An Integration and Analysis Pipeline for Systems Biology in Crop Plant Metabolism

Stephan Weise, Christian Colmsee, Eva Grafahrend-Belau, Björn Junker, Christian Klukas, Matthias Lange, Uwe Scholz and Falk Schreiber

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Gatersleben, Germany
{weise,colmsee,grafahr,junker,klukas,lange, scholz,schreibe}@ipk-gatersleben.de

Abstract. To advance the comprehension of complex biological processes occurring in crop plants (e.g. for improvement of growth or yield) it is of high interest to reconstruct and analyse detailed metabolic models. Therefore, we established a pipeline combining software tools for (1) storage of metabolic pathway data and reconstruction of crop plant metabolic models, (2) simulation and analysis of stoichiometric and kinetic models and (3) visualisation of data generated with these models. The applicability of the approach is demonstrated by a case study of cereal seed metabolism.

1 Introduction

Crop plants form one of the main sources of human and animal nutrition and importantly contribute to chemical or pharmaceutical industry and renewable resources [1–4]. In order to achieve an improvement of growth and yield of crop plants it is necessary to understand biological processes on a detailed level [5, 6].

Due to the increasing amount of data obtained by modern high-throughput technologies it becomes more difficult to perform all necessary experiments conventionally. However, such large amounts of data enable new analysis and modelling techniques. In order to cope with these challenges, systems biology aims to understand biological processes as a whole and to map onto mathematical models [7], thus enabling *in silico* experiments.

Mathematical modelling of metabolism enables analysis of the structure, dynamics and behaviour of metabolic networks. With the help of these models, understanding of complex processes can be verified and extended, new hypotheses can be generated, and suitable targets for metabolic engineering can be identified by exploring *in silico* scenarios. In plant metabolism, different methods for mathematical modelling are constantly gaining attention [8–10].

To deal with such models, several aspects have to be taken into consideration: model representation, model exchange, model analysis and model visualisation.

Data about biological processes in plants is available from various, often ambiguous, sources and thus needs to be integrated and persistently stored in a

central, well-structured repository. To ensure high data quality, the integrated data needs to be curated manually [11]. This comprises, for example, to distinguish different growth and developmental stages using controlled vocabulary. Once the data is prepared this way it can be used together with different tools for visualisation, simulation and analysis purposes. Therefore, standardised exchange formats such as the Systems Biology Markup Language (SBML) [12] or the Biological Pathways Exchange Language (BioPAX) [13] should be used.

Depending on the objective of the modelling process and the available data used for model reconstruction, different model analysis techniques can be applied to the reconstructed metabolic model ranging from purely stoichiometric to kinetic approaches of model analysis. With each of these modelling techniques being characterised by certain advantages and drawbacks, an analysis pipeline offering a variety of different model analysis techniques is needed to support user-friendly and flexible model simulation and analysis.

With increasing amount of large-scale experimental data, simulation results and large biological networks, visualisation methods become more and more important for data analysis. Advanced visualisation methods aim at bringing the data in a form that, on the one hand gives an overview about the overall system, and on the other hand provides sufficient detail. Static visualisation is inadequate for large scale data exploration, thus in order to assist modern biological research dynamic visualisation and user-friendly interaction methods need to be implemented in supportive tools.

This paper describes a pipeline for supporting the research on crop plant metabolic models. It comprises the following steps: (1) data management and individual, user-specific model reconstruction, (2) stoichiometric and kinetic model analysis and (3) model and flux visualisation. Each of the steps is described focusing on tools and methods developed. The applicability of the pipeline is shown by a case study of storage metabolism in developing seeds of the agriculturally important crop species barley.

2 Methods

For the integration and analysis of data describing metabolic networks of plants a pipeline was developed, which is summarised in Figure 1. The respective steps are described in detail below.

2.1 Model Storage and Reconstruction

The reconstruction of plant metabolic models is an important step for a better understanding of biological processes. Therefore, detailed metabolic information up to compartment level needs to be collected and managed in a well-structured way. On this account we developed MetaCrop (http://metacrop.ipk-gatersleben.de) [14], a manually curated repository of high quality data of seven major crop plants with high agronomical importance (*Hordeum vulgare* (barley), *Triticum aestivum* (wheat), *Oryza sativa* (rice), *Zea mays* (maize), *Beta vulgaris*

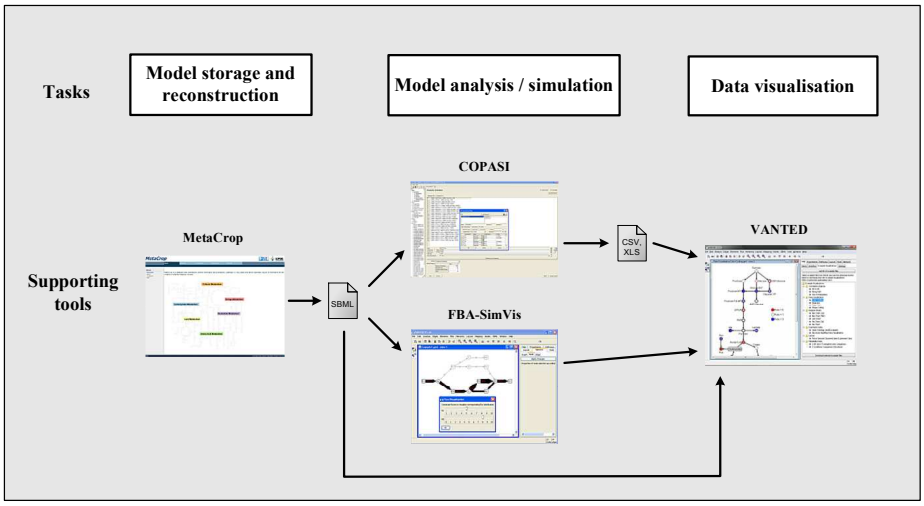

Fig. 1. Comprehensive pipeline of integration and analysis of pathway related data

(beet), *Solanum tuberosum* (potato) and *Brassica napus* (canola)) and two model plants widely used in plant research (*Arabidopsis thaliana* (thale cress) and *Medicago truncatula* (barrel medic)). MetaCrop is available for scientists working in the area of plant research, thus accelerating the process of data curation.

MetaCrop uses the Meta-All software [15]. Major concepts of MetaCrop are *substances* and *conversion processes*. The latter of which should be understood as a reaction or a translocation, which can be either actively or passively. Substances play certain roles within these conversion processes, such as substrate, product, catalyst or inhibitor. To enable the successive construction of metabolic models, conversions can be combined to pathways and pathways to super-pathways.

Fine-grained information can be stored with every conversion element managed in MetaCrop, such as reaction or translocation type, formula, synonyms, EC numbers, literature references and also kinetic data like V_{max} values, affinity or inhibitor constants. Moreover, all this information is assigned with the compartment the respective element is located in, thus considering that conversions take place at different locations inside an organism depending on the developmental state and environmental effects. Pathways can be stored as different parallel versions and furthermore, there is the possibility to assign quality tags due to different quality levels in the data used for MetaCrop.

Since there are only a few models for crop plants existing at all, the time-consuming process of manual data curation is indispensable. Data can be imported into MetaCrop using either the standardised SBML format [12] or a web interface and can then be curated manually. To enable visualisation, simulation and analysis metabolic pathway data can be exported using the SBML format. Therefore, MetaCrop offers an export functionality, which is integrated into the graphical user interface. It enables a user to compose an individual pathway

model by adding elements such as reactions, translocation processes or even whole pathways into a 'shopping cart' step by step. By means of a wizard, export settings can be defined then, e. g. the use of certain kinetics or a restriction to data from a certain species only. The generated SBML file contains function definitions, unit definitions, all involved species, annotations and the reactions including their kinetic data. Additionally, it is possible to export structural models only. Both SBML import and export can also be performed on command line thus enabling batch runs.

2.2 Model Analysis and Simulation

Stoichiometric Model Analysis Stoichiometric model analysis or constraint-based modelling is based on the knowledge about the topological structure of the metabolic network under consideration. Due to the advantage of not requiring the knowledge of enzyme kinetic properties, constraint-based modelling approaches such as Flux balance analysis (FBA) have become an important approach for understanding the capabilities and properties of metabolic networks [16, 17].

To provide a user-friendly environment for the constraint-based analysis of crop plant metabolic models, we developed FBA-SimVis (http://fbasimvis.ipk-gatersleben.de), a VANTED ([18], see section 2.3) plug-in for integrated constraint-based model analysis and visualisation. The plug-in integrates various constraint-based analysis techniques (Flux balance analysis, Knock-out analysis, Robustness analysis and Flux variability analysis) with interactive and dynamic visualisation routines to support the quantitative analysis of stoichiometric models of plant metabolism. Due to the dynamic and visual exploration of simulated flux data resulting from model analysis, aimed at facilitating the analysis and interpretation of metabolic fluxes in response to genetic and/or environmental conditions, FBA-SimVis provides a comprehensive understanding of constraint-based metabolic models in both overview and detail.

To perform constraint-based analysis of a crop plant metabolic model exported from MetaCrop, the model has to be imported into FBA-SimVis as a SBML file, which forms the basis for subsequent stoichiometric model analysis and flux visualisation.

Kinetic Model Analysis Kinetic models are the most detailed models of metabolism and allow most fine-grained predictions on metabolic behaviour. However, several issues arise due to their complexity: kinetic models require detailed kinetic knowledge about the modelled enzymes, and calculations quickly reach the limits of affordable computational power. Furthermore, the establishment of kinetic models is a tedious task requiring acquisition of various types of detailed information. In the integration and analysis pipeline presented in this paper, SBML files that were exported from MetaCrop contain all the necessary information to build a kinetic model and thus can be imported in various simulation software packages. As an example for a simulation software we use

COPASI [19], which is a GUI-based tool that allows easy model editing and simulation also for non-experts. Once the model has been imported into COPASI, it can be edited and refined by the user, so that also data not contained in MetaCrop can be added. COPASI then allows to simulate the model as a time-course, to calculate a steady-state and analyse its stability, to perform parameter estimation based on experimental data, to perform metabolic control analysis, etc. The calculated results are stored in text files which can be opened in spreadsheet programmes and thus passed on to data visualisation tools as explained in the next section. Furthermore, data can be exported in SBML format and can then be used to update MetaCrop.

2.3 Model Visualisation

VANTED (visualisation and analysis of networks containing experimental data) [18] is a platform-independent open source software which enables researchers to evaluate extensive experiment data from multiple *-omics* areas (transcriptomics, proteomics and metabolomics) and in the same way, the results of simulation studies. In order to support the analysis and visualisation of metabolic flux data, the data-mapping method, which initially was targeted at connecting experiment data to graph nodes, now supports the assignment of experimental datasets to graph edges.

Flexible network-integrated visualisation techniques may be used to visualise the data connected to graph nodes or edges. Available approaches include line, bar and pie charts as well as transformation functions, which map observed data to visual properties such as node size, edge width or graph element colours.

To support flexible transfer of graph models with various software tools and data sources, the following file formats are supported: GML, GraphML, DOT, Pajek .NET, KGML, SBML. Experiment data may be loaded into the system in the form of Excel spreadsheet files (XLS) or as comma separated value files (CSV).

3 Case Study

With the aim of getting a systemic understanding of cereal seed storage metabolism, a stoichiometric model of central metabolism in the developing endosperm of barley (*Hordeum vulgare*) was reconstructed by integrating biochemical, physiological and proteomic data derived from literature and databases into MetaCrop. The resulting compartmented stoichiometric model includes 257 conversion processes (193 reactions, 64 transport processes) and 234 metabolites, compartmentalised between the extracellular medium and the intracellular compartments cytosol, mitochondria and plastid.

To study grain yield and metabolic flux distributions under varying growth conditions, the model was subjected to Flux balance analysis using FBA-SimVis. Parameters necessary to perform FBA (e. g. maximum uptake and excretion

rates) were extracted from published experimental results and the model was validated by comparing the simulation results to literature-based findings.

In general, the simulation results were found to be in good agreement with the main qualitative physiological characteristics of cereal seed storage metabolism. For example, the growth rate and the metabolic pathway pattern under fully aerobic growth conditions predicted by the model were in accordance with published experimental results (see Figure 2 for an example). Thus, by providing an initial template for studying seed metabolic behaviour *in silico*, in future applications the model can be used to generate or test hypothesis on ways to improve grain and yield quality.

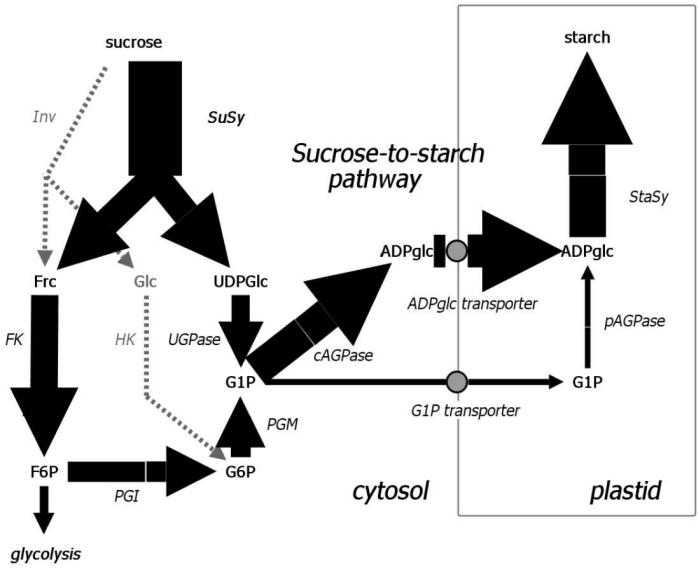

Fig. 2. Carbon flux map of the Sucrose-to-starch pathway predicted by the stoichiometric model of cereal seed metabolism under fully aerobic growth conditions (visualised with VANTED)

4 Conclusion

In this paper we presented a data integration and analysis pipeline for data about crop plant metabolism. It comprises the steps model storage and reconstruction, model analysis and model visualisation applying MetaCrop, FBA-SimVis / COPASI and VANTED, respectively. All tools used are publicly available. As a proof of concept a case study illustrating creation, analysis and visualisation of a model of cereal seed metabolism was shown. The pipeline as described here is applicable to all other crop plant species managed in MetaCrop. In principle,

it could also be used for non-plant species by establishing a separate repository employing the Meta-All software.

Generally, the presented pipeline could be fully automated by applying workflow management systems, such as Taverna [20] or Kepler [21]. But on current data in crop plant biology we reckon the use of professionals' expertise as indispensable. Therefore, manual interaction is needed.

5 Acknowledgements

This work was partly supported by the German Federal Ministry of Education and Research (BMBF) under grants 031270-6A, 0315044A and 0315295.

References

1. Tilman, D., Hill, J., Lehman, C.: Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. Science **314**(5805) (2006) 1598–1600
2. Metzger, J., Bornscheuer, U.: Lipids as renewable resources: current state of chemical and biotechnological conversion and diversification. Applied Microbiology and Biotechnology **71**(1) (2006) 13–22
3. Scheller, J., Guhrs, K.H., Grosse, F., Conrad, U.: Production of spider silk proteins in tobacco and potato. Nature Biotechnology **19**(6) (2001) 573–577
4. DellaPenna, D.: Nutritional Genomics: Manipulating Plant Micronutrients to Improve Human Health. Science **285**(5426) (1999) 375–379
5. Jenner, H.: Transgenesis and yield: what are our targets? Trends in Biotechnology **21**(5) (2003) 190–192
6. Carrari, F., Urbanczyk-Wochniak, E., Willmitzer, L., Fernie, A.: Engineering central metabolism in crop species: learning the system. Metabolic Engineering **5**(3) (2003) 191–200
7. Kitano, H.: Systems biology: A brief overview. Science **295**(5560) (2002) 1662–1664
8. Giersch, C.: Mathematical modelling of metabolism. Current Opinion in Plant Biology **3**(3) (2000) 249–253
9. Morgan, J., Rhodes, D.: Mathematical modeling of plant metabolic pathways. Metabolic Engineering **4**(1) (2002) 80–89
10. Poolman, M., Assmus, H., Fell, D.: Applications of metabolic modelling to plant metabolism. Journal of Experimental Botany **55**(400) (2004) 1177–1186
11. Zhang, P., Foerster, H., Tissier, C., Mueller, L., Paley, S., Karp, P., Rhee, S.: MetaCyc and AraCyc. Metabolic Pathway Databases for Plant Research. Plant Physiology **138**(1) (2005) 27–37
12. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Novere, N.L., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J.: The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics **19**(4) (2003) 524–531

13. Luciano, J.: PAX of mind for pathway researchers. Drug Discovery Today **10**(13) (2005) 937–942
14. Grafahrend-Belau, E., Weise, S., Koschützki, D., Scholz, U., Junker, B., Schreiber, F.: MetaCrop: a detailed database of crop plant metabolism. Nucleic Acids Research **36**(Database issue) (2008) D954–D958
15. Weise, S., Grosse, I., Klukas, C., Koschützki, D., Scholz, U., Schreiber, F., Junker, B.H.: Meta-All: a system for managing metabolic pathway information. BMC Bioinformatics **7**(1) (2006) e465.1–9
16. Varma, A., Palsson, B.: Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nature Biotechnology **12**(10) (1994) 994–998
17. Kauffman, K., Prakash, P., Edwards, J.: Advances in flux balance analysis. Current Opinion in Biotechnology **14**(5) (2003) 491–496
18. Junker, B., Klukas, C., Schreiber, F.: VANTED: A system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics **7**(1) (2006) e109.1–13
19. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U.: COPASI–a COmplex PAthway SImulator. Bioinformatics **22**(24) (2006) 3067–3074
20. Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M., Li, P., Oinn, T.: Taverna: a tool for building and running workflows of services. Nucleic Acids Research **34**(Webserver issue) (2006) W729–W732
21. McPhillips, T., Bowers, S., Zinn, D., Ludäscher, B.: Scientific workflow design for mere mortals. Future Generation Computer Systems **25**(5) (2009) 541–551