



Streamlining Natural Products Biomanufacturing With Omics and Machine Learning Driven Microbial Engineering

Ahmad Bazli Ramzi^{1*}, Syarul Nataqain Baharum¹, Hamidun Bunawan¹ and Nigel S. Scrutton²

¹ Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, Bangi, Malaysia, ² EPSRC/BBSRC Future Biomanufacturing Research Hub, BBSRC/EPSC Synthetic Biology Research Centre, Manchester Institute of Biotechnology and School of Chemistry, The University of Manchester, Manchester, United Kingdom

OPEN ACCESS

Edited by:

Luan Luong Chu,
Yeungnam University, South Korea

Reviewed by:

Dae-Hee Lee,
Korea Research Institute
of Bioscience and Biotechnology
(KRIBB), South Korea
Hiromichi Minami,
Ishikawa Prefectural University, Japan

Jian Zha,
Shaanxi University of Science
and Technology, China

*Correspondence:

Ahmad Bazli Ramzi
bazliramzi@ukm.edu.my

Specialty section:

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

Received: 22 September 2020

Accepted: 18 November 2020

Published: 21 December 2020

Citation:

Ramzi AB, Baharum SN,
Bunawan H and Scrutton NS (2020)
Streamlining Natural Products
Biomanufacturing With Omics
and Machine Learning Driven
Microbial Engineering.
Front. Bioeng. Biotechnol. 8:608918.
doi: 10.3389/fbioe.2020.608918

Increasing demands for the supply of biopharmaceuticals have propelled the advancement of metabolic engineering and synthetic biology strategies for biomanufacturing of bioactive natural products. Using metabolically engineered microbes as the bioproduction hosts, a variety of natural products including terpenes, flavonoids, alkaloids, and cannabinoids have been synthesized through the construction and expression of known and newly found biosynthetic genes primarily from model and non-model plants. The employment of omics technology and machine learning (ML) platforms as high throughput analytical tools has been increasingly leveraged in promoting data-guided optimization of targeted biosynthetic pathways and enhancement of the microbial production capacity, thereby representing a critical debottlenecking approach in improving and streamlining natural products biomanufacturing. To this end, this mini review summarizes recent efforts that utilize omics platforms and ML tools in strain optimization and prototyping and discusses the beneficial uses of omics-enabled discovery of plant biosynthetic genes in the production of complex plant-based natural products by bioengineered microbes.

Keywords: microbial engineering, synthetic biology, omics technology, machine learning, biomanufacturing, systems biology

INTRODUCTION

Omics-Enabled Discovery of Plant Biosynthetic Genes

Plant natural products represent an enormous resource for chemical and biotechnological production of biopharmaceuticals and natural products-based drugs where about 50–70% of all anti-infective agents in clinical use are being provided and inspired by natural products (Newman and Cragg, 2016). As of 2019, up to 41.3% of anti-infective agents including antiviral and anti-malarial drugs were derived from natural products, which underlines the importance of these compounds as therapeutic agents (Newman and Cragg, 2020). With the advent of systems biology and omics research that have focused on investigating biological mechanisms at systems levels (Kitano, 2002; Lister et al., 2009), a plethora of bioactive compounds and relevant biosynthetic

pathways has since been profiled and identified. This has culminated in the steady expansion of plant-based natural products datasets (Rai et al., 2017).

Driven by the increased availability of bioinformatics tools and high throughput instruments, including next generation sequencing (NGS) and mass spectrometry (MS), omics technologies have been prominently used as principal tools in systems biology research aimed at elucidating the underlying molecular mechanisms behind cellular functions and interplays among biomolecules in biological systems (Fridman and Pichersky, 2005; Sheth and Thaker, 2014; O'Brien et al., 2015). Omics technologies including DNA sequencing (genomics), RNA sequencing (RNA-seq; transcriptomics), and MS-based protein (proteomics) and metabolite (metabolomics) analyses have empowered the reconstruction of metabolic networks based on genome annotation and functional characterization of targeted biochemical reactions in a particular organism or system. The use of systems biology approaches in combination with computational methods has contributed to the generation of genome-scale metabolic models (GEMs) that are important in identifying all metabolic reactions and corresponding biosynthetic genes in various microbes and plants (Seaver et al., 2012; O'Brien et al., 2015).

Importantly, the adoption of single or multi-omics in natural products studies has seen the increment of omics-guided discovery of known and novel metabolites, biosynthetic genes, and regulatory elements from model and non-model plants. By employing transcriptome-guided gene mining and microbial engineering strategies, a number of natural products from previously incomplete and gapped pathways, such as opiate alkaloid noscapine and cannabinoids, have since been produced in microbial hosts, thereby opening up new and exciting opportunities in natural products biomanufacturing using bioengineered microbes as the preferred bioproduction platform (Li and Smolke, 2016; Luo et al., 2019; Courdavault et al., 2020). Biomanufacturing and commercialization of fermentation-based bioproducts, such as artemisinin, nootkatone, and β -farnesene, serve to demonstrate the feasibility and the bioeconomy potential of microbial engineering platforms in the production of fine chemicals and biopharmaceuticals (Benjamin et al., 2016; Ekas et al., 2019). In this mini review, recent applications of systems and synthetic biology approaches in the bioproduction of natural products are discussed where the advancement of natural products biomanufacturing using omics-driven microbial engineering and machine learning (ML)-assisted strain optimization strategies was further highlighted.

INTEGRATION OF SYSTEMS AND SYNTHETIC BIOLOGY FOR MICROBIAL PRODUCTION OF NATURAL PRODUCTS

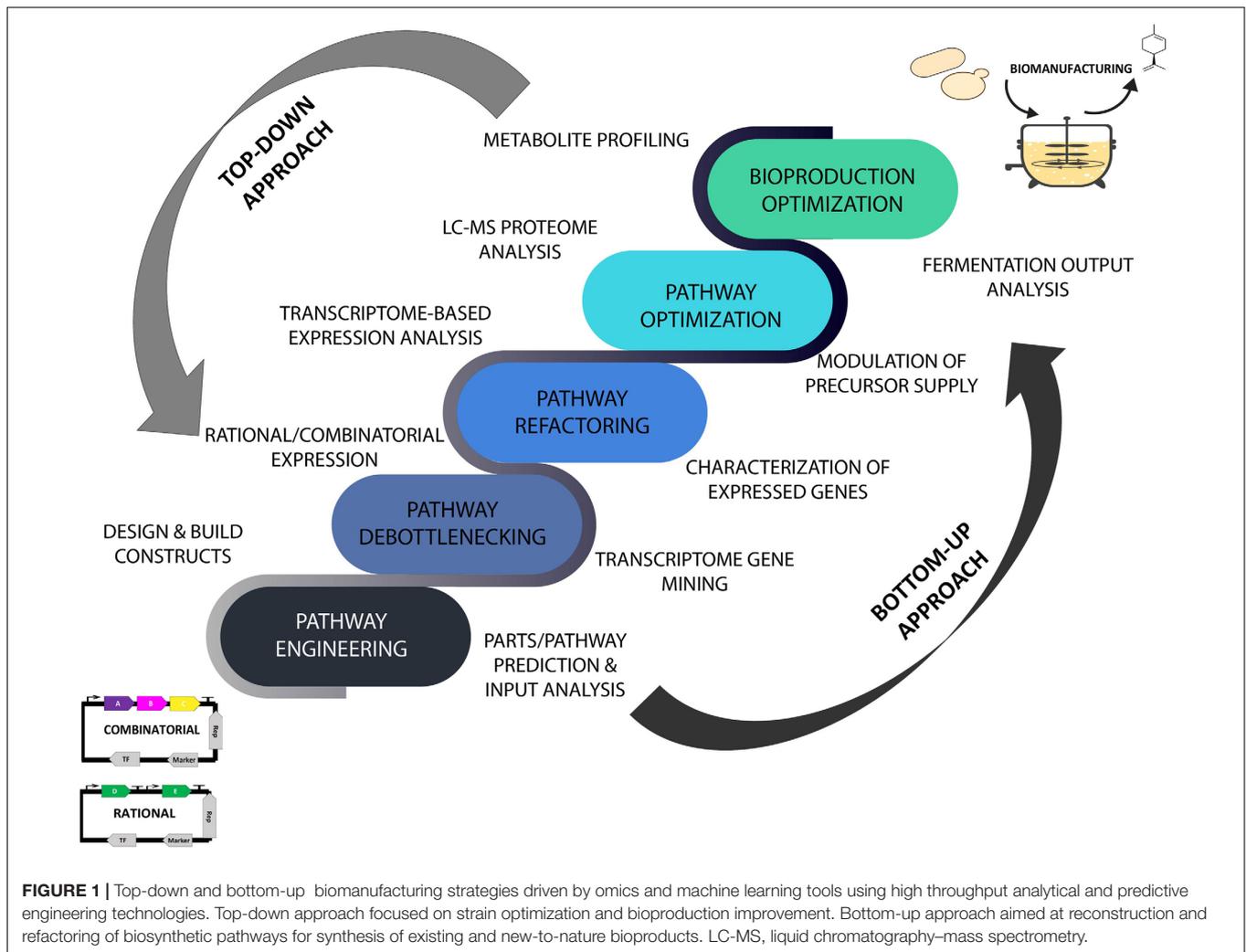
Metabolic engineering and synthetic biology represent advanced bioproduction strategies that have allowed researchers to reprogram and modulate microbial metabolism using genetic and computational tools (Ramzi, 2018; Choi et al., 2019). Multi-omics approaches have been initially established for microbial systems

leading to a growing number of reconstructed GEMs, especially in the universal chassis *Escherichia coli* and *Saccharomyces cerevisiae* where the computational sets of stoichiometric and mass-balanced metabolic reactions in the microbes were derived from genomics-guided experimental analysis including flux balance analysis (FBA) and elementary mode analysis (Gu et al., 2019; Dahal et al., 2020). A host of systems biology, bioinformatics, and computer-aided design (CAD) tools has been developed and utilized to identify cellular metabolic bottleneck, pathway prediction, and gene design with the ultimate aim of enhancing bioproduction titers, rates, and yields (TRYs) by metabolically engineered microbes (Chae et al., 2017; Choi et al., 2019).

The advent of data-driven systems and synthetic biology has brought a renewed and ever-increasing interest in translating laboratory strains into commercial-level microbial prototypes using omics- and *in silico*-guided biomanufacturing platforms that are expected to accelerate the scale-up process and speed up industrial scale production of desired products (Lee and Kim, 2015; Carbonell et al., 2018; Dunstan et al., 2020). The incorporation of the iterative Design-Build-Test-Learn (DBTL) cycle in microbial engineering approaches has provided a biological engineering and *in silico*-assisted framework for strain design and prototyping invaluable for industrial biotechnology applications. As part of the efforts in converging predictive analytics in improving bioproduction capabilities, the employment of metabolome, proteome, transcriptome, and bioinformatics analyses of the plant resources and microbial chassis has provided a comprehensive data-driven means for modulating and streamlining the biomanufacturing process of high-value natural products guided by the DBTL bioengineering framework (Casini et al., 2018; Carqueijeiro et al., 2020). An overview of data-guided bioproduction of natural products using systems and synthetic biology approaches is illustrated in **Figure 1** where the implementation of omics technology and ML tools in improving top-down and bottom-up biomanufacturing strategies is further discussed in the following sections.

TOP-DOWN APPROACH: OMICS-GUIDED STRAIN DESIGN AND PATHWAY OPTIMIZATION

One of the key aspects of strain development using metabolic engineering and synthetic biology tools is the generation and characterization of biosynthetic genes as genetic parts in the pathway design of which the standardization in parts and plasmid assembly allows rapid strain prototyping *via* the DBTL iteration (Nielsen and Keasling, 2016; Robinson et al., 2020). In efforts to maximize TRYs of the natural products and precursor biosynthesis, omics-guided pathway analysis has been applied for a top-down microbial engineering approach by elucidating and identifying affected genes and proteins especially rate-limiting enzymes in engineered metabolic pathways (**Table 1**). In this top-down strain optimization approach, several omics platforms were employed in pathway debottlenecking and optimization in bioengineered microbial



chassis that aimed at improving precursor supply and enhancing targeted natural product biosynthesis in a reverse engineering manner. With the focus on Test and Learn steps, proteome, metabolome, and bioinformatics analyses were conducted for the modulation of endogenous pathway intermediates, such as amino acids and isopentenyl pyrophosphate (IPP)-derived precursors, in bioengineered microbes. In particular, fine-tuning of IPP-related biosynthetic genes was found to be critical in optimizing terpenes bioproduction in engineered *E. coli* and *S. cerevisiae* owing to poor recombinant protein translation and precursor toxicity. Through proteome and transcriptome analyses of terpene-producing strains of *E. coli*, these pathway bottlenecks were debugged through codon optimization of the rate-limiting enzymes and the use of strong and regulated promoters, such as pTrc and pGadE (Redding-Johanson et al., 2011; Dahl et al., 2013). The application of principal component analysis of proteomics (PCAP) and multi-omics approaches in terpene-producing *E. coli* further demonstrated the importance of balanced and optimal protein expression, especially for HMG-CoA reductase, the key enzyme in the IPP-supplying mevalonate (MVA) pathway (Alonso-Gutierrez

et al., 2015). In a seminal report by Brunk et al. (2016) on omics-guided microbial engineering, the combination of GEM, metabolomic, and proteomic analyses has allowed comprehensive pathway mapping and debottlenecking in MVA-derived terpene-overproducing *E. coli* by which several genes in the pentose phosphate pathway, tricarboxylic acid (TCA) cycle, and acetyl-CoA biosynthesis were found to be important in particular by downregulating pyruvate synthase (YDBK) gene that culminated in higher specific production of limonene. In genome engineered *S. cerevisiae*, the use of flux and metabolomic analysis has aided the functional expression of a heterologous 1-deoxy-D-xylulose 5-phosphate (DXP) pathway, the alternative IPP-producing pathway by which combinatorial expression of IspG (2-C-methyl-D-erythritol-2,4-cyclodiphosphate reductase) and IspH (4-hydroxyl-3-methylbut-2-enyl diphosphate reductase) enzymes was tested to overcome the poor conversion of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (MEcPP) and the limited NADPH coenzyme availability (Kirby et al., 2016).

Omics tools have also been utilized in elucidating cellular changes in yeast chassis engineered to produce aromatic phenylpropanoids *via* the shikimate pathway using aromatic

amino acid L-phenylalanine or L-tyrosine as the main entry routes for phenylpropanoid biosynthesis. Metabolomic and transcriptomic analyses of p-coumaric acid (p-CA)-overproducing *S. cerevisiae* revealed distinct transcriptional changes of genes related to sugars and amino acids transport in S288c and CEN.PK background strains that aided in the efforts of systematically modulating the final production of p-CA with up to 20–50% improvement (Rodriguez et al., 2017). Further transcriptome-guided pathway optimization enabled enhanced p-CA bioproduction from xylose in which deletion of the tyrosine and tryptophan amino acid transporter TAT1 resulted in 50% increased of the p-CA titer (Borja et al., 2019). Similar transcriptome-assisted bioengineering strategies were employed to build and test multiple sets of yeast promoters including pINO1, pSED1, and pCCW12 that conferred increased naringenin production from p-CA in engineered *S. cerevisiae* (Gao et al., 2020). Evidently, the application of omic technologies in chassis optimization, especially in the Test and Learn synthetic biology cycle, is inordinately advantageous in pathway debottlenecking and increasing the TRYs of the desired natural products.

BOTTOM-UP APPROACH: OMICS-ENABLED PATHWAY ENGINEERING AND REFACTORIZING FOR NATURAL PRODUCTS BIOMANUFACTURING

The employment of single- and multi-omic tools has brought about a systematic biology-informed pipeline for discovering and biomanufacturing of new-to-nature plant-derived compounds using systems and synthetic biology platforms (Goh, 2018; Chen et al., 2020; Jamil et al., 2020). In the Design and Build steps, genes involved in plant and microbial natural products pathways are considered as important genetic parts by which reconstruction and combinatorial expression of the corresponding biosynthetic pathways have yielded a plethora of industrially important natural products and biochemicals in bioengineered microbes. Discovery of key and missing enzymes in plant biosynthetic pathways has been greatly expedited with transcriptome gene mining of non-model plants and expression of the candidate genes in microbial systems (Goh et al., 2018; Ku Bahaudin et al., 2018; Pyne et al., 2019). Two alkaloid-enriched plants specifically *Papaver somniferum* (opium poppy) and *Catharanthus roseus* (Madagascar periwinkle) have emerged as the model medicinal plants with regard to the employment of multi-omics approaches in the comprehensive analysis of the benzylisoquinoline alkaloids (BIAs) and monoterpene indole alkaloids (MIAs) biosynthetic pathways, respectively (Facchini and De Luca, 2008; Scossa et al., 2018). Using multi-omics strategies, the complete biosynthetic pathway of the anticancer drug vinblastine in *C. roseus* has been finally elucidated where a total of 31 steps are required for MIA compound synthesis from geranyl pyrophosphate (GPP) where the key redox and hydrolase enzymes for the conversion of stemmadenine to

tabersonine or catharanthine were successfully identified via proteome analysis and transcriptome gene mining (Caputi et al., 2018). These omics-driven strategies were similarly employed for the identification and expression of terpene and phenylpropanoid biosynthetic genes from the aromatic plant *Polygonum minus* (*Persicaria minor*) essential for pathway reconstruction and natural product biosynthesis in engineered microbes (Ramzi et al., 2018; Rusdi et al., 2018; Tan et al., 2018).

Transcriptomic-Driven Design and Build of High-Value Natural Products in Microbial Chassis

One of the prominent examples of omics-enabled discovery and production of high-value natural products is the bioproduction of BIAs where candidate genes were obtained from the transcriptome datasets of BIA-accumulating plants, thereby representing a bottom-up approach in natural products biomanufacturing. The production of (S)- and (R)-reticuline was first demonstrated in engineered *S. cerevisiae* through BIA pathway reconstitution that includes the expression of the enzymes norcoclaurine synthase (NCS) and reticuline epimerase (CYP82Y2) from opium poppy *P. somniferum* (DeLoache et al., 2015; Farrow et al., 2015; Table 1). Through gene mining of *P. somniferum* transcriptome datasets, microbial expression of long and complex pathway of BIAs allowed the bioproduction of bioactive dihydrosanguinarine, thebaine, and hydrocodone compounds in engineered *S. cerevisiae* (Fossati et al., 2014; Galanie et al., 2015). Reconstruction and implantation of plant biosynthetic pathways can be modulated and programmed to exploit intrinsic amino acid pathways, such as L-phenylalanine, L-tryptophan, and L-ornithine, thereby removing the metabolic barriers for precursor and energy supply. Combinatorial and rational design strategies have enabled the biosynthesis of tropane alkaloids where *de novo* production of *N*-methylpyrrolinium, tropine, and cinnamoyl tropine has been attained through the incorporation and conversion of L-ornithine- and L-phenylalanine-derived intermediates, respectively, through the expression of corresponding *N*-methyl putrescine oxidase (MPO) from *Nicotiana tabacum* and tropane alkaloid biosynthesis genes from *Anisodus acutangulus* (Ping et al., 2019a,b; Srinivasan and Smolke, 2019).

Transcriptome analysis of antioxidant-rich medicinal plants, including *P. minus* and *Erigeron breviscapus*, revealed the candidate biosynthetic genes for phenylpropanoid-derived flavonoids and breviscapine that shared L-phenylalanine as the main intermediate compound in the plant biosynthetic pathway (Loke et al., 2017; Liu et al., 2018). The introduction of key biosynthetic genes, such as cinnamate-4-hydroxylase (C4H), flavone-6-hydroxylase (F6H), and flavonoid-7-O-glucuronosyltransferase (F7GAT), enabled pathway reconstruction and directed biosynthesis of the desired phenylpropanoid compounds in engineered *S. cerevisiae* using glucose as carbon source (Liu et al., 2018; Ramzi et al., 2018). Interestingly, the presence of endogenous MVA and squalene biosynthetic pathways in *S. cerevisiae* serves as a starting platform for transcriptome-enabled biosynthesis of

TABLE 1 | Omics-guided microbial engineering approaches for natural product and precursor biomanufacturing. Top-down approach mainly represented pathway debottlenecking and strain optimization for increasing bioproduction capacity. Bottom-up approach utilized transcriptome-enabled gene discovery for pathway engineering, refactoring, and bioproduction of industrially important natural products. MVA, mevalonate; PP, pentose-phosphate; TCA, tricarboxylic acid; DXP, 1-deoxy-D-xylulose 5-phosphate; BIA, benzylisoquinoline alkaloid.

Approach	Target metabolite (Chassis)	Key biosynthetic genes and parts	Omics-guided strategy	References
Top-down	Terpene (<i>E. coli</i>)	<p>MVA pathway Mevalonate kinase (MK) and phosphomevalonate kinase (PMK) from <i>S. cerevisiae</i> under the control of <i>trc</i> promoter (<i>E. coli</i>)</p> <p>MVA pathway Farnesyl pyrophosphate (FPP)-responsive promoters PybrL, PgdE, and PrstA controlling FPP biosynthetic genes</p> <p>MVA pathway HMG-CoA synthase (HMGS) and HMG-CoA reductase (HMGR) from <i>Staphylococcus aureus</i>; Terpene synthase from <i>Mentha spicata</i> and <i>Abies grandis</i></p> <p>PP pathway Phosphogluconate dehydrogenase (GND), glucose-6-phosphate dehydrogenase (G6PDH2r) in <i>S. cerevisiae</i></p> <p>TCA cycle Isocitrate dehydrogenase (ICDHγ), alpha-ketoglutarate dehydrogenase (AKGDH) in <i>S. cerevisiae</i></p> <p>Acetyl-CoA biosynthesis Pyruvate synthase (YDBK) in <i>S. cerevisiae</i></p>	<p>Proteome-guided promoter characterization and pathway bottleneck debugging via codon optimization</p> <p>Promoter characterization and pathway intermediate toxicity measurement based on proteome and transcriptome dataset analysis of engineered <i>E. coli</i></p> <p>Application of principal component analysis for enzyme characterization and improvement based on proteome dataset of engineered <i>E. coli</i></p> <p>Pathway precursor supply mapping using multi-omics (Metabolomics and proteomics) and GEM analysis</p>	<p>Redding-Johanson et al., 2011</p> <p>Dahl et al., 2013</p> <p>Alonso-Gutierrez et al., 2015</p> <p>Brunk et al., 2016</p>
		<p>MVA-associated pathway HMG-CoA synthase (ERG13) and membrane protein (PRM10) in <i>S. cerevisiae</i></p> <p>DXP pathway DXP biosynthetic genes (<i>Dxs</i>, <i>IspC</i>, <i>IspD</i>, <i>IspE</i>, <i>IspF</i>, <i>IspH</i>) from <i>E. coli</i>; 2-methyl-buteryl-4-diphosphate (HMBPP) synthase (<i>IspG</i>) from <i>E. coli</i>, <i>Bacillus subtilis</i>, <i>B. thuringiensis</i> and <i>Thermus thermophilus</i></p> <p>DXP-related redox system Flavodoxin/ferredoxin NADP⁺-reductase (AtrFNR) from <i>Arabidopsis thaliana</i> and flavodoxin (Fld) from <i>E. coli</i> and <i>B. subtilis</i></p> <p>Iron-sulfur cluster (ISC) machinery ISC operon (<i>HscA</i>, <i>iscA</i>, <i>cyaY</i>, <i>iscS</i>, <i>iscU</i>, <i>hscB</i>, <i>fdx</i>) and respiratory protein A (<i>erpA</i>) from <i>E. coli</i></p>	<p>Genomics and metabolomics-assisted DXP pathway optimization</p>	<p>Kirby et al., 2016</p>
	Terpene precursor (<i>S. cerevisiae</i>)			
	Phenylpropanoid precursor (<i>S. cerevisiae</i>)	<p>Coumaric acid biosynthesis Tyrosine ammonia lyase (TAL) from <i>Flavobacterium johnsoniae</i>, shikimate kinase (<i>aroL</i>) from <i>E. coli</i>, tyrosine biosynthetic genes (<i>mARO7</i>, <i>mARO4</i>, <i>ARO10</i>) and pyruvate decarboxylase (<i>PDC5</i>) in <i>S. cerevisiae</i></p> <p>Amino acid and sugar transport Tyrosine and tryptophan amino acid transporter (TAT1), polyamine transporter (TPO1), arginine transporter (ALP1), amino acids transporters (BAP2, AGP3), acetate transporter (ADY2) and galactose transporter (GAL2) in <i>S. cerevisiae</i></p> <p>Coumaric acid biosynthesis Shikimate kinase (<i>aroL</i>) from <i>E. coli</i>, tyrosine ammonia-lyase (TAL) from <i>F. johnsoniae</i>, tyrosine biosynthetic genes (<i>mARO7</i>, <i>mARO4</i>, <i>ARO10</i>) and pyruvate decarboxylase (<i>PDC5</i>) from <i>S. cerevisiae</i></p> <p>PP pathway Glucose-6-phosphate dehydrogenase (ZWF1), 6-phosphogluconolactonase (SOL3) and 6-phosphogluconate dehydrogenase (GND1) in <i>S. cerevisiae</i></p>	<p>Metabolic pathway characterization and optimization based on metabolomic and transcriptomic analysis</p> <p>Transcriptome-guided metabolic pathway characterization and optimization</p>	<p>Rodriguez et al., 2017</p> <p>Borja et al., 2019</p>

(Continued)

TABLE 1 | Continued

Approach	Target metabolite (Chassis)	Key biosynthetic genes and parts	Omic-guided strategy	References
	Flavonoid (S. cerevisiae)	Characterized promoters pTDH1, pPGK1, pINO1, pSED1 and pCCW12 in <i>S. cerevisiae</i> Naringenin biosynthesis 4-coumarate:CoA ligase (Ps4CL) from <i>Petroselinum crispum</i> , CHS chalcone synthase from <i>Petunia x hybrida</i> (PhCHS) and CHI chalcone isomerase from <i>Medicago sativa</i> (MsCHI)	Promoter characterization and yield improvement via transcriptomic analysis	Gao et al., 2020
Bottom-up	BIA (S. cerevisiae)	(S)-reticuline biosynthesis Norcoclaurine synthase (PsNCS) from <i>Papaver somniferum</i> (R)-reticuline biosynthesis Reticuline epimerase (PsCYP82Y2) from <i>P. somniferum</i> Dihydrosanguinarine biosynthesis 6-O-methyltransferase (6OMT), coclaurine N-methyltransferase (CNMT), 4'-O-methyltransferase 2 (4'OMT2), truncated berberine bridge enzyme (BBEΔN), cheilanthifoline synthase (PsCFS) and stylophine synthase (PsSPS), cytochrome P450 reductase (PsCPR), tetrahydroprotoberberine cis-N-methyltransferase (TNMT), (S)-cis-N-methylstylophine 14-hydroxylase (MSH) from <i>P. somniferum</i> Opioids biosynthesis 1,2-dehydroreticuline synthase-1,2-dehydroreticuline reductase (DRS-DRR), salutaridine synthase (SalSyn), salutaridine reductase (SalR), salutaridinol 7-O-acetyltransferase (SalAT), thebaine 6-O-demethylase (T6ODM) from <i>P. somniferum</i>	Identification and functional expression of norcoclaurine synthase in <i>S. cerevisiae</i> for (S)-reticuline production from L-tyrosine Identification and functional expression of reticuline epimerase in <i>S. cerevisiae</i> for conversion of (S) to (R)-reticuline Transcriptome gene mining and expression of 10-gene pathway from <i>P. somniferum</i> for biosynthesis of dihydrosanguinarine in <i>S. cerevisiae</i> Transcriptome gene mining, characterization and complete biosynthesis of opioids thebaine and hydrocodone in bioengineered <i>S. cerevisiae</i>	Xiao et al., 2013; DeLoache et al., 2015 Desgagné-Penix et al., 2012; Farrow et al., 2015 Xiao et al., 2013; Fossati et al., 2014 Xiao et al., 2013; Matasci et al., 2014; Galanie et al., 2015
	Tropane alkaloids (S. cerevisiae)	Tropane alkaloid biosynthesis N-methyl putrescine oxidase (MPO) from <i>Nicotiana tabacum</i> Tropane alkaloid biosynthesis Polyketide synthase (AaPYKS), cytochrome p450 (AaP450), tropinone reductase (AaTRI, AaTRII) from <i>Anisodus acutangulus</i>	Transcriptome gene mining, characterization and functional expression of putrescine oxidase in <i>S. cerevisiae</i> Transcriptome gene mining, characterization and functional expression of tropane alkaloid biosynthetic genes in bioengineered <i>S. cerevisiae</i>	Matasci et al., 2014; Srinivasan and Smolke, 2019 Cui et al., 2015; Ping et al., 2019b
	Tropane alkaloids precursor (S. cerevisiae)	Tropane alkaloid biosynthesis Diamine oxidase (DAO) from <i>A. acutangulus</i>	Transcriptome gene mining, characterization and functional expression of diamine oxidase in <i>S. cerevisiae</i>	Cui et al., 2015; Ping et al., 2019a
	Phenylpropanoid precursor (S. cerevisiae)	Phenylpropanoid biosynthesis Cinnamate-4-hydroxylase (C4H) from <i>P. minus</i>	Transcriptome gene mining and expression of cinnamate-4-hydroxylase in <i>S. cerevisiae</i>	Loke et al., 2017; Ramzi et al., 2018
	Breviscapine flavonoid (S. cerevisiae)	Breviscapine flavonoid biosynthesis Flavonoid-7-O-glucuronosyltransferase (F7GAT) and flavone-6-hydroxylase (F6H) from <i>E. breviscapus</i>	Transcriptome gene mining, characterization and reconstitution of complete breviscapine flavonoid pathway from <i>E. breviscapus</i> in bioengineered <i>S. cerevisiae</i>	Liu et al., 2018
	Triterpenoid saponin (S. cerevisiae)	Triterpenoid saponin biosynthesis Cucurbitadienol synthase (SgCDS), epoxide hydrolase (SgEPH3EPH), cytochrome p450 (SgCYP87D18) from <i>Siraitia grosvenorii</i>	Production of mogroside V compounds by bioengineered <i>S. cerevisiae</i> expressing <i>S. grosvenorii</i> enzymes	Tang et al., 2011; Itkin et al., 2016
	Cannabinoids (S. cerevisiae)	Cannabinoids biosynthesis Prenyltransferases (CsPT), tetraketide synthase (<i>C. sativa</i> TKS; CsTKS), olivetolic acid cyclase (CsOAC), acyl activating enzyme (AAE). Cannabinoid synthases THCAS and CBDAS from <i>C. sativa</i>	Transcriptome gene mining, characterization and reconstitution of cannabinoid biosynthetic pathway in bioengineered <i>S. cerevisiae</i>	van Bakel et al., 2011; Luo et al., 2019

cannabinoids and triterpenoid saponin that were naturally derived from *Cannabis sativa* L. and *Siraitia grosvenorii*, respectively. Complete biosynthesis of cannabinoids was demonstrated through the expression of *Cannabis* enzymes that include newly identified *Cannabis* candidate prenyltransferases that are responsible for the conversion of olivetolic acid and GPP supplied by native MVA and heterologous hexanoyl-CoA biosynthetic pathways, respectively (Luo et al., 2019). Using a MVA-dependent squalene pathway in *S. cerevisiae*, the biosynthesis of triterpenoid mogrol compounds was achieved *via* pathway reconstitution and heterologous expression of cucurbitadienol synthase, epoxide hydrolase, and cytochrome p450 identified from *S. grosvenorii* transcriptome (Itkin et al., 2016). Overall, the utilization and expression of transcriptome-derived plant biosynthetic genes represent an increasingly valuable and feasible strategy in pathway engineering and natural product biomanufacturing using bioengineered microbes as cell factories.

THE WAY FORWARD: STREAMLINING NATURAL PRODUCTS BIOMANUFACTURING WITH OMICS AND ML PLATFORMS

To date, model microbes, especially *E. coli* and *S. cerevisiae*, represent the most suitable natural product chassis for strain improvement and biological engineering using DBTL iteration and upscaling processes owing to increased availability of genetic parts and biological data, including GEMs and omics datasets. As discussed earlier, omics technologies have been valuable in enhancing synthetic biology applications, but progress remains in accelerating the Learn step needed to inform the next Design phase and consequent DBTL cycles important in improving the desirable specification and biomanufacturing capacities. Recent progress in advanced genomics and synthetic biology has seen the increased adoption of ML-based data training and non-biased predictive tools for analyzing biological datasets to complement the biology-informed systems biology approaches. The predictive ability of ML tools is empowered through training and learning of experimental data *via* statistical linkage and modeling of independent and dependent variables as input and output data, respectively (Radivojević et al., 2020). Critically, the employment of ML approaches in strain design and optimization is gaining much interest, which is expected to address limitations in biology-informed approaches and circumvent the needs for detailed mechanistic understanding and resource constraints (Carbonell et al., 2019; Presnell and Alper, 2019).

Improving DBTL Performance and Predictive Capacities With ML and Omics Tools

ML-based training of biological datasets has been successfully used in microbial hosts in the efforts to improve gene annotation, metabolic pathway optimization, and fermentation bioprocess parameters (Kim et al., 2020). The bioproduction of specialty

and fine chemicals, such as dodecanol and limonene, has been demonstrated in engineered *E. coli* and *S. cerevisiae* using ML-generated predictive models, which enabled unbiased genetic designs and combination (Zhou et al., 2018; Jervis et al., 2019; Opgenorth et al., 2019). A key advantage of utilizing ML tools is the development of a pure *in silico* system applicable for the Design and Learn phases that enable the selection of high-performing biological system without the needs to perform extensive and costly *in vivo* screening experiments. In the Design step, several ML tools have been developed for optimizing gene expression and cellular protein synthesis through *de novo* and quantitative design of genetic parts including promoter, 5'-untranslated region (5'UTR), and ribosomal binding site (RBS) in addition to the use of ML-assisted directed evolution and semi-rational protein engineering strategies (Decoene et al., 2018; Jervis et al., 2019; Wu et al., 2019). By training of partial least square (PLS) regression model on fluorescence output of a yeast UTR (yUTR) library, a newly constructed yUTR calculator was used to accurately predict the outcome of translation initiation rates in *S. cerevisiae* (Decoene et al., 2018). Employment of the predictive yUTR calculator enabled a tailored *in vivo* p-CA production in tyrosine ammonia lyase (TAL1)-expressing *S. cerevisiae* in accordance to the strengths of *de novo* and native 5'UTR with weak and high predicted protein abundance (Decoene et al., 2018).

Modulation and improvement of terpenes production has also been demonstrated through ML-enabled fine-tuning of gene expression by synthetic promoters and RBS of the MVA and non-MVA DXP biosynthetic genes (Meng et al., 2013; Jervis et al., 2019). Using a mutated Trc promoter and RBS sequences for artificial neural network (NN)-based model training and test, the expression of 1-deoxy-D-xylulose-5-phosphate synthase (DXS) gene under the control low-strength synthetic s14 promoter enhanced the production of amorphaadiene in engineered *E. coli* (Meng et al., 2013). A recent report (Jervis et al., 2019) has expanded the use of a feedforward NN-based ML model on *de novo* design and screening of synthetic RBS for MVA pathway engineering and bioproduction of limonene where model training was conducted on expression levels of HMG-CoA synthase (HMGS), HMG-CoA reductase (HMGR), MVA kinase (MK), and IPP delta-isomerase (IDI) using multiple combinations of RBS sequences. The constructed library of 32 RBS combinations was then built and tested in combination with terpene-synthesizing pGL403 plasmid construct that resulted in the identification and selection of high-performing *E. coli* strains with improved limonene titer over 1.5–3-fold (Jervis et al., 2019).

The combination of omics datasets and ML strategies is expected to drive the production of natural products and other biobased chemicals especially in terms of biosynthetic pathway inference, refactoring, and optimization. The self-organizing map (SOM) approach represents an unsupervised NN method useful in the identification of new enzymes using plant transcriptome datasets to complement common gene co-expression analysis, such as differentially expressed genes (DEG) method (Dugé de Bernonville et al., 2020). The SOM-assisted co-expression analysis of *Rauvolfia serpentina* transcriptome has led to the identification of sarpagan bridge enzyme (SBE) and vinorine

hydroxylase (VH) essential in sarpagan and ajmalan alkaloid biosynthesis that could be useful in the Build and Test of these high-value bioproducts in engineered microbial chassis (Dang et al., 2017, 2018). A supervised ML platform has been developed and tested using proteome and metabolome datasets of biofuel- and terpene-producing *E. coli* where the ML-driven model predictions yielded an accurate *in silico* pathway design and outperformed classical Michaelis–Menten kinetic modeling (Costello and Martin, 2018). In their report, a Tree-based Pipeline Optimization Tool (TPOT) was used for training data and succeeded in generating models for dynamically predicting medium level limonene-producing *E. coli* strains using experimental omics datasets, thus providing a pure ML and omic dataset-based virtual strain simulation and pathway construction (Costello and Martin, 2018). Interestingly, another recent report by Radivojević et al. (2020) leveraged on ensemble approach and probabilistic modeling methods to construct a ML-based Automated Recommendation Tool (ART) useful for improving microbial engineering and DBTL bioproduction performance by training of proteome datasets among a host of experimental data as input variables. By comparing limonene bioproduction improvement in engineered *E. coli* guided by experimentally tested PCAP, the ML models generated by ART were suggested to be able to match and further enhance the production of a given product through the DBTL cycle by recommending new inputs, such as transcriptome datasets and promoter strengths in the next Design phase. Following this, the integration of transcriptome, proteome, and/or metabolome datasets with ML methods is particularly useful in the development of mathematical models in the Test and Learn cycle that would guide and facilitate *in silico* optimization of the DBTL pipeline (Presnell and Alper, 2019; St. John and Bomble, 2019; Volk et al., 2020). Thanks to the growing list of genome, transcriptome, and GEM resources, further adoption and implementation of *in silico* and ML tools on these biological datasets are expected to bring about a markedly improved and accurate predictive engineering and retrosynthetic design of metabolic pathways to existing and new-to-nature chemicals (Lin et al., 2019; Zhang

et al., 2020). In line with the emergence of data-driven 4th Industrial Revolution (4IR), the applications of omics and ML tools in strain and bioproduct development are set to be the cornerstone in industrial biomanufacturing of biobased chemicals and pharmaceuticals.

CONCLUSION AND PERSPECTIVES

Overall, it is envisioned that the employment of data-centered omics and ML platforms will lead to more streamlined and less resource-intensive biomanufacturing strategies and accelerate strain prototyping pipelines that have been a major stumbling block in the translation of bioproduct development from laboratory to market. Omics-guided microbial engineering and ML-assisted biomanufacturing will therefore bring about data-driven biomanufacturing pipelines that can be expanded to include metagenome datasets and accelerate the bioproduction of industrially relevant biomolecules and drugs tailored to the pressing needs of medical, agricultural, environmental, and industrial sectors.

AUTHOR CONTRIBUTIONS

ABR, SNB, HB, and NSS all contributed toward the writing and editing of this manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Yayasan Penyelidikan Antartika Sultan Mizan (YPASM) Fellowship Scheme (RB-2019-001) and Geran Universiti Penyelidikan (GUP-2019-065) research grants. The article processing fee for this work is funded by Dana Pecutan Penerbitan INBIOSIS UKM (PP-INBIOSIS-2020), a publication initiative grant awarded to the Institute of Systems Biology (INBIOSIS) under UKM Research University Grant.

REFERENCES

- Alonso-Gutierrez, J., Kim, E.-M., Batth, T. S., Cho, N., Hu, Q., Chan, L. J. G., et al. (2015). Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metab. Eng.* 28, 123–133. doi: 10.1016/j.ymben.2014.11.011
- Benjamin, K. R., Silva, I. R., Cherubim, J. P., McPhee, D., and Paddon, C. J. (2016). Developing commercial production of semi-synthetic artemisinin, and of β -farnesene, an isoprenoid produced by fermentation of Brazilian sugar. *J. Braz. Chem. Soc.* 27, 1339–1345. doi: 10.5935/0103-5053.2016.0119
- Borja, G. M., Rodriguez, A., Campbell, K., Borodina, I., Chen, Y., and Nielsen, J. (2019). Metabolic engineering and transcriptomic analysis of *Saccharomyces cerevisiae* producing p-coumaric acid from xylose. *Microb. Cell Fact.* 18, 191. doi: 10.1186/s12934-019-1244-4
- 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
- Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T.-D., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* 360, 1235–1239. doi: 10.1126/science.aat4100
- 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
- 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
- Carqueijeiro, I., Langley, C., Grzech, D., Koudounas, K., Papon, N., O'Connor, S. E., et al. (2020). Beyond the semi-synthetic artemisinin: metabolic engineering of plant-derived anti-cancer drugs. *Curr. Opin. Biotechnol.* 65, 17–24. doi: 10.1016/j.copbio.2019.11.017
- Casini, A., Chang, F.-Y., Eluere, R., King, A. M., Young, E. M., Dudley, Q. M., et al. (2018). A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology. *J. Am. Chem. Soc.* 140, 4302–4316. doi: 10.1021/jacs.7b13292

- Chae, T. U., Choi, S. Y., Kim, J. W., Ko, Y.-S., and Lee, S. Y. (2017). Recent advances in systems metabolic engineering tools and strategies. *Curr. Opin. Biotechnol.* 47, 67–82. doi: 10.1016/j.copbio.2017.06.007
- Chen, Y., Banerjee, D., Mukhopadhyay, A., and Petzold, C. J. (2020). Systems and synthetic biology tools for advanced bioproduction hosts. *Curr. Opin. Biotechnol.* 64, 101–109. doi: 10.1016/j.copbio.2019.12.007
- Choi, K. R., Jang, W. D., Yang, D., Cho, J. S., Park, D., and Lee, S. Y. (2019). Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.* 37, 817–837. doi: 10.1016/j.tibtech.2019.01.003
- Costello, Z., and Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst. Biol. Appl.* 4:19. doi: 10.1038/s41540-018-0054-3
- Courdavault, V., O'Connor, S. E., Oudin, A., Besseau, S., and Papon, N. (2020). Towards the microbial production of plant-derived anticancer drugs. *Trends Cancer* 6, 444–448. doi: 10.1016/j.trecan.2020.02.004
- Cui, L., Huang, F., Zhang, D., Lin, Y., Liao, P., Zong, J., et al. (2015). Transcriptome exploration for further understanding of the tropane alkaloids biosynthesis in *Anisodus acutangulus*. *Mol. Genet. Genomics.* 290, 1367–1377. doi: 10.1007/s00438-015-1005-y
- Dahal, S., Yurkovich, J. T., Xu, H., Palsson, B. O., and Yang, L. (2020). Synthesizing systems biology knowledge from omics using genome-scale models. *Proteomics* 20:e1900282. doi: 10.1002/psm.1000282
- Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., et al. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* 31, 1039–1046. doi: 10.1038/nbt.2689
- Dang, T. T. T., Franke, J., Carqueijeiro, I. S. T., Langley, C., Courdavault, V., and O'Connor, S. E. (2018). Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat. Chem. Biol.* 14, 760–763. doi: 10.1038/s41589-018-0078-4
- Dang, T. T. T., Franke, J., Tatsis, E., and O'Connor, S. E. (2017). Dual catalytic activity of a cytochrome p450 controls bifurcation at a metabolic branch point of alkaloid biosynthesis in *Rauwolfia serpentina*. *Angew. Chem. Int. Ed. Engl.* 56, 9440–9444. doi: 10.1002/anie.201705010
- Decoene, T., Peters, G., De Maeseneire, S. L., and De Mey, M. (2018). Toward predictable 5'UTRs in *Saccharomyces cerevisiae*: development of a yUTR calculator. *ACS Synth. Biol.* 7, 622–634. doi: 10.1021/acssynbio.7b00366
- DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. J., and Dueber, J. E. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* 11, 465–471. doi: 10.1038/nchembio.1816
- Desagné-Penix, I., Farrow, S. C., Cram, D., Nowak, J., and Facchini, P. J. (2012). Integration of deep transcript and targeted metabolite profiles for eight cultivars of opium poppy. *Plant Mol. Biol.* 79, 295–313. doi: 10.1007/s11103-012-9913-2
- Dugé de Bernonville, T., Papon, N., Clastre, M., O'Connor, S. E., and Courdavault, V. (2020). Identifying missing biosynthesis enzymes of plant natural products. *Trends Pharmacol. Sci.* 41, 142–146. doi: 10.1016/j.tips.2019.12.006
- Dunstan, M. S., Robinson, C. J., Jervis, A. J., Yan, C., Carbonell, P., Hollywood, K. A., et al. (2020). Engineering *Escherichia coli* towards de novo production of gatekeeper (2S)-flavanones: naringenin, pinocembrin, eriodictyol and homoeriodictyol. *Synth. Biol.* 5:ysaa012. doi: 10.1093/synbio/ysaa012
- Ekas, H., Deaner, M., and Alper, H. S. (2019). Recent advancements in fungal-derived fuel and chemical production and commercialization. *Curr. Opin. Biotechnol.* 57, 1–9. doi: 10.1016/j.copbio.2018.08.014
- Facchini, P. J., and De Luca, V. (2008). Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J.* 54, 763–784. doi: 10.1111/j.1365-313X.2008.03438.x
- Farrow, S. C., Hagel, J. M., Beaudoin, G. A. W. W., Burns, D. C., and Facchini, P. J. (2015). Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* 11, 728–732. doi: 10.1038/nchembio.1879
- Fossati, E., Ekins, A., Narcross, L., Zhu, Y., Falgoutret, J. P., Beaudoin, G. A. W., et al. (2014). Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydroanguinarine in *Saccharomyces cerevisiae*. *Nat. Commun.* 5:3283. doi: 10.1038/ncomms4283
- Fridman, E., and Pichersky, E. (2005). Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr. Opin. Plant Biol.* 8, 242–248. doi: 10.1016/j.pbi.2005.03.004
- Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., Smolke, C. D., Interrante, M. F., et al. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373
- Gao, S., Zhou, H., Zhou, J., and Chen, J. (2020). Promoter-library-based pathway optimization for efficient (2S)-naringenin production from p-coumaric acid in *Saccharomyces cerevisiae*. *J. Agric. Food Chem.* 68, 6884–6891. doi: 10.1021/acs.jafc.0c01130
- Goh, H. H. (2018). Integrative multi-omics through bioinformatics. *Adv. Exp. Med. Biol.* 1102, 69–80. doi: 10.1007/978-3-319-98758-3_5
- Goh, H. H., Ng, C. L., and Loke, K. K. (2018). Functional genomics. *Adv. Exp. Med. Biol.* 1102, 11–30. doi: 10.1007/978-3-319-98758-3_2
- Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biol.* 20:121. doi: 10.1186/s13059-019-1730-3
- Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., et al. (2016). The biosynthetic pathway of the nonsugar, high-intensity sweetener mogrosin V from *Siraitia grosvenorii*. *Proc. Natl. Acad. Sci.* 113, E7619–E7628. doi: 10.1073/pnas.1604828113
- Jamil, I. N., Remali, J., Azizan, K. A., Nor Muhammad, N. A., Arita, M., Goh, H.-H., et al. (2020). Systematic multi-omics integration (MOI) approach in plant systems biology. *Front. Plant Sci.* 11:944. doi: 10.3389/fpls.2020.00944
- Jervis, A. J., Carbonell, P., Vinaixa, M., Dunstan, M. S., Hollywood, K. A., Robinson, C. J., et al. (2019). Machine learning of designed translational control allows predictive pathway optimization in *Escherichia coli*. *ACS Synth. Biol.* 8, 127–136. doi: 10.1021/acssynbio.8b00398
- Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2020). Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.* 64, 1–9. doi: 10.1016/j.copbio.2019.08.010
- Kirby, J., Dietzel, K. L., Wichmann, G., Chan, R., Antipov, E., Moss, N., et al. (2016). Engineering a functional 1-deoxy-D-xylulose 5-phosphate (DXP) pathway in *Saccharomyces cerevisiae*. *Metab. Eng.* 38, 494–503. doi: 10.1016/j.ymben.2016.10.017
- Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. doi: 10.1126/science.1069492
- Ku Bahaudin, K. N. A., Ramzi, A. B., Baharum, S. N., Sabri, S., Thean Chor, A. L., Sabri, S., et al. (2018). Current progress in production of flavonoids using systems and synthetic biology platforms. *Sains Malays.* 47, 3077–3084. doi: 10.17576/jism-2018-4712-18
- Lee, S. Y., and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061–1072. doi: 10.1038/nbt.3365
- Li, Y., and Smolke, C. D. (2016). Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* 7:12137. doi: 10.1038/ncomms12137
- Lin, G. M., Warden-Rothman, R., and Voigt, C. A. (2019). Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Curr. Opin. Syst. Biol.* 14, 82–107. doi: 10.1016/j.coisb.2019.04.004
- Lister, R., Gregory, B. D., and Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.* 12, 107–118. doi: 10.1016/j.pbi.2008.11.004
- Liu, X., Cheng, J., Zhang, G., Ding, W., Duan, L., Yang, J., et al. (2018). Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. *Nat. Commun.* 9:448. doi: 10.1038/s41467-018-02883-z
- Loke, K.-K., Rahnamaie-Tajadod, R., Yeoh, C.-C., Goh, H.-H., Mohamed-Hussein, Z. A., Zainal, Z., et al. (2017). Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. *PeerJ* 5:e2938. doi: 10.7717/peerj.2938
- 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
- Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 Plants (1KP) project. *Gigascience*. 3:2047-217X-3-17. doi: 10.1186/2047-217X-3-17
- Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G., and Wang, Y. (2013). Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. *PLoS One* 8:e60288. doi: 10.1371/journal.pone.0060288
- Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi: 10.1021/acs.jnatprod.5b01055

- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285
- Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004
- O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019
- 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
- Ping, Y., Li, X., Xu, B., Wei, W., Kai, G., et al. (2019a). Building microbial hosts for heterologous production of N-methylpyrrolinium. *ACS Synth. Biol.* 8, 257–263. doi: 10.1021/acssynbio.8b00483
- Ping, Y., Li, X., You, W., Li, G., Yang, M., Wei, W., et al. (2019b). *De novo* production of the plant-derived tropine and pseudotropine in yeast. *ACS Synth. Biol.* 8, 1257–1262. doi: 10.1021/acssynbio.9b00152
- Presnell, K. V., and Alper, H. S. (2019). Systems metabolic engineering meets machine learning: a new era for data-driven metabolic engineering. *Biotechnol. J.* 14:e1800416. doi: 10.1002/biot.201800416
- Pyne, M. E., Narcross, L., and Martin, V. J. J. (2019). Engineering plant secondary metabolism in microbial systems. *Plant Physiol.* 179, 844–861. doi: 10.1104/pp.18.01291
- 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
- Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* 90, 764–787. doi: 10.1111/tpj.13485
- Ramzi, A. B. (2018). Metabolic engineering and synthetic biology. *Adv. Exp. Med. Biol.* 1102, 81–95. doi: 10.1007/978-3-319-98758-3_6
- Ramzi, A. B., Ku Bahaudin, K. N. A., Baharum, S. N., Che Me, M. L., Goh, H.-H., Hassan, M., et al. (2018). Rapid assembly of yeast expression cassettes for phenylpropanoid biosynthesis in *Saccharomyces cerevisiae*. *Sains Malays.* 47, 2969–2974. doi: 10.17576/jsm-2018-4712-05
- Redding-Johanson, A. M., Bath, T. S., Chan, R., Krupa, R., Szmidi, H. L., Adams, P. D., et al. (2011). Targeted proteomics for metabolic pathway optimization: application to terpene production. *Metab. Eng.* 13, 194–203. doi: 10.1016/j.ymben.2010.12.005
- 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
- Rodriguez, A., Chen, Y., Khoomrung, S., Özdemir, E., Borodina, I., and Nielsen, J. (2017). Comparison of the metabolic response to over-production of p-coumaric acid in two yeast strains. *Metab. Eng.* 44, 265–272. doi: 10.1016/j.ymben.2017.10.013
- Rusdi, N. A., Goh, H.-H., Sabri, S., Ramzi, A. B., Mohd Noor, N., and Baharum, S. N. (2018). Functional characterisation of new sesquiterpene synthase from the Malaysian herbal plant, *Polygonum minus*. *Molecules* 23, 1370. doi: 10.3390/molecules23061370
- Scossa, F., Benina, M., Alseekh, S., Zhang, Y., and Fernie, A. R. (2018). The integration of metabolomics and next-generation sequencing data to elucidate the pathways of natural product metabolism in medicinal plants. *Planta Med.* 84, 855–873. doi: 10.1055/a-0630-1899
- Seaver, S. M. D., Henry, C. S., and Hanson, A. D. (2012). Frontiers in metabolic reconstruction and modeling of plant genomes. *J. Exp. Bot.* 63, 2247–2258. doi: 10.1093/jxb/err371
- Sheth, B. P., and Thaker, V. S. (2014). Plant systems biology: insights, advances and challenges. *Planta* 240, 33–54. doi: 10.1007/s00425-014-2059-5
- Srinivasan, P., and Smolke, C. D. (2019). Engineering a microbial biosynthesis platform for de novo production of tropane alkaloids. *Nat. Commun.* 10:3634. doi: 10.1038/s41467-019-11588-w
- St. John, P. C., and Bomble, Y. J. (2019). Approaches to computational strain design in the multiomics era. *Front. Microbiol.* 10:597. doi: 10.3389/fmicb.2019.00597
- Tan, C. S., Hassan, M., Mohamed Hussein, Z. A., Ismail, I., Ho, K. L., Ng, C. L., et al. (2018). Structural and kinetic studies of a novel nerol dehydrogenase from *Persicaria minor*, a nerol-specific enzyme for citral biosynthesis. *Plant Physiol. Biochem.* 123, 359–368. doi: 10.1016/j.plaphy.2017.12.033
- Tang, Q., Ma, X., Mo, C., Wilson, I. W., Song, C., Zhao, H., et al. (2011). An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics* 12:343. doi: 10.1186/1471-2164-12-343
- van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-r102
- Volk, M. J., Lourentzou, I., Mishra, S., Vo, L. T., Zhai, C., and Zhao, H. (2020). Biosystems design by machine learning. *ACS Synth. Biol.* 9, 1514–1533. doi: 10.1021/acssynbio.0c00129
- Wu, Z., Jennifer Kan, S. B., Lewis, R. D., Wittmann, B. J., and Arnold, F. H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.* 116, 8852–8858. doi: 10.1073/pnas.1901979116
- Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J. S., Chakrabarty, R., et al. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* 166, 122–134. doi: 10.1016/j.jbiotec.2013.04.004
- 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
- Zhou, Y., Li, G., Dong, J., Xing, X. H., Dai, J., et al. (2018). MiYA, an efficient machine-learning workflow in conjunction with the YeastFab assembly strategy for combinatorial optimization of heterologous metabolic pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 47, 294–302. doi: 10.1016/j.ymben.2018.03.020

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.

Copyright © 2020 Ramzi, Baharum, Bunawan and Scrutton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.