Metabolomics, standards, and metabolic modeling for synthetic biology in plants

Camilla Beate Hill[1]*, Tobias Czauderna[2], Matthias Klapperstück[2], Ute Roessner[1] and Falk Schreiber[2,3]

[1] School of BioSciences, University of Melbourne, Parkville, VIC, Australia, [2] Faculty of Information Technology, Monash University, Clayton, VIC, Australia, [3] Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany

Life on earth depends on dynamic chemical transformations that enable cellular functions, including electron transfer reactions, as well as synthesis and degradation of biomolecules. Biochemical reactions are coordinated in metabolic pathways that interact in a complex way to allow adequate regulation. Biotechnology, food, biofuel, agricultural, and pharmaceutical industries are highly interested in metabolic engineering as an enabling technology of synthetic biology to exploit cells for the controlled production of metabolites of interest. These approaches have only recently been extended to plants due to their greater metabolic complexity (such as primary and secondary metabolism) and highly compartmentalized cellular structures and functions (including plant-specific organelles) compared with bacteria and other microorganisms. Technological advances in analytical instrumentation in combination with advances in data analysis and modeling have opened up new approaches to engineer plant metabolic pathways and allow the impact of modifications to be predicted more accurately. In this article, we review challenges in the integration and analysis of large-scale metabolic data, present an overview of current bioinformatics methods for the modeling and visualization of metabolic networks, and discuss approaches for interfacing bioinformatics approaches with metabolic models of cellular processes and flux distributions in order to predict phenotypes derived from specific genetic modifications or subjected to different environmental conditions.

Keywords: metabolic engineering, synthetic biology, metabolic modeling, systems biology, metabolomics

INTRODUCTION

One of the greatest challenges for scientists is to understand the genetics, physiology, and biochemistry of plants and the interaction of genes with the environment in order to provide strategies to manipulate these processes to improve plant growth and performance and to prevent diseases. There have been immense advances in the past decades allowing scientists to sequence genomes of organisms addressing many questions, such as how the genome determines a plant's response to any environmental stimuli. Simultaneously, the development of analytical technologies allows us to take a comprehensive and unbiased glance at the gene products, such as gene transcripts (mRNA), proteins, lipids, and metabolites. The post-genomics era was born with the establishment of transcriptomics, proteomics, lipidomics, and metabolomics, each

with their associated computational advances providing the path for data analysis, visualization, and integration to establish the relationships between the genome and gene products under certain conditions.

Metabolites are synthesized by enzyme-catalyzed reactions in any living cell. They are important for the maintenance and survival of cells, most importantly for energy storage and provision, and they also contribute to building and maintaining the cell's structural components. Metabolites and their functionalities are indispensable in the interaction of a cell with the environment, and it has been argued that the metabolome of any biological system represents the final "read-out" of the expression of many genes in that system in a particular situation, and reflecting gene × environment relationships (Hill et al., 2014).

In comparison to transcripts or proteins, metabolites have a vast range of different chemical structures with an astonishing array of different functional groups that lead to differences in their physical and chemical properties, such as solubility, reactivity, stability, and polarity (Trethewey, 2004). This sheer diversity presents challenges to assay these compounds in a multiparallel fashion. Firstly, a number of different solvent extraction procedures need to be utilized to extract metabolites efficiently from any given plant tissue. In addition, no single analytical approach is capable of detecting and quantifying such chemical diversity; therefore, a range of different approaches (more detail below) has to be employed to analyze as many metabolites as possible. Today, metabolomics is considered as the science combining modern and sophisticated analytical instrumentation for metabolite detection and quantification with appropriate computational and statistical approaches to extract, mine, and interpret metabolomics data.

Metabolomics is now also becoming an important tool for biotechnological and metabolic engineering approaches, which aim to manipulate biochemical pathways to enhance the accumulation of compounds of interest (Dromms and Styczynski, 2012). Since metabolomics can provide a more complete picture of the biological system studied, it has been argued that it can be applied to identify metabolite markers that indicate a particular phenotype [e.g., level of target compound(s)] to allow the assessment of the successes of engineering steps to provide further guidance for future engineering strategies (Beckles and Roessner, 2011). Historically, metabolic engineers have used the analysis of the levels of the target compound(s) and potentially a few closely related metabolites to define metabolic engineering strategies. However, the potential metabolomics offers, which measures hundreds of metabolites rather than just a few, has only been rarely explored in metabolic engineering approaches, particularly in plants (Rios-Estepa and Lange, 2007; Fernie and Morgan, 2013). The unbiased and broad approach of metabolomics helps to assess how plants maintain energy, carbon, and nutrient resources and provides information how these resources may potentially be redirected into the synthesis of the desired metabolites, therefore allowing smart engineering strategies to be developed.

In the following, we review current metabolomics technologies, information resources for metabolomics, as well as computational analysis and modeling approaches with a focus on plant related research.

TOOLS AND TECHNOLOGIES TO STUDY PLANT METABOLISM

Analytical Technologies

The plant metabolome, compared to the metabolome of other organisms, is represented by a particularly vast variety of chemical structures with an enormous diversity of chemical and physical properties (Villas-Boas et al., 2007). In the past decade, researchers have developed and validated a number of complementary analytical approaches to extract, separate, detect, and quantify this diversity. Different solvent extraction procedures may need to be employed to cover the range of polarity of metabolites (Dias et al., 2012). However, most routine metabolite extractions are based on a methanol/water/chloroform biphasic extraction, which captures a large complement of the plant metabolome.

Once metabolites are extracted, the complex mixtures need to be separated allowing individual detection and quantification of compounds. The polarity of metabolites also influences the choice of the separation approach. In liquid chromatography (LC)-based separations, researchers now commonly use two types of separation chemistries, such as C18 reverse phase, which separates the more hydrophobic complement of a metabolite extract, such as many secondary metabolites and lipids, and hydrophilic interaction chromatography, which is better suited for the polar metabolites (Callahan et al., 2009; Hill and Roessner, 2015). In addition, many metabolites are either positively or negatively charged molecules, therefore molecules need to be ionized using both positive and negative ionization mode (Beckles and Roessner, 2011; Hill et al., 2014). There are a number of different ionization techniques available, such as electrospray ionization or atmospheric pressure ionization, again each better suited for a particular subclass of metabolites. However, the most common technique used in LC-MS-based metabolomics is electrospray ionization, which allows reliable ionization of thousands of compounds (Hill et al., 2014).

An alternative separation technique is gas chromatography (GC), which is known for its superior separation power and reproducibility (Dias et al., 2015). The Metabolomics Standards Initiative (MSI; Fiehn et al., 2007) has developed minimal reporting standards for metabolomics data, and strategies to further enhance reproducibility, experimental and data standardization are continuously developed (Allwood et al., 2009). However, GC requires compounds to be volatile to be amenable for analysis. Most metabolites are not volatile and therefore require chemical derivatization to make them volatile. This limits the utility of GC-based separation in metabolomics applications; however, GC coupled to MS still remains the "work horse" in metabolomics due to its reproducibility and its ease of use (Hill and Roessner, 2013).

Mass spectrometry (MS) is the most commonly used detector in metabolomics approaches. The power of MS is that it can distinguish the size of the ionized molecule (by determining the mass to charge ratio of each detected ion) and allows the determination of the number of individual ions detected. The mass to charge ratio determined by MS is used for compound identification and the number of ions detected can be related back to initial concentration of the molecule in the metabolite extract. Another

advantage of MS is that it can fragment an ion once (MS/MS) or multiple times (MS^n), a feature used for structural elucidation of unknown compounds. MS also has excellent capabilities to detect and distinguish isotopic patterns of each ion under analysis. This is particularly important for stable isotope labeling experiments used for metabolic flux analysis. MS analysis is able to determine enrichments in stable isotope labels in all the analyzed compounds and therefore allows determining metabolic fluxes through particular pathways. Most commonly ^{13}C labeled metabolic precursors are used, and the distribution and enrichment of the label across pathways of interest are quantified (O'Grady et al., 2012; Chokkathukalam et al., 2014). ^{13}C-labeling patterns can also be detected by nuclear magnetic resonance spectroscopy (NMR), which also represents an important platform for untargeted, nondestructive metabolite profiling (Eisenreich and Bacher, 2007).

Metabolomics Data Analysis and Visualization

Metabolic data give us a snapshot of the current state of an organism (Hill and Roessner, 2013). It represents the outcome of a preceding gene expression profile, which influences the activity of pathways, transport processes, as well as production and consumption of metabolites. The resulting metabolic profile can be used to classify organisms, for example, by genotype or treatment. Furthermore, these profiles enable comparative analysis between selected treatments or genotypes (Hill et al., 2013a,b, 2015) and to obtain information about metabolites with most or least changes as a result of changes in the gene expression profile.

There are two general methods to analyze metabolic data, which can also be combined. The first, analytical method uses commonly known statistics and clustering algorithms, the second method implies the use of networks to visualize spatial and temporal properties of the data. Performing data statistics in the scope of networks for visualization purposes represents the combination of both methods.

Scientists can choose from a broad repository of statistical methods with respect to the objective at hand. Methods such as frequency distribution, analysis of variance (ANOVA), min−max, and Pearson's and Spearman-rank correlation are examples for univariate data analysis. Principal component analysis (PCA), partial least squares regression (PLS), and multivariate analysis of variance (MANOVA) are commonly used for multivariate data analysis. Self-organizing maps (SOM), support vector machines (SVM), and k-means are popular methods for cluster analysis. Generated results can be visualized using different kinds of diagrams such as plots, histograms, cluster diagrams, and heat maps.

To perform the above-mentioned analytical methods, several software tools can be used. There is a distinction between low-level tools to provide the actual set of algorithms, which are in turn used by high-level tools to provide user-friendly application of those algorithms. **Table 1** shows a selection of frequently used software tools.

Graphical representations of metabolic pathways and networks have been used for long time to represent knowledge about metabolic processes, and with the availability of pathways in databases,

TABLE 1 | Software tools for metabolomics data analysis.

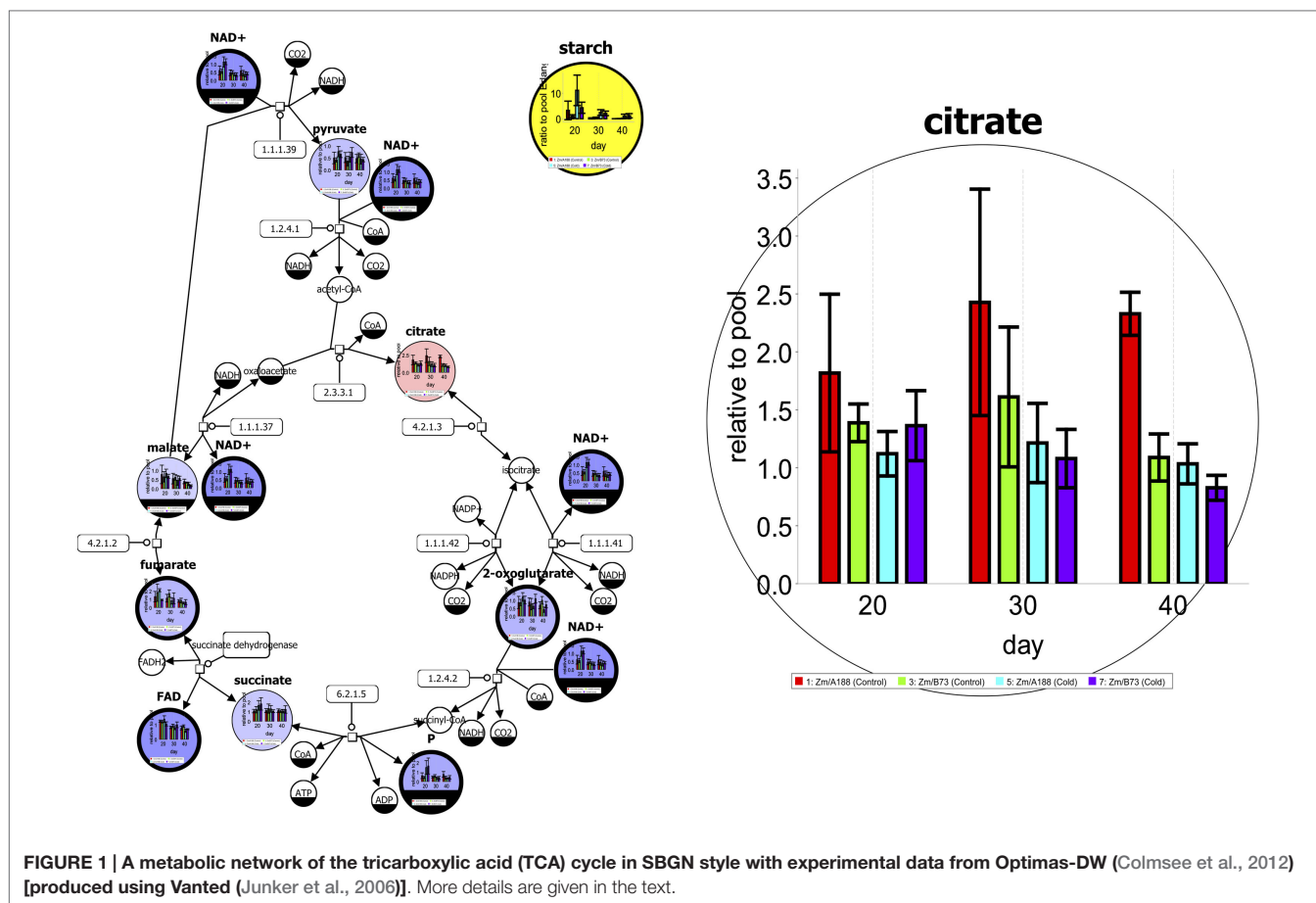| Name	Reference	URL
HIGH LEVEL		
MetaboAnalyst	Xia et al. (2015)	http://www.metaboanalyst.ca/
XCMS	Gowda et al. (2014)	https://xcmsonline.scripps.edu/
MetATT (Metabolomics tool for analyzing two-factor and time-series data)	Xia et al. (2011)	http://metatt.metabolomics.ca/MetATT/
metaP-Server	Kastenmüller et al. (2011)	http://metap.helmholtz-muenchen.de/metap2/
LOW LEVEL		
Matlab (PLS toolbox, msalign, etc.)	MathWorks (2012)	http://ch.mathworks.com/products/matlab/
R-packages (AmsRPM, apLCMS, metabolomics, muma, etc.)	R Development Core Team (2008), Kirchner et al. (2007), Yu et al. (2009), Gaude et al. (2013)	http://www.r-project.org/

TABLE 2 | Network visualization software tools that support metabolic data.

Name	Reference	URL
MetScape Plugin for Cytoscape	Karnovsky et al. (2012)	http://metscape.ncibi.org/
MetaMapp (MS data in context of metabolic networks using Cytoscape)	Barupal et al. (2012)	http://metamapp.fiehnlab.ucdavis.edu
MAVEN (LC-MS)	Clasquin et al. (2012)	http://maven.princeton.edu
VANTED	Rohn et al. (2012)	http://www.vanted.org
Pathomx	Fitzpatrick et al. (2014)	http://pathomx.org/

several tools have been developed to visualize metabolic data in the context of networks. These tools mostly support tasks such as data mapping and network analysis but often also try to help with layout and exploration of data. The latter touches the field of visual analytics for metabolic information (Kerren and Schreiber, 2012, 2014). **Table 2** shows a selection of tools supporting visualization of metabolic data as diagrams or heat maps in the context of biological networks.

As a result of applying common procedures of extracting and measuring metabolic abundance and concentration, usually metabolomic data possesses no spatial information. With the advent of modern methods such as imaging MS (Kaspar et al., 2011; Miura et al., 2012), additional spatial information can be gathered and displayed using 2D (Rohn et al., 2011) or 3D immersive techniques (Sommer et al., 2011).

Figure 1 shows an exemplary metabolic network in SBGN style (see section Standards for Systems and Synthetic Biology). Time series data for different genotypes and treatments have been mapped on the tricarboxylic acid (TCA) cycle pathway. The left figure not only shows the data using box plots but also a Pearson's correlation analysis for starch production (yellow). Red- and blue-colored elements have a positive or negative correlation, respectively. For further investigation, single elements can be

FIGURE 1 | A metabolic network of the tricarboxylic acid (TCA) cycle in SBGN style with experimental data from Optimas-DW (Colmsee et al., 2012**) [produced using Vanted (**Junker et al., 2006**)]**. More details are given in the text.

enlarged to support additional exploration of data as shown on the right side.

Databases and Repositories for Metabolites and Metabolism

Given the tools introduced before, metabolite and pathway databases and repositories are valuable resources that can be used to manage, explore, and export knowledge about metabolites and reactions in meaningful ways. They thereby deliver an encyclopedia on metabolic information as well as a base for integration of complex data into metabolic pathways in the context of graphical pathway representations (e.g., data about compound levels, reaction flow, enzyme activity, and gene expression, see Metabolomics Data Analysis and Visualization). Often these databases and repositories also provide or allow building metabolic models which can be analyzed and simulated using mathematical modeling techniques, see Section "Methods for Metabolic Modeling in Plants." A typical example of information provided by databases is shown in **Figure 2**.

A range of pathway databases and repositories are currently available; an overview is available from the Pathguide resource (Bader et al., 2006). In **Table 3,** we provide a summary of important databases and repositories especially for plant research. It should be noted that in addition to general metabolic pathway databases such as Reactome, KEGG (Kyoto Encyclopedia of

Genes and Genomes), and PANTHER Pathway, there are also general plant metabolic pathway databases (see **Table 3**) and several species-specific plant metabolic pathway databases available, for example, for *Arabidopsis* AraCyc (Mueller et al., 2003) and MetNetDB (Yang et al., 2005).

Examples for multispecies plant metabolic databases are:

- *PlantCyc*, which contains curated information on pathways, metabolites, reactions, genes, and enzymes,
- *Arabidopsis Reactome*, which has been extended from the initial representation of *Arabidopsis* pathways to further plant species and contains curated information about pathways, experimental evidence, literature citations, and further pathways imported from KEGG and AraCyc, and
- *MetaCrop*, which focuses on crop plant species with agronomical importance and contains curated information about pathways, metabolites, reactions, enzymes (including kinetic information), location, and developmental stage.

CURRENT STATUS OF ENGINEERING SYNTHETIC METABOLIC NETWORKS

For a long time, the use of biological organisms for the production of chemicals was limited by the repertoire of biosynthetic pathways naturally present in these organisms. Microbial
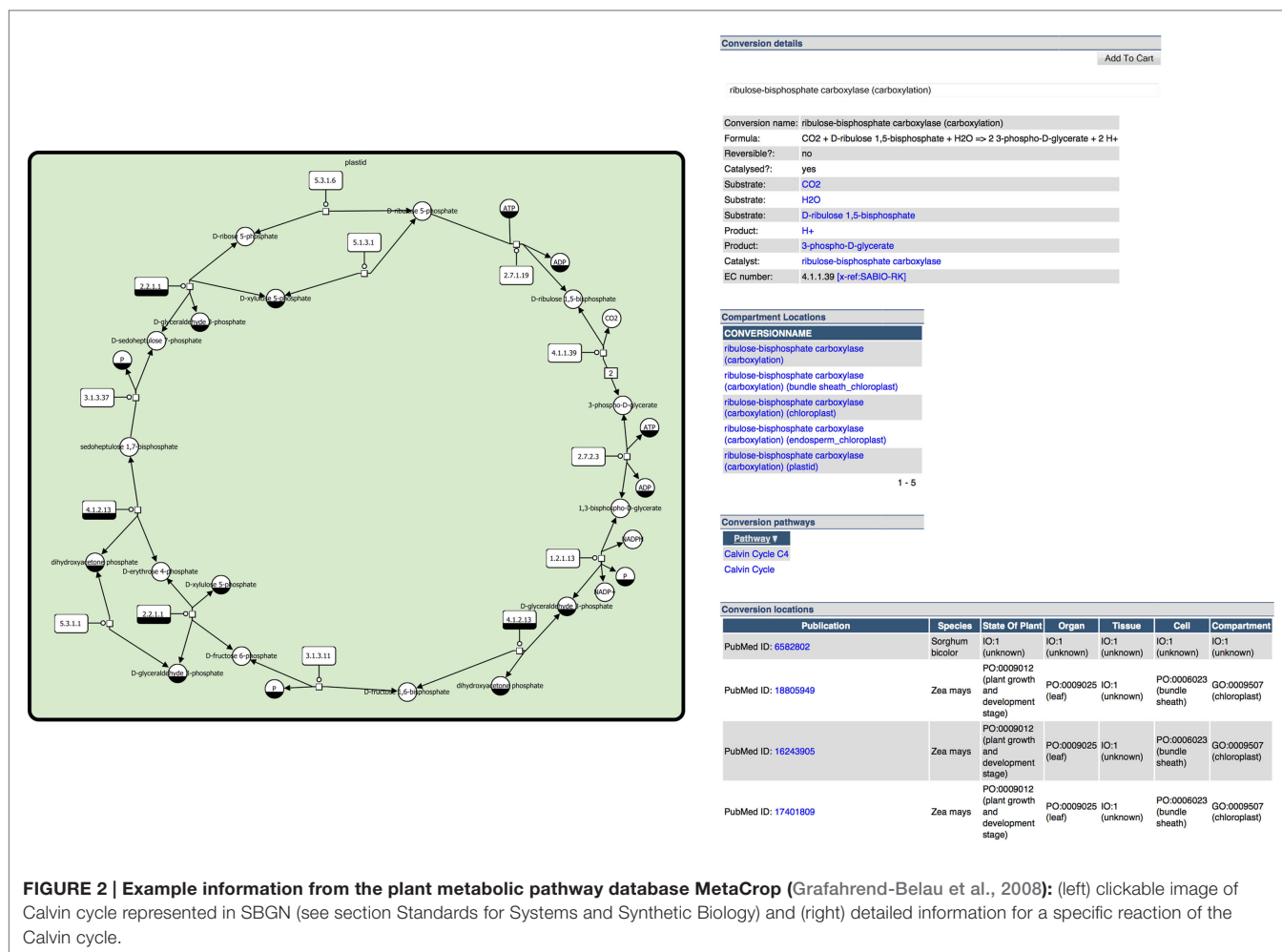
FIGURE 2 | Example information from the plant metabolic pathway database MetaCrop (Grafahrend-Belau et al., 2008): (left) clickable image of Calvin cycle represented in SBGN (see section Standards for Systems and Synthetic Biology) and (right) detailed information for a specific reaction of the Calvin cycle.

low-molecular-weight metabolic end products have long been used as commodity chemicals for medical applications: for example, antibiotics, such as penicillin, or immunosuppressive drugs, such as cyclosporine, have been discovered in 1929 and 1972, respectively (Drews, 2000). Microbes have also been used to produce chemical compounds via biotransformation, a process in which the compound of interest is produced by the microorganism through enzymatic conversion of an external substrate added to the microbial culture medium, used, for example, for the production of acrylamide by nitrile-assimilating bacteria (Asano et al., 1982).

With the advent of recombinant DNA technology in the early 1990s, it was possible to engineer specific genes in biological organisms, which has significantly reduced the time required for mutagenesis and selection of desirable traits. Genetic engineering made it possible to use heterologous hosts for the production of chemical compounds that are not naturally present in the organism. The emergence of the clustered, regularly interspaced, short palindromic repeat (CRISPR) and related technologies that use targeted genome editing via engineered nucleases are the latest developments to introduce alterations of genome sequences and gene expression, which can be ultimately used to also introduce modifications to existing metabolic pathways and to transfer

novel traits in agricultural crops (Shan et al., 2013; Sander and Joung, 2014).

The development of new sophisticated genomic sequencing and other enabling technologies for synthetic biology facilitated the production of naturally present chemicals at levels that made extraction economically feasible, and the field of metabolic engineering began to emerge. Metabolic engineering is defined as the targeted modification of metabolic pathways of biological organisms for metabolite overproduction or the improvement of cellular properties (Lessard, 1996). Since the last decade, significant progress was made to engineer the metabolism of plants to produce specific lipids, secondary metabolites, derivatives of complex natural products, and even vaccines (Mortimer et al., 2012). Many recent studies show that it is often not sufficient to modify existing metabolic pathways, but rather it is required to design metabolic pathways *de novo* from other plants or bacteria. Editing or redesigning existing plant metabolic networks is a challenging task that will benefit from advances in targeted genome modification, tissue-, cell-, and organelle-specific gene expression, the controlled expression of multigene pathways, and improvements in analytical technologies (as described in Analytical Technologies) as well as computational analysis and modeling methods (as described in Metabolomics Data Analysis

TABLE 3 | A summary of important databases and repositories for plant research.

Database	Exchange (download) formats	Reference	URL
Metabolite databases (provide information about metabolites/compounds such as names, chemical structures, molecular weight, occurrence in pathways, EC number, and mass spectrum references)			
ChEBI	XML, SDF, Tab-delimited	De Matos et al. (2009)	http://www.ebi.ac.uk/chebi/
GMD	MS formats	Kopka et al. (2005)	http://gmd.mpimp-golm.mpg.de/
KEGG compound	Jmol, MDL/MOL, KCF, KegDraw	Kanehisa et al. (2014)	http://www.genome.jp/kegg/compound/
PubChem	XML, SDF, SMILES	Bolton et al. (2008)	http://www.ncbi.nlm.nih.gov/pccompound
Reaction databases (provide information about reactions and enzymes such as names, reaction diagrams, reaction mechanisms, enzymatic parameters, occurrence in pathways, and links to encoding genes)			
BRENDA	SBML, Fasta, CVS	Scheer et al. (2010)	http://www.brenda-enzymes.org/
ExPASy-enzyme	–	Gasteiger et al. (2003)	http://enzyme.expasy.org/
KEGG enzyme/ KEGG reaction	–	Kanehisa et al. (2014)	http://www.genome.jp/kegg/reaction/
Rhea	BioPAX, Tab-delimited	Morgat et al. (2015)	http://www.ebi.ac.uk/rhea/
Sabio-RK	SBML	Rojas et al. (2007)	http://sabio.villa-bosch.de/
Pathway databases (provide information about plant-specific metabolic pathways such as names, involved reactions, metabolites and enzymes, and pathway structure)			
KEGG pathway	KGML, BioPAX	Kanehisa et al. (2014)	http://www.genome.jp/kegg/pathway.html
MetaCrop	SBML, SBGN-ML	Schreiber et al. (2012)	http://metacrop.ipk-gatersleben.de/
MetaCyc	SBML, BioPAX	Krieger et al. (2004), Caspi et al. (2006)	http://metacyc.org/
PANTHER pathway	SBML, BioPAX, SBGN-ML	Mi et al. (2013)	http://www.pantherdb.org/pathway/
PlantCyc	SBML, BioPAX		http://www.plantcyc.org/
Reactome	SBML, BioPAX	Croft et al. (2014)	http://www.reactome.org/

and Visualization, Databases and Repositories for Metabolites and Metabolism, and Computational Approaches for Metabolic Engineering).

The ultimate goal of synthetic biology is the efficient design of biological systems (Heinemann and Panke, 2006). In this section, we will discuss the current status of engineering synthetic metabolic networks using recent examples for synthetic biology endeavors in plants: engineering of synthetic metabolic networks

of plant lipids to provide an alternative and sustainable source of nutrients (see Metabolic Engineering of Plant Lipids to Provide an Alternative and Sustainable Source of Nutrients) and for the production of fuels from renewable resources (see Metabolic Engineering of Plant Lipids for the Production of Fuels from Renewable Resources), and plant secondary metabolites including alkaloids and lignins (see Metabolic Engineering of Plant Secondary Metabolites).

Metabolic Engineering of Plant Lipids to Provide an Alternative and Sustainable Source of Nutrients

Plant oils are a major component of human diets, comprising as much as 25% of average caloric intake (Broun et al., 1999). However, certain fatty acids such as omega-3 long-chain polyunsaturated fatty acids (ω3 LC-PUFA) are present predominantly in fish and have important functions for human health, as deficiencies in these fatty acids can increase the risk or severity of cardiovascular and inflammatory diseases (Abeywardena and Patten, 2011). Until recently, the chemical composition of plant oils was constrained by the repertoire of naturally present lipid biosynthetic pathways. Novel opportunities have emerged to tailor the composition of plant-derived lipids so that they are optimized with respect to food functionality and human dietary needs. For example, Petrie et al. (2010, 2012) have recently described metabolic engineering of ω3 LC-PUFA in plants: after inserting seven biosynthesis genes of the docosahexaenoic acid (DHA) biosynthesis pathway from microalgae into the genome of *Arabidopsis thaliana*, they were able to obtain ω3 LC-PUFA levels in seeds similar to that observed in bulk fish oil. If applied to oilseed crops such as *Brassica napus*, this technology could potentially form the basis of a plant-based sustainable source to complement the existing marine fish oil supply.

Metabolic Engineering of Plant Lipids for the Production of Fuels from Renewable Resources

Fossil fuels are the primary source of many industrial products, but reserves are decreasing rapidly and are non-renewable, and their widespread use has contributed to environmental problems arising from increased CO_2 levels in the atmosphere (Le Quéré et al., 2009). Currently, biologically derived fuels from plant oils represent one of the main strategies to provide renewable and sustainable source material that can potentially substitute fossil fuels in some industrial applications. Among the many proposed solutions, algal biofuels are seen as one of the most promising: algal biomass is less resistant to conversion into simple sugars than plant biomass due to lack of lignin, and there is no issue arising from the food versus feed dilemma as no farmland has to be diverted for the production of biofuels (Daroch et al., 2013). Over the last few years, progress was made in bioethanol production through fermentation from algal feedstock (Kim et al., 2012) as well as biodiesel production from algal oils (Singh and Dhar, 2011). Increasingly, research efforts are focusing to metabolically engineer lipid pathways to increase lipid accumulation without

compromising growth (Trentacoste et al., 2013). Although previously *A. thaliana* mutants of lipid catabolism were found to be linked with impaired growth (Graham, 2008), Trentacoste et al. (2013) demonstrated that disrupting lipid catabolism via the knockdown of a multifunctional lipase/phospholipase/acyltransferase in the microalgae *Thalassiosira pseudonana* led to an increased lipid accumulation without compromised algal growth. Further elucidation of lipid metabolism has the potential to lead to new strategies to engineer improved algal strains for their fuel molecules.

Metabolic Engineering of Plant Secondary Metabolites

Plant secondary metabolites, such as alkaloids, flavonoids, terpenes, and phenylpropanoids (Hill et al., 2014), are considered to be non-essential for normal growth and development but play important roles in plant defense against pathogens and other environmental stresses. Additionally, plant secondary metabolites are of great interest to pharmaceutical industries, as they often have beneficial medicinal effects on humans. For example, many plant alkaloids are currently in medical use, such as atropine derived from the nightshade *Atropa belladonna*, morphine from the opium poppy *Papaver somniferum*, and quinine from the *Cinchona* tree (Roberts and Wink, 1998). Recent progress has been made in the metabolic engineering of morphine, a medicinally important benzylisoquinoline alkaloid: Runguphan et al. (2012) reengineered a codeine *O*-demethylase mutant that selectively demethylates codeine instead of both codeine and thebaine, as is common in the wild-type morphinan biosynthesis pathway. The integration of this highly selective mutant enzyme into commercial poppy plants as part of a future metabolic engineering effort has the potential to increase yields of morphine and codeine.

The phenylpropanoid pathway is conserved in all terrestrial plants and is responsible for the biosynthesis of many compounds that are involved in plant cell wall structure and integrity, water transport, and plant defense. They are required for the biosynthesis of lignins, aromatic natural polymers in secondary cell walls derived from the oxidative polymerization of monolignols. Decreasing or altering lignin structure provides enhanced cell wall digestibility and can greatly increase the utilization of lignin itself or cell wall polysaccharides. Due to the importance of lignin in agriculture and industry, the genes participating in lignin biosynthesis have been identified and modified in many plant species including switchgrass (Fu et al., 2011), *A. thaliana* (Gallego-Giraldo et al., 2011), and sugarcane (Jung et al., 2012). In a recent study, Zhang et al. (2012) were able to manipulate lignification in *A. thaliana* without compromising plant growth by introducing an artificial enzyme that esterifies the para-hydroxyl of phenols. The modified 4-*O*-methyl lignin monomers deprive the products of participation in oxidative dehydrogenation, leading to a decreased level of available monolignols for lignin polymerization and thus to depressed lignin biosynthesis. Further metabolic engineering efforts are currently underway to integrate this artificial enzyme into poplar with the potential to manipulate lignin levels.

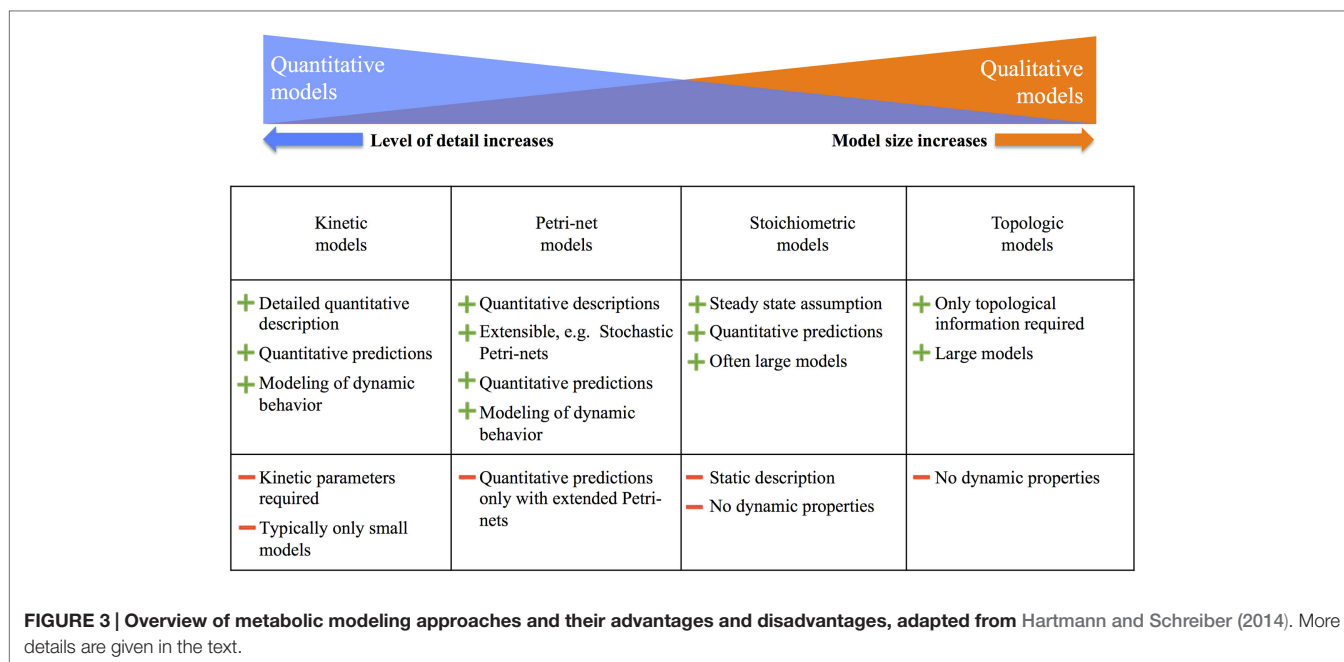# COMPUTATIONAL APPROACHES FOR METABOLIC ENGINEERING

A variety of different methods and approaches to collect experimental data can be used to quantify metabolites and other components of regulatory networks in plants, such as metabolomics (see Analytical Technologies). Computational modeling is an important tool for metabolic engineering as it facilitates the integration and analysis of experimental datasets to quantify metabolic fluxes and model metabolic networks. Section "Methods for Metabolic Modeling in Plants" presents a brief overview of modeling approaches. There are many tools and databases available for computational modeling and therefore a standardized exchange of models is highly relevant; Section "Standards for Systems and Synthetic Biology" provides an introduction to major standards in systems in synthetic biology.

Methods for Metabolic Modeling in Plants

Several approaches have been developed to qualitatively and quantitatively model and simulate metabolic systems *in silico*. This ranges from topological analysis of network models (which looks at the interconnections between metabolites) to stoichiometric models (where constraints can be applied to define the potential metabolic flux state space or which can be analyzed using Petri-nets) to detailed kinetic models (which model changes of metabolite concentration over time). A current review (Baghalian et al., 2014) discusses the different modeling approaches, modeling software, and metabolic models of several plants in detail. A particular challenge in plants compared to prokaryotic cells is the number of different compartments, which needs to be considered in metabolic models. In addition, plants have a greater complexity of metabolic pathways and especially a large number of special pathways for secondary metabolites.

For an overview, **Figure 3** summarizes the major modeling approaches and their advantages and disadvantages. The most detailed are kinetic models, which allow for a comprehensive quantitative description and prediction of metabolic fluxes. However, in plants, these models are common only in the size of 10–20 reactions. As we move further to the right in **Figure 3,** the model size increases, but the level of detailed descriptions and predictions decreases. As the other extreme, topological models allow covering the complete metabolism in plants, but predictions are restricted to qualitative information such as reachability of metabolites.

As detailed in (Baghalian et al., 2014), plant-specific computational models of metabolism can be used for different purposes such as predicting the behavior of the metabolism under different conditions, analyzing the effect of mutations, and investigating the effect of changes due to manipulation of the metabolic system, for example, via the introduction of new metabolic pathways. Computational models usually allow investigating effects much faster and cheaper than by running wet laboratory experiments (Rohwer, 2012). In addition, using a computational model can often generate a set of alternative strategies (Copeland et al., 2012). Finally, computational models allow integrating additional data such as transcriptomics and proteomics data sets, which together with bioinformatics approaches can support a better understanding of metabolic behavior in plants (Töpfer et al., 2012, 2013).

FIGURE 3 | Overview of metabolic modeling approaches and their advantages and disadvantages, adapted from Hartmann and Schreiber (2014). More details are given in the text.

Standards for Systems and Synthetic Biology

This section presents a short introduction to major standards in systems and synthetic biology related to software infrastructure (see also **Table 4**). Software infrastructure plays an important role in systems biology research (Kitano, 2002), in particular in supporting standardized exchange of information between different tools and databases. The major standards are Systems Biology Markup Language (SBML), CellML, Systems Biology Graphical Notation (SBGN), and Synthetic Biology Open Language (SBOL), Language (SBOL), see (Schreiber et al., 2015) for detailed specifications.

Systems Biology Markup Language

Systems Biology Markup Language is a machine-readable format for representation and exchange of computational models in systems biology. It can represent models of metabolism, signal transduction, and gene regulation. The main goals of SBML are (1) sharing and publication of models, (2) reusability of models, and (3) surviving of models beyond the lifetime of the software used to create them. The SBML web page currently lists more than 270 software applications that support SBML, and thousands of SBML-encoded models are available from public repositories such as BioModels including Path2Models (Büchel et al., 2013; Chelliah et al., 2013).

An SBML model consists of hierarchical lists of conceptual elements: (1) species (biological entities taking part in reactions), (2) compartments (physical containers for species), and (3) reactions (transformation, transport, or binding processes occurring over time). For the analysis and simulation of a model, more properties need to be defined such as stoichiometries, rate laws, local and global parameters, as well as units on quantities. A formal

TABLE 4 | Overview of major standards in systems and synthetic biology.

Name	Reference	URL
SBML	Hucka et al. (2003)	http://sbml.org/
CellML	Cuellar et al. (2003)	http://www.cellml.org/
SBGN	Le Novère et al. (2009)	http://sbgn.org/
SBOL	Galdzicki et al. (2014)	http://sbolstandard.org/

description of SBML can be found in the detailed specification (Hucka et al., 2015).

CellML

CellML is a machine-readable format for representation, publication, and sharing of mathematical models of cellular function. In comparison to SBML, the focus of CellML is on the representation of a variety of models such as models of biological pathways, electrophysiological models, and mechanical models. The CellML web page lists a couple of software tools which support the CellML format and the development of models. The CellML Model Repository (Lloyd et al., 2008) contains several hundred models including a subrepository providing SVPs (Standard Virtual Biological Parts) for the composition of synthetic biology models (Cooling et al., 2010).

A CellML model description consists of components and lists of connections between the components. A component contains at least one variable and mathematical equations describing its behavior. Connections are mappings of variables between components enabling information exchange between them. Components and connections can be imported from an existing model, CellML allows reusing of parts of other models. A detailed description of CellML can be found in the specification (Cuellar et al., 2006).

Systems Biology Graphical Notation

Systems Biology Graphical Notation is a standard for the graphical representation of processes and networks studied in systems biology. Three SBGN languages (Process Description, PD; Entity Relationship, ER; and Activity Flow, AF) allow for the representation of different aspects of biological systems at different levels of detail as SBGN maps, thus providing corresponding views on the underlying biological system. The PD language (Moodie et al., 2015) describes biological entities and processes between these entities, the ER language (Sorokin et al., 2015) focuses on interactions between biological entities, and the AF language (Mi et al., 2015) depicts information flow between biological activities. An example of an SBGN PD map is shown in **Figure 1**.

The standardization of graphical representations helps to exchange biological knowledge more efficiently and accurately between different research communities, industry, and other players in systems biology. Several databases already provide maps in SBGN, e.g., BioModels Database including Path2Models (Büchel et al., 2013; Chelliah et al., 2013), MetaCrop (Schreiber et al., 2012), PANTHER Pathway (Mi et al., 2013), Reactome (Croft et al., 2014), and RIMAS (Junker et al., 2010). The SBGN web page lists more than 20 software tools that support creating, editing, and viewing of SBGN maps, some of these tools allow to visualize SBML models in SBGN PD.

Synthetic Biology Open Language

Synthetic Biology Open Language is a data format for sharing and exchanging synthetic biology designs. It allows synthetic biologists to provide an unambiguous description of a design in a hierarchical and fully annotated form with the goal to improve designing, building, testing, and dissemination of synthetic biology designs. For the visualization of synthetic biology designs in SBOL, SBOL Visual (Synthetic Biology Open Language Visual) has been developed. It is a graphical notation allowing depiction of the structure of a design using glyphs to specify genetic parts, devices, modules, and systems.

The SBOL web page currently lists more than 20 software applications supporting SBOL and SBOL Visual. Some of these applications allow the generation of SBML models from synthetic biology designs in SBOL (Roehner et al., 2015) as well as the creation of synthetic biology designs in SBOL by automatically generating DNA sequences from annotated SBML and CellML models (Misirli et al., 2011). More detailed information

about SBOL and SBOL Visual can be found in the specifications (Quinn et al., 2013; Bartley et al., 2015).

CONCLUSION

The total number of metabolites in the plant kingdom is estimated to be between 100,000 and 200,000 and can be highly variable depending on the physiological and environmental conditions as well as the genetic background of the plant (Hill et al., 2013a,b). Such great metabolic diversity holds great promise for expanding our repertoire of known beneficial plant compounds, as many metabolic pathways and regulatory mechanisms are still awaiting discovery. Reaching significant benchmarks toward attaining these goals will be possible with better analytical tools. More accurate representations of metabolite identities and quantities will require analytical instruments and improved techniques for sample extraction and data analysis. The engineering of synthetic metabolic networks of plants will require further advances in targeted genome modification such as the application of the CRISPR/Cas system, as well as tissue-, cell-, and organelle-specific gene expression, and the controlled expression of multigene pathways. The development of methods for measuring metabolic flux directly, the quantification of metabolites in individual plant compartments, and the analysis of metabolites and activities between compartments *in vivo* will be very important next steps to further enhance the predictive capabilities of existing metabolic models. Continuous development of more user-friendly software, databases, languages, and computer models that incorporate and interpret complex information will be crucial to handle the acquired data and to aid interpretation in a biological context. We are just at the beginning of a new area of synthetic biology in plants based on metabolomics and metabolic modeling.

ACKNOWLEDGMENTS

REFERENCES

Abeywardena, M. Y., and Patten, G. S. (2011). Role of v3 longchain polyunsaturated fatty acids in reducing cardio-metabolic risk factors. *Endocr. Metab. Immune Disord. Drug Targets* 11, 232–246. doi:10.2174/187153011796429817

Allwood, J. W., Erban, A., de Koning, S., Dunn, W. B., Luedemann, A., Lommen, A., et al. (2009). Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics* 5, 479–496. doi:10.1007/s11306-009-0169-z

Asano, Y., Yasuda, T., Tani, Y., and Yamada, H. (1982). A new enzymatic method of acrylamide production. *Agric. Biol. Chem.* 46, 1183–1189. doi:10.1271/bbb1961.46.1183

Bader, G. D., Cary, M. P., and Sander, C. (2006). Pathguide: a pathway resource list. *Nucleic Acids Res.* 34, D504–D506. doi:10.1093/nar/gkt1076

Baghalian, K., Hajirezaei, M. R., and Schreiber, F. (2014). Plant metabolic modeling: achieving new insight into metabolism and metabolic engineering. *Plant Cell* 26, 3847–3866. doi:10.1105/tpc.114.130328

Bartley, B., Beal, J., Clancy, K., Misirli, G., Roehner, N., Oberortner, E., et al. (2015). Synthetic Biology Open Language (SBOL) Version 2.0.0 *J. Int. Bioinformatics* 12, 272. doi:10.2390/biecoll-jib-2015-272

Barupal, D. K., Haldiya, P. K., Wohlgemuth, G., Kind, T., Kothari, S. L., Pinkerton, K. E., et al. (2012). MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 13:99. doi:10.1186/1471-2105-13-99

Beckles, D., and Roessner, U. (2011). "Plant metabolomics – applications and opportunities for agricultural biotechnology," in *Plant Biotechnology and Agriculture: Prospects for the 21st Century*, eds Altmann A. and Hasegawa P. M. (London, UK: Elsevier), p. 67–80.

Bolton, E. E., Wang, Y., Thiessen, P. A., and Bryant, S. H. (2008). PubChem: integrated platform of small molecules and biological activities. *Ann. Rep. Comput. Chem.* 4, 217–241. doi:10.1016/s1574-1400(08)00012-1

Broun, P., Gettner, S., and Somerville, C. (1999). Genetic engineering of plant lipids. *Ann. Rev. Nutr.* 19, 197–216. doi:10.1146/annurev.nutr.19.1.197

Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., et al. (2013). Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* 7:116. doi:10.1186/1752-0509-7-116

Callahan, D. L., Souza, D. D., Bacic, A., and Roessner, U. (2009). Profiling of polar metabolites in biological extracts using diamond hydride-based aqueous normal phase chromatography. *J. Sep. Sci.* 32, 2273–2280. doi:10.1002/jssc.200900171

Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., et al. (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 34(Suppl. 1), D511–D516. doi:10.1093/nar/gkj128

Chelliah, V., Laibe, C., and Le Novère, N. (2013). "BioModels database: a repository of mathematical models of biological processes," in *Silico Systems Biology (Methods in Molecular Biology)*, ed. Schneider M. V. (New York, US: Humana Press), 189–199.

Chokkathukalam, A., Kim, D. H., Barrett, M. P., Breitling, R., and Creek, D. J. (2014). Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis* 6, 511–524. doi:10.4155/bio.13.348

Clasquin, M. F., Melamud, E., and Rabinowitz, J. D. (2012). LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* 14, 11. doi:10.1002/0471250953.bi1411s37

Colmsee, C., Mascher, M., Czauderna, T., Hartmann, A., Schlüter, U., Zellerhoff, N., et al. (2012). OPTIMAS-DW: a comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize. *BMC Plant Biol.* 12:245. doi:10.1186/1471-2229-12-245

Cooling, M. T., Rouilly, V., Misirli, G., Lawson, J., Yu, T., Hallinan, J., et al. (2010). Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics* 26, 925–931. doi:10.1093/bioinformatics/btq063

Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., et al. (2012). Computational tools for metabolic engineering. *Metab. Eng.* 14, 270–280. doi:10.1016/j.ymben.2012.03.001

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., et al. (2014). The reactome pathway knowledgebase. *Nucleic Acids Res.* 42, D472–D477. doi:10.1093/nar/gkt1102

Cuellar, A. A., Lloyd, C. M., Nielsen, P. F., Bullivant, D. P., Nickerson, D. P., and Hunter, P. J. (2003). An overview of CellML 1.1, a biological model description language. *Simulation* 79, 740–747. doi:10.1177/0037549703040939

Cuellar, A. A., Nielsen, P. F., Halstead, M., Bullivant, D. P., Nickerson, D. P., Hedley, W., et al. (2006). *CellML 1.1 Specification.* Available at: http://www.cellml.org/specifications/cellml_1.1/index_html

Daroch, M., Geng, S., and Wang, G. (2013). Recent advances in liquid biofuel production from algal feedstocks. *Appl. Energy* 102, 1371–1381. doi:10.1016/j.apenergy.2012.07.031

De Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., et al. (2009). Chemical entities of biological interest: an update. *Nucleic Acids Res.* 38, D249–D254. doi:10.1093/nar/gkp886

Dias, D. A., Hill, C. B., Jayasinghe, N. S., Atieno, J., Sutton, T., and Roessner, U. (2015). Quantitative profiling of polar primary metabolites of two chickpea cultivars with contrasting responses to salinity. *J. Chromatogr. B* 1000, 1–13. doi:10.1016/j.jchromb.2015.07.002

Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi:10.3390/metabo2020303

Drews, J. (2000). Drug discovery: a historical perspective. *Science* 287, 1960–1964. doi:10.1126/science.287.5460.1960

Dromms, R. A., and Styczynski, M. P. (2012). Systematic applications of metabolomics in metabolic engineering. *Metabolites* 2, 1090–1122. doi:10.3390/metabo2041090

Eisenreich, W., and Bacher, A. (2007). Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry* 68, 2799–2815. doi:10.1016/j.phytochem.2007.09.028

Fernie, A. R., and Morgan, J. A. (2013). Analysis of metabolic flux using dynamic labelling and metabolic modelling. *Plant Cell Environ.* 36, 1738–1750. doi:10.1111/pce.12083

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., et al. (2007). The metabolomics standards initiative (MSI). *Metabolomics* 3, 175–178. doi:10.1007/s11306-007-0070-6

Fitzpatrick, M. A., McGrath, C. M., and Young, S. P. (2014). Pathomx: an interactive workflow-based tool for the analysis of metabolomic data. *BMC Bioinformatics* 15:396. doi:10.1186/s12859-014-0396-9

Fu, C., Mielenz, J. R., Xiao, X., Ge, Y., Hamilton, C. Y., Rodriguez, M., et al. (2011). Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3803–3808. doi:10.1073/pnas.1100310108

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., et al. (2014). The synthetic biology open language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* 32, 545–550. doi:10.1038/nbt.2891

Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L. A., and Dixon, R. A. (2011). Salicylic acid mediates the reduced growth of lignin down-regulated plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20814–20819. doi:10.1073/pnas.1117873108

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi:10.1093/nar/gkg563

Gaude, E., Chignola, F., Spiliotopoulos, D., Spitaleri, A., Ghitti, M., Garcia-Manteiga, J. M., et al. (2013). muma, an R package for metabolomics univariate and multivariate statistical analysis. *Curr. Metabolomics* 1, 180–189. doi:10.2174/2213235x11301020005

Gowda, H., Ivanisevic, J., Johnson, C. H., Kurczy, M. E., Benton, H. P., Rinehart, D., et al. (2014). Interactive XCMS online: simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* 86, 6931–6939. doi:10.1021/ac500734c

Grafahrend-Belau, E., Weise, S., Koschützki, D., Scholz, U., Junker, B. H., and Schreiber, F. (2008). MetaCrop: a detailed database of crop plant metabolism. *Nucleic Acids Res.* 36(Suppl. 1), D954–D958. doi:10.1093/nar/gkm835

Graham, I. A. (2008). Seed storage oil mobilization. *Annu. Rev. Plant Biol.* 59, 115–142. doi:10.1146/annurev.arplant.59.032607.092938

Hartmann, A., and Schreiber, F. (2014). Integrative analysis of metabolic models – from structure to dynamics. *Front. Bioeng. Biotechnol.* 2:91. doi:10.3389/fbioe.2014.00091

Heinemann, M., and Panke, S. (2006). Synthetic biology – putting engineering into biology. *Bioinformatics* 22, 2790–2799. doi:10.1002/9780470437988.ch11

Hill, C. B., Bacic, A., and Roessner, U. (2014). "LC-MS profiling to link metabolic and phenotypic diversity in plant mapping populations," in *Mass Spectrometry in Metabolomics*, ed. Raftery D. (New York, NY: Springer), 29–41.

Hill, C. B., Jha, D., Bacic, A., Tester, M., and Roessner, U. (2013a). Characterization of ion contents and metabolic responses to salt stress of different *Arabidopsis AtHKT1;1* genotypes and their parental strains. *Mol. Plant* 6, 350–368. doi:10.1093/mp/sss125

Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Bacic, A., Langridge, P., et al. (2013b). Whole-genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment. *Plant Physiol.* 162, 1266–1281. doi:10.1104/pp.113.217851

Hill, C. B., and Roessner, U. (2013). "Metabolic profiling of plants by GC–MS," in *The Handbook of Plant Metabolomics*, eds Weckwerth W. and Kahl G. (Hoboken, NJ: John Wiley & Sons), 1–23.

Hill, C. B., and Roessner, U. (2015). "Advances in high-throughput LC-MS analysis for plant metabolomics," in *Advanced LC-MS Applications for Metabolomics*, ed. De Vooght-Johnson R. (London, UK: Future Science Ltd), 58–71.

Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Langridge, P., Bacic, A., et al. (2015). Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. *Plant Sci.* 233, 143–154. doi:10.1016/j.plantsci.2015.01.008

Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., Schaff, J. C., et al. (2015). The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core. *J. Int. Bioinformatics* 12, 266. doi:10.2390/biecoll-jib-2015-266

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531. doi:10.1093/bioinformatics/btg015

Jung, J. H., Fouad, W. M., Vermerris, W., Gallo, M., and Altpeter, F. (2012). RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel

production from lignocellulosic biomass. *Plant Biotechnol. J.* 10, 1067–1076. doi:10.1111/j.1467-7652.2012.00734.x

Junker, A., Hartmann, A., Schreiber, F., and Bäumlein, H. (2010). An engineer's view on regulation of seed development. *Trends Plant Sci.* 15, 303–307. doi:10.1016/j.tplants.2010.03.005

Junker, B. H., Klukas, C., and Schreiber, F. (2006). VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7:109. doi:10.1186/1471-2105-7-109

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi:10.1093/nar/gkt1076

Karnovsky, A., Weymouth, T., Hull, T., Tarcea, V. G., Scardoni, G., Laudanna, C., et al. (2012). Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* 28, 373–380. doi:10.1093/bioinformatics/btr661

Kaspar, S., Peukert, M., Svatos, A., Matros, A., and Mock, H. P. (2011). MALDI-imaging mass spectrometry – an emerging technique in plant biology. *Proteomics* 11, 1840–1850. doi:10.1002/pmic.201000756

Kastenmüller, G., Römisch-Margl, W., Wägele, B., Altmaier, E., and Suhre, K. (2011). metaP-server: a web-based metabolomics data analysis tool. *Biomed. Res. Int.* 2011, 839862. doi:10.1155/2011/839862

Kerren, A., and Schreiber, F. (2012). "Towards the role of interaction in visual analytics," in *Proc. Winter Simulation Conference – WSC* (Berlin: IEEE), Vol. 420, 1–13. doi:10.1109/WSC.2012.6465208

Kerren, A., and Schreiber, F. (2014). 'Why integrate InfoVis and SciVis? An example from systems biology'. *IEEE Comput. Graph. Appl.* 34, 6–10. doi:10.1109/mcg.2014.122

Kim, J., Um, B.-H., and Kim, T. (2012). Bioethanol production from micro-algae, *Schizocytrium* sp., using hydrothermal treatment and biological conversion. *Korean J. Chem. Eng.* 29, 209–214. doi:10.1007/s11814-011-0169-3

Kirchner, M., Saussen, B., Steen, H., Steen, J. A. J., and Hamprecht, F. A. (2007). amsrpm: robust point matching for retention time alignment of LC/MS data with R. *J. Stat. Softw.* 18, 1–12. doi:10.18637/jss.v018.i04

Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. doi:10.1126/science.1069492

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2005). GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* 21, 1635–1638. doi:10.1093/bioinformatics/bti236

Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., et al. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32(Suppl. 1), D438–D442. doi:10.1093/nar/gkh100

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., et al. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741. doi:10.1038/nbt.1558

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nat. Geosci.* 2, 831–836. doi:10.1038/ngeo689

Lessard, P. (1996). Metabolic engineering: the concept coalesces. *Nat. Biotechnol.* 14, 1654–1655. doi:10.1038/nbt1296-1654

Lloyd, C. M., Lawson, J. R., Hunter, P. J., and Nielsen, P. F. (2008). The CellML model repository. *Bioinformatics* 24, 2122–2123. doi:10.1093/bioinformatics/btn390

MathWorks. (2012). *MATLAB and Statistics Toolbox Release*. Natick, MA: The MathWorks, Inc.

Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D377–D386. doi:10.1093/nar/gks1118

Mi, H., Schreiber, F., Moodie, S., Czauderna, T., Demir, E., Haw, R., et al. (2015). Systems Biology Graphical Notation: activity flow language level 1 version 1.2. *J. Int. Bioinformatics* 12, 265. doi:10.2390/biecoll-jib-2015-265

Misirli, G., Hallinan, J. S., Yu, T., Lawson, J. R., Wimalaratne, S. M., Cooling, M. T., et al. (2011). Model annotation for synthetic biology: automating model to nucleotide sequence conversion. *Bioinformatics* 27, 973–979. doi:10.1093/bioinformatics/btr048

Miura, D., Fujimura, Y., and Wariishi, H. (2012). In situ metabolomic mass spectrometry imaging: recent advances and difficulties. *J. Proteomics* 75, 5052–5060. doi:10.1016/j.jprot.2012.02.011

Moodie, S., Le Novère, N., Demir, E., Mi, H., and Villéger, A. (2015). Systems Biology Graphical Notation: process description language level 1 version 1.3. *J. Int. Bioinformatics* 12, 263. doi:10.1038/npre.2009.3721.1

Morgat, A., Axelsen, K. B., Lombardot, T., Alcántara, R., Aimo, L., Zerara, M., et al. (2015). Updates in RHEA – a manually curated resource of biochemical reactions. *Nucleic Acids Res.* 43, D459–D464. doi:10.1093/nar/gku961

Mortimer, E., Maclean, J. M., Mbewana, S., Buys, A., Williamson, A. L., Hitzeroth, I. I., et al. (2012). Setting up a platform for plant-based influenza virus vaccine production in South Africa. *BMC Biotechnol.* 12:14. doi:10.1186/1472-6750-12-14

Mueller, L. A., Zhang, P., and Rhee, S. Y. (2003). AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132, 453–460. doi:10.1104/pp.102.017236

O'Grady, J., Schwender, J., Shachar-Hill, Y., and Morgan, J. A. (2012). Metabolic cartography: experimental quantification of metabolic fluxes from isotopic labelling studies. *J. Exp. Bot.* 63, 2293–2308. doi:10.1093/jxb/ers032

Petrie, J. R., Shrestha, P., Liu, Q., Mansour, M. P., Wood, C. C., Zhou, X. R., et al. (2010). Rapid expression of transgenes driven by seed-specific constructs in leaf tissue: DHA production. *Plant Methods* 6, 8–13. doi:10.1186/1746-4811-6-8

Petrie, J. R., Shrestha, P., Zhou, X. R., Mansour, M. P., Liu, Q., Belide, S., et al. (2012). Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLoS One* 7:e49165. doi:10.1371/journal.pone.0049165

Quinn, J., Beal, J., Bhatia, S., Cai, P., Chen, J., Clancy, K., et al. (2013). *Synthetic Biology Open Language Visual (SBOL Visual) Version 1.0.0 BBF RFC #93*.

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rios-Estepa, R., and Lange, B. M. (2007). Experimental and mathematical approaches to modeling plant metabolic networks. *Phytochemistry* 68, 2351–2374. doi:10.1016/j.phytochem.2007.04.021

Roberts, M. F., and Wink, M. (eds) (1998). *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*. New York, USA: Springer Science & Business Media.

Roehner, N., Zhang, Z., Nguyen, T., and Myers, C. J. (2015). Generating systems biology markup language models from the synthetic biology open language. *ACS Synth. Biol.* 4, 873–879. doi:10.1021/sb5003289

Rohn, H., Junker, A., Hartmann, A., Grafahrend-Belau, E., Treutler, H., Klapperstück, M., et al. (2012). VANTED v2: a framework for systems biology applications. *BMC Syst. Biol.* 6:139. doi:10.1186/1752-0509-6-139

Rohn, H., Klukas, C., and Schreiber, F. (2011). Creating views on integrated multidomain data. *Bioinformatics* 27, 1839–1845. doi:10.1093/bioinformatics/btr282

Rohwer, J. M. (2012). Kinetic modelling of plant metabolic pathways. *J. Exp. Bot.* 63, 2275–2292. doi:10.1093/jxb/ers080

Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., et al. (2007). Storing and annotating of kinetic data. *In Silico Biol.* 7, 37–44.

Runguphan, W., Glenn, W. S., and O'Connor, S. E. (2012). Redesign of a dioxygenase in morphine biosynthesis. *Chem. Biol.* 19, 674–678. doi:10.1016/j.chembiol.2012.04.017

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi:10.1038/nbt.2842

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., et al. (2010). BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* 39, D670–D676. doi:10.1093/nar/gkq1089

Schreiber, F., Colmsee, C., Czauderna, T., Grafahrend-Belau, E., Hartmann, A., Junker, A., et al. (2012). MetaCrop 2.0: managing and exploring information about crop plant metabolism. *Nucleic Acids Res.* 40, D1173–D1177. doi:10.1093/nar/gkr1004

Schreiber, F., Bader, G. D., Golebiewski, M., Hucka, M., Kormeier, B., Le Novère, N., et al. (2015). Specifications of standards in systems and synthetic biology. *J. Int. Bioinformatics* 12, 258. doi:10.2390/biecoll-jib-2015-258

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Singh, N. K., and Dhar, D. W. (2011). Microalgae as second generation biofuel. A review. *Agronomy Sustain. Dev.* 31, 605–629. doi:10.1007/s13593-011-0018-0

Sommer, B., Dingersen, T., Gamroth, C., Schneider, S. E., Rubert, S., Krüger, J., et al. (2011). CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous membrane packing problems. *J. Chem. Inf. Model.* 51, 1165–1182. doi:10.1021/ci1003619

Sorokin, A., Le Novère, N., Luna, A., Czauderna, T., Demir, E., Haw, R., et al. (2015). Systems Biology Graphical Notation: entity relationship language level 1 version 2. *J. Int. Bioinformatics* 12, 264. doi:10.2390/biecoll-jib-2015-264

Töpfer, N., Caldana, C., Grimbs, S., Willmitzer, L., Fernie, A. R., and Nikoloski, Z. (2013). Integration of genome-scale modeling and transcript profiling

reveals metabolic pathways underlying light and temperature acclimation in *Arabidopsis. Plant Cell* 25, 1197–1211. doi:10.1105/tpc.112.108852

Töpfer, N., Jozefczuk, S., and Nikoloski, Z. (2012). Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli. BMC Syst. Biol.* 6:148. doi:10.1186/1752-0509-6-148

Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Glé, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19748–19753. doi:10.1073/pnas.1309299110

Trethewey, R. N. (2004). Metabolite profiling as an aid to metabolic engineering in plants. *Curr. Opin. Plant Biol.* 7, 196–201. doi:10.1016/j.pbi.2003.12.003

Villas-Boas, S. G., Nielsen, J., Smedsgaard, J., Hansen, M. A., and Roessner-Tunali, U. (2007). *Metabolome Analysis: An Introduction*, Vol. 24. Hoboken, USA: John Wiley & Sons.

Xia, J., Sinelnikov, I. V., Han, B., and Wishart, D. S. (2015). MetaboAnalyst 3.0 – making metabolomics more meaningful. *Nucleic Acids Res.* 43, W251–W257. doi:10.1093/nar/gkv380

Xia, J., Sinelnikov, I. V., and Wishart, D. S. (2011). MetATT: a web-based metabolomics tool for analyzing time-series and two-factor datasets. *Bioinformatics* 27, 2455–2456. doi:10.1093/bioinformatics/btr392

Yang, Y., Engin, L., Wurtele, E. S., Cruz-Neira, C., and Dickerson, J. A. (2005). Integration of metabolic networks and gene expression in virtual reality. *Bioinformatics* 21, 3645–3650. doi:10.1093/bioinformatics/bti581

Yu, T., Park, Y., Johnson, J. M., and Jones, D. P. (2009). apLCMS – adaptive processing of high-resolution LC/MS data. *Bioinformatics* 25, 1930–1936. doi:10.1093/bioinformatics/btp291

Zhang, K., Bhuiya, M. W., Pazo, J. R., Miao, Y., Kim, H., Ralph, J., et al. (2012). An engineered monolignol 4-O-methyltransferase depresses lignin biosynthesis and confers novel metabolic capability in *Arabidopsis. Plant Cell* 24, 3135–3152. doi:10.1105/tpc.112.101287