4 g/(mol s^{-1})] than glucose metabolism at the beginning of fermentation, which indicates a protein burden for *Z. mobilis* to initiate xylose metabolism when co-fermenting with glucose. It was also found that, likely due to the high K_m value, the exogenous enzyme acetolactate synthase (als) accounts for the dominant amount of protein cost in both pathways: 41% and 25% of total pathway protein requirement for glucose and xylose, respectively. Consistent with a previous study (Yang et al., 2016b), our results suggests that als is the bottleneck in acetolactate generation from pyruvate, and because the limitation is kinetic rather than thermodynamic, represents a potential metabolic engineering target to improve 2,3-BDO production in *Z. mobilis*.

We then extended the MDF optimization and enzyme protein cost estimation to the entire fermentation process. As shown in **Figure 3** as well as **Supplementary Figures 1**, **2**, the MDF and protein requirements of most enzymes vary with depletion of the two substrates. With respect to glucose metabolism, the MDF of majority of the pathway enzymes rises steeply (approaching zero from the negative direction) after 25 h when the glucose concentration was below 3 mM with the exception of phosphogluconate dehydratase (edd), als and acetolactate decarboxylase (aldc) (**Supplementary Figure 1A**), whereas only a gentle increase was seen in enzymes located at the beginning and end of the xylose metabolism pathway (**Supplementary Figure 2A**). In total reaction driving force, glucose utilization maintained thermodynamic favorability over xylose until the glucose concentration was below 2 mM, when xylose achieved a maximum uptake rate.

Similar results were obtained in the time course enzyme protein requirement for both pathways. The protein cost increases drastically with the depletion of substrates to sustain a given uptake rate. As the transporter of glucose and xylose, glf is influenced the most by the substrate exhaustion since protein cost scales inversely proportional to substrate concentration. It is noteworthy that the protein requirement of als is barely affected by the substrate concentration in both pathways and dominates the total protein requirement with sufficient substrate concentrations (**Supplementary Figures 1B**, **2B**). Like the thermodynamic driving force, glucose metabolism is always more economical in total enzyme protein than xylose metabolism until glucose is depleted.

Growth Simulation of *Zymomonas mobilis* With Dynamic Flux Balance Modeling

To further account for cell growth dynamics and predict cellular metabolism of *Z. mobilis*, we extended the stoichiometric model from the pathway-level to the system-level and performed dynamic FBA (Mahadevan et al., 2002). In dynamic FBA, the

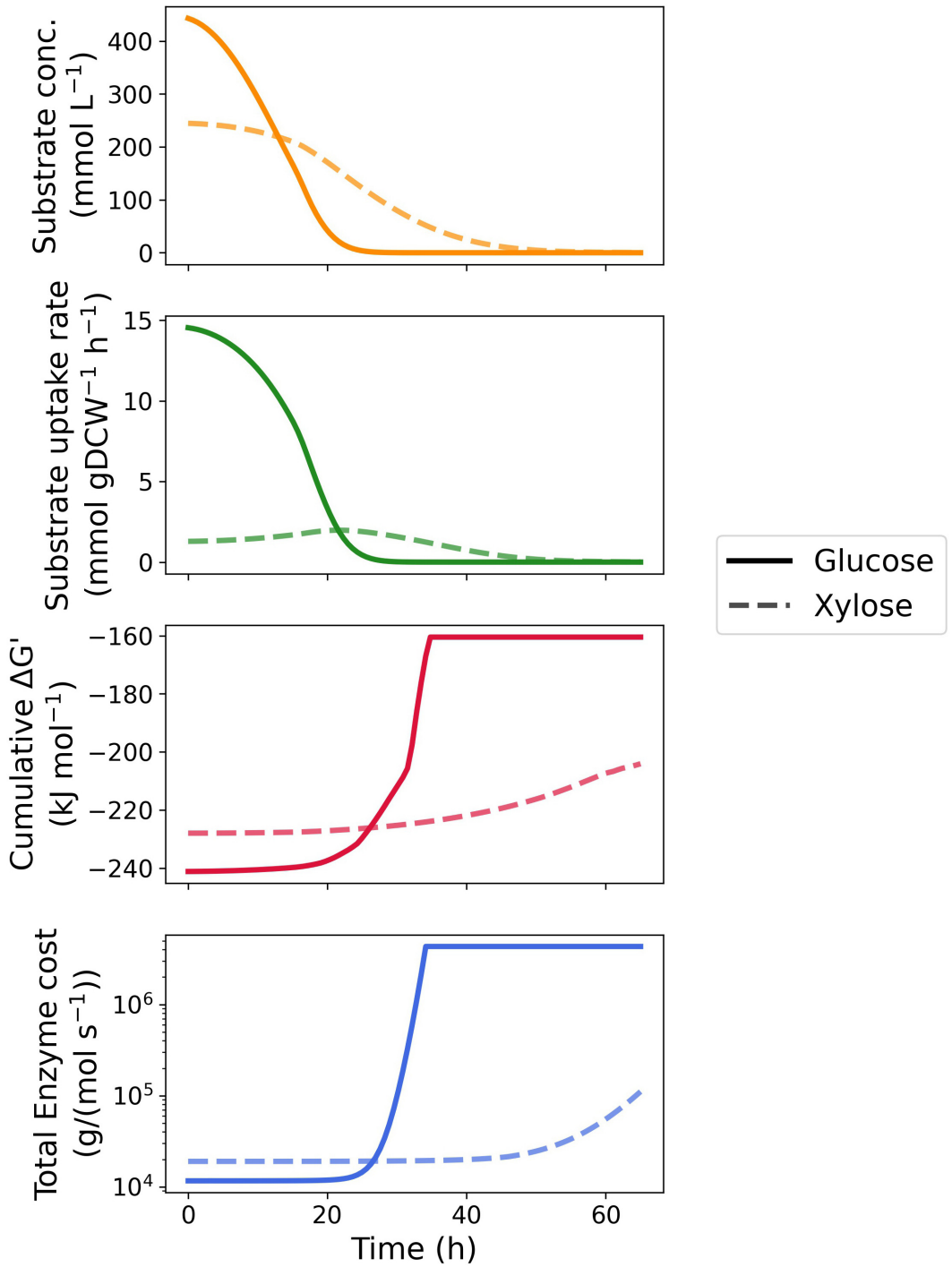


FIGURE 3 | Time-course kinetic and thermodynamic profile of glucose and xylose metabolism by *Z. mobilis* BC42C Δpdc. Substrate concentrations constituted the fitting curve of experimental data. Specific substrate uptake rates were calculated as the first derivative of substrate concentrations divided by corresponding biomass concentrations. MDF optimization and enzyme protein cost estimation were performed at 100 equally spaced timepoints during fermentation with substrate concentration fixed at corresponding values. Only total cumulative ΔG′ and protein cost were illustrated, see **Supplementary Figures 1**, **2** for results of individual enzymes.

timescales associated with cell growth and substrate uptake are assumed to be much faster than the dynamics of intracellular metabolic conversions. This a smaller dynamic system of only the extracellular metabolites to be considered while maintaining a pseudo-steady-state for intracellular metabolite concentrations for FBA. The core reaction network, consisting of 79 reactions

and 70 metabolites, was built based on a genome-scale metabolic reconstructions of *Z. mobilis* (Lee et al., 2010; Pentjuss et al., 2013; Nouri et al., 2020), and includes an incomplete EMP pathway (with phosphofructokinase missing), the ED pathway, the pentose phosphate pathway, and an incomplete TCA cycle (with α-ketoglutarate dehydrogenase and malate dehydrogenase missing, **Figure 1A**). Estimated kinetic parameters were used to describe substrate uptake kinetics of the model. The lower and upper bounds of metabolic flux through pyruvate decarboxylase were set to zero to mimic the knockout of the associated gene.

The measured time course for extracellular metabolites and predictions by the dynamic FBA model for batch aerobic growth of *Z. mobilis* BC42C Δpdc on mixed glucose and xylose media are shown in **Figure 4**. The model simulations show good agreement with the experimental data, except for an overprediction of biomass accumulation. It could be attributable to the unmeasured metabolite leak in the fermentation. Nevertheless, the leaked metabolites will have less effect on modeling of the formation of other products since metabolic flux is additive, and the excess biomass predicted by model could have been converted into the unmeasured metabolites. Diauxic growth was not evident since the two substrates, glucose, and xylose, were consumed simultaneously. The accumulation of acetoin was delayed until glucose was exhausted, and 2,3-BDO concentration began to decrease when both sugar substrates were depleted. As a reduced product, 2,3-BDO can be re-oxidized to generate NADH, which with an external electron acceptor, can be used to produce ATP for cell maintenance. This phenomenon is implemented by addition of reuptake kinetics of 2,3-BDO in the dynamic flux balance model.

Identification of Optimal Growth Conditions for 2,3-BDO Production

Since our dynamic flux balance model was validated to make accurate prediction of growth on dual substrates, the model was further used to identify *in silico* the optimal fermentation conditions for 2,3-BDO production. For this purpose, maximum 2,3-BDO productivity was used as the optimization objective, defined as the maximal ratio of extracellular concentration of 2,3-BDO divided by the fermentation time. First, the initial concentration of glucose and xylose was tuned to see whether a simultaneous depletion of the two sugars can be achieved. The ratio of glucose and xylose was varied from 0.5 to 4 while the total C-mol of substrates was fixed as a constant. As shown in **Figure 5A**, the biomass accumulates with little change between different glucose:xylose ratios; however, the consumption time of the two substrates rarely deviates either. We found that formation of reduced products, glycerol and 2,3-BDO, increase with the ratio of glucose to xylose in the media, which can be attributed to the more efficient production of reducing equivalents by glucose metabolism. Accordingly, *in silico* optimization shows that the highest maximum 2,3-BDO productivity of 14.6 mmol L^{-1} h^{-1} at 36.2 h can be reached with a glucose:xylose ratio of 4, though the maximum productivity is insensitive to the ratio changes.

The effect of different levels of oxygen supply was also studied for which the oxygen transfer coefficient, k_La, was selected as the tuning parameter. With an increased oxygen supply, our model predicts faster growth and substrate consumption (**Figure 5B**). In the model, enhanced uptake of oxygen consumes extra reducing power as an effective acceptor of electrons, thereby stimulating the regeneration of redox cofactors by sugar metabolism. Changing the oxygen supply also diverts carbon flux from glycerol to acetoin and 2,3-BDO, since acetoin and 2,3-BDO are produced through the generation of NADH and NADPH, whereas glycerol production primarily serves to capture excess electrons generated through g6pdh when the cell lacks an external electron acceptor. Accordingly, the optimization of oxygen supply suggests that a high agitation rate will benefit the production of 2,3-BDO by *Z. mobilis*. On the other hand, if maximum 2,3-BDO titer is used as the optimization objective, a moderate aeration will be the optimal condition for the formation of the product which is consistent with the previous study (Yang et al., 2016b).

DISCUSSION

To investigate the preferable utilization of glucose over xylose by *Z. mobilis*, a thermodynamic analysis was first performed which showed a continuously stronger driving force of glucose metabolism until the substrate is exhausted. Thermodynamics is the key factor determining the feasibility and efficiency of a metabolic pathway. Rutkis et al. (2013) has reported that ATP consumption exerts a high level of control on the ED glycolysis activity of *Z. mobilis* through kinetic modeling and metabolic control analysis (Kalnenieks et al., 2014). Interestingly, our thermodynamic model also shows an enhanced driving force with a decreased ratio of ATP and ADP in both glucose and xylose metabolism pathway (**Supplementary Figures 3A,B**). Our results support their findings since a decreased ATP/ADP ratio usually suggests an increased ATP requirement, whereas a larger negative $\Delta G'$ indicates a higher ratio between the forward and backward fluxes, that is, stronger metabolic activity in a pathway according to the flux-force relationship (Beard and Qian, 2007; Noor et al., 2014). Similar results were noticed by Jacobson et al. (2019) who quantified the intracellular metabolite concentrations with a ^{13}C labeling approach (Bennett et al., 2008), and found a significant lower ATP/ADP ratio in *Z. mobilis* than in *E. coli* and *Saccharomyces cerevisiae*. *Z. mobilis* thus suffers an energy shortage and sustains a high ED pathway activity for ATP production (Seki et al., 1990; Kalnenieks et al., 2014; Jacobson et al., 2019). Our results also show that the ratio between reducing equivalents and their oxidized form positively correlates with the cumulative $\Delta G'$ of both glucose (**Supplementary Figures 3C,E**) and xylose (**Supplementary Figures 3D,F**) metabolism. This suggests reducing power might have a controlling effect on *Z. mobilis* glycolytic activity as well, while more theoretical and experimental evidence is required.

Next, enzyme protein cost was estimated for both substrate utilization pathways at the initial state and during the

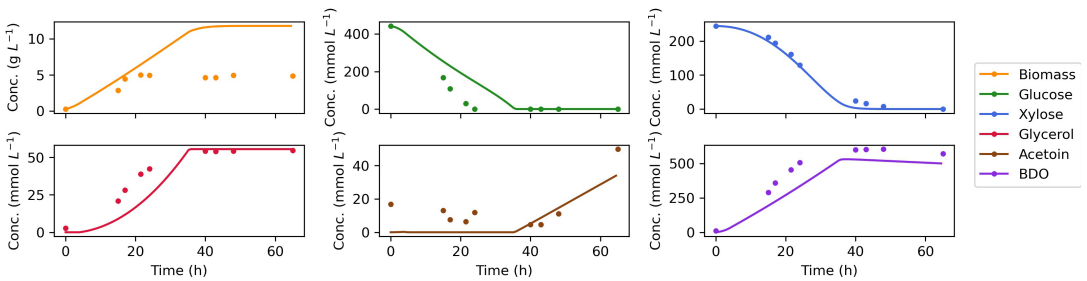


FIGURE 4 | Comparison of dynamic flux balance model predictions (lines) and experimental data (dots) for *Z. mobilis* BC42C Δpdc co-utilizing glucose and xylose. The estimated kinetic parameters are used in substrate uptake kinetics. The lower bound of non-growth associated maintenance (NGAM) is set to zero. Reuptake of products acetoin and BDO are allowed for a better match of the experimental results.

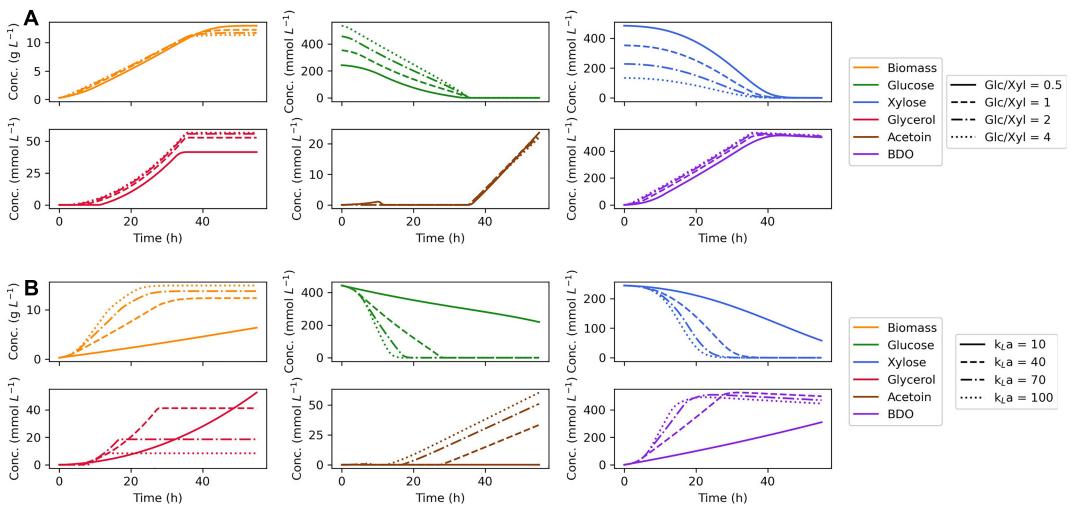


FIGURE 5 | Effects of initial substrate concentration ratio **(A)** and oxygen supply **(B)** on predictions of dynamic flux balance model. In the investigation of initial substrate concentration, a gradient of glucose:xylose ratios are tested keeping the total substrate C mol as a constant. Different agitation and oxygen supply levels are studied by varying the mass transfer coefficient ($k_L a$).

fermentation process. A number of assumptions must be made in building the protein cost model. Besides those basic ones for modeling pathway protein cost (Flamholz et al., 2013) and those for application of the common rate law (Wu et al., 2020), two more assumptions are required for the time course analysis. First, the entire fermentation process is divided into small enough time intervals, and metabolic steady state is achieved and sustained during each interval, which means that the transition between two steady states is instantaneous. This assumption is similar to the one used in the static optimization approach to solve a dynamic flux balance model (Mahadevan et al., 2002), and thus is reasonable for our time course analysis. Second, we assume that the expression level of native enzymes is not affected by introduction of exogenous enzymes. This assumption might be violated due to interactions between the introduced enzymes and local proteins or cofactors; however, it serves as the best approximation until further experimental evidence is available. Notably, the protein cost has the unit g/(mol s^{-1}), which yields mass proportion of pathway enzyme to cell dry weight multiplied by metabolic flux in mmol $gDCW^{-1}$ s^{-1}. It suggests that our time course protein cost estimation could be further validated with dynamic proteomics analyses which is, however, beyond the scope of this research.

Finally, a global stoichiometric model was constructed to simulate and optimize the growth of *Z. mobilis* by dynamic FBA. Our results show that higher 2,3-BDO productivity can be achieved with increased oxygen supply. However, fully aerobic growth for 2,3-BDO productivity is not possible since there exists a balance between excess oxygen supply where producing acetoin is favorable to 2,3-BDO and insufficient aeration that benefits the formation of glycerol. More experiment trials are therefore required to identify and specify the balance. Experimental investigation of whether increased oxygen concentrations in microaerobic fermentations improves growth rates for our engineering *Z. mobilis* strain remains as future work, as growth for ethanol-producing *Z. mobilis* typically does not benefit from oxygen availability. Taken together, with an insight of the dynamic behavior of metabolic networks, the approach shows the potential of providing profound guidance in industrial applications of biofuel production.

MATERIALS AND METHODS

The Strain and Measurement of Growth, Substrate, and Product Kinetics in Batch Fermentations

The *Z. mobilis* strain used in batch fermentations is BC42C Δpdc which is capable of 2,3-BDO production without ethanol biosynthesis by a pyruvate decarboxylase knockout (Zhang and Himmel, 2019; Zhang et al., 2019a). The biological cultures were kept frozen in a −80°C freezer using a 20% glycerol solution. The strains were revived from frozen culture on rich media (10 g L^{-1} yeast extract, 2 g L^{-1} potassium phosphate) and 50 g L^{-1} glucose (RMG 5%) in a 50 mL baffled flask with 10 mL media. The flasks were incubated for 6 h at 30°C and 180 rpm in a shaker incubator. The revived culture was used to start the seed for the primary fermentation. The seed media was prepared in a 125 mL baffled shake flask with RMG 8% (80 g L^{-1} glucose) at a 50 mL working volume. The revived culture was transferred to the flask at a volume to achieve an initial optical density (OD) of 0.1 when measured with a spectrometer at 600 nm wavelength. The seed flask was incubated at 30°C and 180 rpm overnight in a shaking incubator. Fermentations to generate experimental design data for 2,3-BDO production were conducted in BioStat-Q plus fermentors with a 300 mL working volume. The specific fermentation parameters that were used in the described kinetic model were a constant temperature of 30°C, pH 5.8 controlled with KOH (4 N), 100 mL air flowrate through overlay rings, and an agitation at 700 rpm while glucose was present that was reduced to 300 rpm when only xylose was remaining.

Samples were taken throughout the batch fermentations for analyses. Fermentation samples were centrifuged, and supernatants were filtered through a 0.2-μm syringe filter before placing in high-pressure liquid chromatography (HPLC) vials. The samples were analyzed for carbohydrates and BDO (both Meso and SS stereoisomers), acetoin, glycerol, and other by-products. Carbohydrate analysis was done on the Shodex SP0810 carbohydrate column, and organic acids analysis was performed using the Bio-Rad Aminex HPX-87H organic acids column. Sugar utilization, 2,3-BDO, acetoin, and glycerol titers and yields were calculated based on the HPLC data.

Estimation of Kinetic Parameters of Substrate Uptake

The glucose and xylose uptake of *Z. mobilis* is described by Michaelis–Menten kinetics with terms for their mutual competitive inhibition (Dimarco and Romano, 1985), which is formulated as:

$$v_{glc} = \frac{V_{max,glc} \cdot [glc]}{K_{m,glc} \cdot (1 + \frac{(xyl)}{K_{i,xyl}}) + [glc]} \quad (1)$$

$$v_{xyl} = \frac{V_{max,xyl} \cdot [xyl]}{K_{m,xyl} \cdot (1 + \frac{(glc)}{K_{i,glc}}) + [xyl]} \quad (2)$$

where [glc] and [xyl] are extracellular concentrations of glucose and xylose, kinetic parameters $V_{max,glc}$ and $V_{max,xyl}$ are the maximum uptake rate of each sugar, $K_{m,glc}$ and $K_{m,xyl}$ are corresponding Michaelis-Menten constants that indicate the affinities of the reactants for enzyme, and $K_{i,glc}$ and $K_{i,xyl}$ are inhibition constants. The dynamics of substrate uptake is described with the following ordinary differential equations (ODEs):

$$\frac{d[glc]}{dt} = -v_{glc} \cdot [X] \quad (3)$$

$$\frac{d[xyl]}{dt} = -v_{xyl} \cdot [X] \quad (4)$$

where [X] is the biomass concentration, for which values were obtained by a simple linear interpolation between experimental datapoints. The ODEs were solved using odeint in the Python package scipy, and the kinetic parameters were estimated by solving the least squares problem minimizing the difference between experimentally measured and simulated substrate concentrations using openopt. The optimization was repeated several times and always converged to a stable solution. The coefficient of determination R^2 was also calculated to evaluate the goodness of fit.

Max-Min Driving Force Optimization and Enzyme Protein Cost Estimation

The MDF and protein cost analysis were conducted to assess the thermodynamic feasibility and protein synthesis expense of both glucose and xylose utilization pathways using a Python-based pathway analysis tool PathParser (Wu et al., 2020). The enzymes involved as well as their thermodynamic and kinetic parameters used for calculation were listed in **Supplementary Tables 2**, **3**. Standard reaction Gibbs free energies were searched in eQuilibrator (Flamholz et al., 2012). Michaelis-Menten constants, catalyst rate constants and molecular weights of *Z. mobilis* enzymes were taken from BRENDA (Jeske et al., 2019). Geometric mean was used if multiple values were available. For heterologous proteins, parameters were taken for enzymes in their host species. Kinetic parameters for acetolactate synthase, acetolactate decarboxylase and butanediol dehydrogenase were taken from *E. cloacae* (Yang et al., 2016b), while values from *E. coli* were used for xylose isomerase, xylulokinase, transketolase and transaldolase (Zhang et al., 1995). For the substrate uptake reaction, irreversible Michaelis-Menten kinetics was applied to sugar transporter glf with zero Gibbs free energy change. During optimization, extracellular sugar concentrations were fixed at experimentally measured values, and intracellular metabolite concentrations were allowed to vary between 1 μM and 10 mM (Flamholz et al., 2013). To perform the time course MDF optimization and enzyme protein cost estimation, the batch time was divided into many time intervals in which a metabolic steady state was assumed. Accordingly, the extracellular glucose and xylose concentrations were set to the lowest values obtained by fitting in each time interval.

Dynamic Flux Balance Model

In addition to glucose and xylose metabolism and 2,3-BDO synthesis pathways, the core metabolic reactions of *Z. mobilis* BC42C Δpdc was expanded to include an incomplete EMP pathway (with phosphofructokinase missing), the ED pathway, pentose phosphate pathway and an incomplete TCA cycle (with α-ketoglutarate dehydrogenase and malate dehydrogenase missing) (Lee et al., 2010; Pentjuss et al., 2013; Nouri et al., 2020). The constructed flux balance model consists of 79 reactions with 70 related metabolites, which is provided as **Supplementary Data**.

For dynamic flux balance analysis, the uptake kinetics of glucose and xylose was described by Equations 1, 2, while oxygen uptake was described by the following equation:

$$v_{o_2} = \frac{V_{max,o_2} \cdot [o_2]}{K_{m,o_2} + [o_2]} \tag{5}$$

To better match the experimental data, reuptake of fermentation products acetoin and 2,3-BDO were allowed through the following kinetics:

$$v_{actn} = \frac{V_{max,actn} \cdot [actn]}{K_{m,actn} + [actn]} \tag{6}$$

$$v_{bdo} = \frac{V_{max,bdo} \cdot [bdo]}{K_{m,bdo} + [bdo]} \tag{7}$$

where $[O_2]$, [actn], and [bdo] are extracellular concentrations of oxygen, acetoin and 2,3-BDO, $V_{max,o2}$, $V_{max,actn}$ and $V_{max,bdo}$ as well as $K_{m,o2}$, $K_{m,actn}$ and $K_{m,bdo}$ are corresponding maximum uptake rate and Michaelis-Menten constants. The kinetic parameters of glucose and xylose were estimated by fitting the model to experimental data as described above. The maximum uptake rate of oxygen was assumed to be 15 mmol $gDCW^{-1}$ h^{-1} (Mahadevan et al., 2002) and its Michaelis-Menten constant was set to 1.24 μM by experimental measurement (Balodite et al., 2014). V_{max} and K_m of acetoin and 2,3-BDO were assumed to be 10 mmol $gDCW^{-1}$ h^{-1} and 5 mM, respectively.

The extracellular mass balance of above the species, as well as biomass and glycerol, are described by:

$$\frac{d[o_2]}{dt} = -v_{o_2} \cdot [X] + k_L a \cdot [(o_2^*) - (o_2)] \tag{8}$$

$$\frac{d[actn]}{dt} = -v_{actn} \cdot [X] \tag{9}$$

$$\frac{d[bdo]}{dt} = -v_{bdo} \cdot [X] \tag{10}$$

$$\frac{d[X]}{dt} = \mu \cdot [X] \tag{11}$$

$$\frac{d[glyc]}{dt} = v_{glyc} \cdot [X] \tag{12}$$

where $k_L a$ is the mass transfer coefficient for oxygen which is positively correlated with the impeller speed during fermentation (Mastroeni et al., 2003) and assumed to be 30 h^{-1} at 300 rpm. $[O_2^*]$ denotes the oxygen concentration in gas phase and assumed to be a constant 0.21 mM (Mahadevan et al., 2002).

Growth Simulation

Dynamic flux balance analysis of *Z. mobilis* BC42C Δpdc was performed using the metabolic modeling package COBRApy in Python (Ebrahim et al., 2013), and was implemented in a static optimization approach. Maximization of the production of biomass and products and minimization of the consumption of substrates and oxygen were used as lexicographic constraints. Growth was simulated by integration of extracellular mass balance Equations 3, 4, 6–10 using odeint in SciPy package. All growth simulations were performed with inoculum OD of 0.84, and initial glucose and xylose concentrations of 443 and 245 mM, respectively, matching experimental conditions. The batch time for simulation was set to be 65 h, after which both substrates were depleted.

To investigate the effect of the initial sugar ratio on fermentation performance, growth was simulated at different glucose:xylose ratios of 1:2, 1:1, 2:1 and 4:1 with the total C-mol of the substrates as a constant. To investigate the effect of agitation on fermentation performance, growth simulations were performed by varying the mass transfer coefficient, $k_L a$. The fermentation performance was quantified through the maximum 2,3-BDO productivity, defined as the maximal concentration of 2,3-BDO divided by fermentation time.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The scripts and model used in this paper are available at https://github.com/Chaowu88/zymomonas_modeling. The pathway tool PathParser can be found at https://github.com/Chaowu88/PathParser.

AUTHOR CONTRIBUTIONS

CW performed the computational experiments and drafted the manuscript. RS performed the batch fermentations. ND assisted in planning the fermentation experiments. YB and PS helped plan the computational results. All authors assisted in editing the final manuscript.

FUNDING

This work was authored by Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory for the United States Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by United States Department of Energy Office of Energy Efficiency and Renewable Energy Bioenergy Technologies Office.

ACKNOWLEDGMENTS

We would like to thank Min Zhang and Yat-Chen Chou at NREL for sharing the *Zymomonas mobilis* strains. The views expressed in the article do not necessarily represent the views of the DOE or the United States Government. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.707749/full#supplementary-material

REFERENCES

Alberty, R. A. (2006). Relations between biochemical thermodynamics and biochemical kinetics. *Biophysical. Chemistry* 124, 11–17. doi: 10.1016/j.bpc.2006.05.024

Balodite, xE., Strazdina, I., Galinina, N., McLean, S., Rutkis, R., Poole, R. K., et al. (2014). Structure of the Zymomonas mobilis respiratory chain: oxygen affinity of electron transport and the role of cytochrome c peroxidase. *Microbiology* 160, 2045–2052. doi: 10.1099/mic.0.081612-0

Beard, D. A., and Qian, H. (2007). Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS One* 2:e144. doi: 10.1371/journal.pone.0000144

Bennett, B. D., Yuan, J., Kimball, E. H., and Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat. Protoc.* 3, 1299–1311. doi: 10.1038/nprot.2008.107

Celinska, E., and Grajek, W. (2009). Biotechnological production of 2,3-butanediol–current state and prospects. *Biotechnol. Adv.* 27, 715–725. doi: 10.1016/j.biotechadv.2009.05.002

De Graaf, A. A., Striegel, K., Wittig, R. M., Laufer, B., Schmitz, G., Wiechert, W., et al. (1999). Metabolic state of Zymomonas mobilis in glucose-, fructose-, and xylose-fed continuous cultures as analysed by 13C- and 31P-NMR spectroscopy. *Arch. Microbiol.* 171, 371–385. doi: 10.1007/s002030050724

Dimarco, A. A., and Romano, A. H. (1985). d-Glucose transport system of Zymomonas mobilis. *Appl. Environ. Microbiol.* 49, 151–157. doi: 10.1128/aem.49.1.151-157.1985

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Felczak, M. M., Jacobson, T. B., Ong, W. K., Amador-Noguez, D., and TerAvest, M. A. (2019). Expression of phosphofructokinase is not sufficient to enable embden-meyerhof-parnas glycolysis in zymomonas mobilis ZM4. *Front. Microbiol.* 10:2270. doi: 10.3389/fmicb.2019.02270

Feldmann, S. D., Sahm, H., and Sprenger, G. A. (1992). Pentose metabolism in Zymomonas mobilis wild-type and recombinant strains. *Appl. Microbiol. Biotechnol.* 38, 354–361.

Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). eQuilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res.* 40, D770–D775.

Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci.* 110, 10039–10044. doi: 10.1073/pnas.1215283110

Fuhrer, T., Fischer, E., and Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.* 187, 1581–1590. doi: 10.1128/jb.187.5.1581-1590.2005

Gunasekaran, P., and Raj, K. C. (1999). Ethanol fermentation technology – Zymomonas mobilis. *Curr. Sci.* 77, 56–68.

He, M. X., Wu, B., Qin, H., Ruan, Z. Y., Tan, F. R., Wang, J. L., et al. (2014). Zymomonas mobilis: a novel platform for future biorefineries. *Biotechnol. Biofuels* 7:101. doi: 10.1186/1754-6834-7-101

Jacobson, T. B., Adamczyk, P. A., Stevenson, D. M., Regner, M., Ralph, J., Reed, J. L., et al. (2019). (2)H and (13)C metabolic flux analysis elucidates in vivo thermodynamics of the ED pathway in Zymomonas mobilis. *Metab Eng.* 54, 301–316. doi: 10.1016/j.ymben.2019.05.006

Janasch, M., Asplund-Samuelsson, J., Steuer, R., and Hudson, E. P. (2019). Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in Synechocystis carbon fixation. *J. Exp. Bot.* 70, 973–983.

Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47, D542–D549.

Ji, X. J., Huang, H., and Ouyang, P. K. (2011). Microbial 2,3-butanediol production: a state-of-the-art review. *Biotechnol. Adv.* 29, 351–364. doi: 10.1016/j.biotechadv.2011.01.007

Kalnenieks, U., Balodite, E., and Rutkis, R. (2019). Metabolic engineering of bacterial respiration: high vs. low P/O and the Case of Zymomonas mobilis. *Front. Bioeng. Biotechnol.* 7:327. doi: 10.3389/fbioe.2019.00327

Kalnenieks, U., Pentjuss, A., Rutkis, R., Stalidzans, E., and Fell, D. A. (2014). Modeling of Zymomonas mobilis central metabolism for novel metabolic engineering strategies. *Front. Microbiol.* 5:42. doi: 10.3389/fmicb.2014.00042

Lee, K. Y., Park, J. M., Kim, T. Y., Yun, H., and Lee, S. Y. (2010). The genome-scale metabolic network analysis of Zymomonas mobilis ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microb Cell Fact.* 9:94. doi: 10.1186/1475-2859-9-94

Liebermeister, W., Uhlendorf, J., and Klipp, E. (2010). Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26, 1528–1534. doi: 10.1093/bioinformatics/btq141

Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002). 3rd, Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J.* 83, 1331–1340. doi: 10.1016/s0006-3495(02)73903-9

Mastroeni, M. F., Gurgel, P. V., Silveira, M. M., Mancilha, I. M. D., and Jonas, R. (2003). The influence of oxygen supply on the production of acetaldehyde by Zymomonas mobilis. *Brazilian J. Chem. Eng.* 20, 87–93. doi: 10.1590/s0104-66322003000200001

Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., and Milo, R. (2014). Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* 10:e1003483. doi: 10.1371/journal.pcbi.1003483

Nouri, H., Fouladiha, H., Moghimi, H., and Marashi, S. A. (2020). A reconciliation of genome-scale metabolic network model of Zymomonas mobilis ZM4. *Sci. Rep.* 10:7782.

Pentjuss, A., Odzina, I., Kostromins, A., Fell, D. A., Stalidzans, E., and Kalnenieks, U. (2013). Biotechnological potential of respiring Zymomonas mobilis: a stoichiometric analysis of its central metabolism. *J. Biotechnol.* 165, 1–10. doi: 10.1016/j.jbiotec.2013.02.014

Ren, C., Chen, T., Zhang, J., Liang, L., and Lin, Z. (2009). An evolved xylose transporter from Zymomonas mobilis enhances sugar transport in *Escherichia coli*. *Microb Cell Fact* 8:66. doi: 10.1186/1475-2859-8-66

Rutkis, R., Kalnenieks, U., Stalidzans, E., and Fell, D. A. (2013). Kinetic modelling of the Zymomonas mobilis Entner-Doudoroff pathway: insights into control and functionality. *Microbiology* 159, 2674–2689. doi: 10.1099/mic.0.071340-0

Schoberth, S. M., and de Graaf, A. A. (1993). Use of in vivo 13C nuclear magnetic resonance spectroscopy to follow sugar uptake in Zymomonas mobilis. *Anal. Biochem.* 210, 123–128. doi: 10.1006/abio.1993.1161

Seki, M., Furusaki, S., and Shigematsu, K. (1990). Cell growth and reaction characteristics of immobilized zymomonas mobilis. *Ann. N. Y. Acad. Sci.* 613, 290–300. doi: 10.1111/j.1749-6632.1990.tb18170.x

Seo, J. S., Chong, H., Park, H. S., Yoon, K. O., Jung, C., Kim, J. J., et al. (2005). The genome sequence of the ethanologenic bacterium Zymomonas mobilis ZM4. *Nat. Biotechnol.* 23, 63–68.

Skotnicki, M. L., Lee, K. J., Tribe, D. E., and Rogers, P. L. (1982). Genetic alteration of Zymomonas mobilis for ethanol production. *Basic Life Sci.* 19, 271–290. doi: 10.1007/978-1-4684-4142-0_22

Sprenger, G. A. (1996). Carbohydrate metabolism in Zymomonas mobilis: a catabolic highway with some scenic routes. *FEMS Microbiol. Lett.* 145, 301–307. doi: 10.1111/j.1574-6968.1996.tb08593.x

Swings, J., and De Ley, J. (1977). The biology of Zymomonas. *Bacteriol. Rev.* 41, 1–46. doi: 10.1128/mmbr.41.1.1-46.1977

Syu, M. J. (2001). Biological production of 2,3-butanediol. *Appl. Microbiol. Biotechnol.* 55, 10–18. doi: 10.1007/s002530000486

Wu, C., Jiang, H., Kalra, I., Wang, X., Cano, M., Maness, P., et al. (2020). A generalized computational framework to streamline thermodynamics and kinetics analysis of metabolic pathways. *Metab Eng.* 57, 140–150. doi: 10.1016/j.ymben.2019.08.006

Yang, S., Fei, Q., Zhang, Y., Contreras, L. M., Utturkar, S. M., Brown, S. D., et al. (2016a). Zymomonas mobilis as a model system for production of biofuels and biochemicals. *Microb Biotechnol.* 9, 699–717. doi: 10.1111/1751-7915.12408

Yang, S., Mohagheghi, A., Franden, M. A., Chou, Y. C., Chen, X., Dowe, N., et al. (2016b). Metabolic engineering of Zymomonas mobilis for 2,3-butanediol production from lignocellulosic biomass sugars. *Biotechnol. Biofuels* 9:189.

Zhang, K., Lu, X., Li, Y., Jiang, X., Liu, L., and Wang, H. (2019b). New technologies provide more metabolic engineering strategies for bioethanol production in Zymomonas mobilis. *Appl. Microbiol. Biotechnol.* 103, 2087–2099. doi: 10.1007/s00253-019-09620-6

Zhang, M., and Himmel, M. E. (2019). *Targeted Microbial Development (TMD) WBS 2.4.3.102*.

Zhang, M., Chou, Y. C., Franden, M. A., and Himmel, M. E. (2019a). *Engineered Zymomonas for the Production of 2,3-Butanediol*. Alliance for Sustainable Energy LLC. Golden, Colorado.

Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., and Picataggio, S. (1995). Metabolic engineering of a pentose metabolism pathway in ethanologenic zymomonas mobilis. *Science* 267, 240–243. doi: 10.1126/science.267.5195.240

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership