Australian Journal of Crop Science AJCS

AJCS 13(04):605-615 (2019) ISSN:1835-2707
doi: 10.21475/ajcs.19.13.04.p1639

Development of a compartmentalized model for insight into the structured metabolic pathway of carbon metabolism in cassava leaves

Nattharat Punyasu[1], Saowalak Kalapanulak[2, 3], Wanatsanan Siriwat[3] and Treenut Saithong[2, 3 *]

[1]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, and School of Information Technology, King Mongkut's University of Technology Thonburi (Bang Khun Thian), Bangkok, 10150, Thailand
[2]Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi (Bang Khun Thian), Bangkok, 10150, Thailand
[3]Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi (Bang Khun Thian), Bangkok, 10150, Thailand

[*]Corresponding author: treenut.sai@kmutt.ac.th

Abstract

In silico metabolic modeling has enabled systematic study of complicated metabolic processes underlying phenotypes of organisms. Modeling of plant metabolism is often hampered by the network complexity and lack of adequate knowledge. The existing metabolic networks of cassava only cover broad metabolism and are not compartmentalized to truly represent metabolism in photosynthetic tissues. To address the aforementioned limitations and develop a robust metabolic network, physiological and genomic data derived from cassava and leaf models of Arabidopsis and rice were to extend the scope of the existing model. The proposed compartmentalized network of metabolism in photosynthetic tissues of cassava, **ph-MeRecon** (**ph**otosynthetic-***M****anihot* ***e****sculenta* Metabolic Pathway **Recon**struction) was developed based on the information resulting of the comparative study of multiple model plants and cassava genome. The ph-MeRecon covers primary carbon metabolism and comprises 461 metabolites, 550 reactions, and 1,037 metabolic genes. Enzymatic genes on the network were validated using RNA-expression data, and the reactions and pathways were compartmentalized into cytoplasm, chloroplast, mitochondria, and peroxisome. To ensure network connectivity, metabolic gaps were filled using gap reactions obtained from literature and metabolic pathway omnibus. In addition, information on plant physiology, including photosynthetic light-dependent reactions, carboxylase and oxygenase activity of RuBisCO enzyme, and phosphoenolpyruvate carboxylase enzyme activity was incorporated into ph-MeRecon to mimic cellular metabolism in cassava leaves. Thus, ph-MeRecon offers a multi-level platform for system analysis of cellular mechanisms underlying phenotypes of interest in cassava. The ph-MeRecon metabolic model is available at http://bml.sbi.kmutt.ac.th/ph-MeRecon/.

Keywords: carbon metabolism, data-integrated pathway reconstruction, *Manihot esculenta*, metabolic pathway reconstruction, photosynthetic metabolism, qualitative model.
Abbreviations: ph-MeRecon-d1_photosynthetic-*Manihot esculenta* Metabolic Pathway Reconstruction draft 1; ph-MeRecon-d2_photosynthetic-*Manihot esculenta* Metabolic Pathway Reconstruction draft 2; ph-MeRecon_photosynthetic-*Manihot esculenta* Metabolic Pathway Reconstruction. Abbreviations of metabolite name are provided in Supplementary data 1.

Introduction

Metabolism is a highly dynamic process and needs to continuously adapt to prevailing environmental conditions to maintain homeostasis in organisms. *In silico* metabolic modeling has enabled the characterization and prediction of metabolic potential and behavior, thereby providing insight into how individual components cooperate to influence cellular functions (Oberhardt et al., 2009; Raman and Chandra, 2009). To date, over 50 genome-scale metabolic models have been constructed in a broad range of organisms, from simple prokaryotes to complex eukaryotes (Kim et al., 2012). However, due to the following difficulties: (i) lack of complete genome information, (ii) insufficient knowledge about the metabolic network and missing components, (iii) duplication of metabolic pathways and transport of metabolites across compartments, (iv) metabolic differences at cell and tissue levels, and varied metabolic response to environmental change, only a few exist for plants, such as Arabidopsis (Beckers et al., 2016; de Oliveira Dal'Molin et al., 2010a; Mintz-Oron et al., 2012), maize (*Zea mays*) (de Oliveira Dal'Molin et al., 2010b; Saha et al., 2011), and rice (*Oryza sativa*) (Lakshmanan et al., 2015; Lakshmanan et al., 2013). Metabolic modeling has been employed to study metabolic adaptation to abiotic stress (Lakshmanan et al., 2016; Lakshmanan et al., 2013) and yield improvement in plants, e.g. oil and carbohydrate syntheses in rapeseed (Hay and Schwender, 2011) and rice (Shaw and Kundu, 2015), respectively. Regrettably, not much is known about metabolism in cassava, a strategic food security crop upon which many vulnerable households in the tropics rely for daily calorie needs. Cassava (*Manihot esculenta*) is one of the most important crops in this era and feeds approximately 800 million people yearly (Howeler et

al., 2013). It is recognized as king of starchy staple crops due to the huge amount of starch synthesized and stored in its roots. Research has sought to understand the metabolic advantage of cassava over other crops in relation to starch synthesis. Moreover, a comprehensive understanding of its metabolism is crucial for developing high-yielding cultivars that are resilient to climate change (Rossi et al., 2015). The pathways of starch and sucrose metabolism in cassava were first reported in 2007 and were based on full-length cDNA expression (Sakurai et al., 2007). The release of the cassava genome sequence has enabled a large-scale prediction of enzymatic functions of proteins as well as their corresponding metabolic reactions and pathways. The extended metabolic pathways are available in the Plant Metabolic Network (PMN) database (Schlapfer et al., 2017). Comparative genomics approach, in addition, has made it possible to evaluate the convergence and divergence of metabolic pathways in cassava and evolutionarily related plants, and has deepened our understanding of carbon assimilation in cassava (Rongsirikul et al., 2010; Saithong et al., 2013; Siriwat, 2012).

Reconstruction of metabolic network merely based on metabolic components, i.e. enzymatic genes, reactions, etc., is inadequate. Given the multi-cellular nature of plants and the many compartments therein, reconstruction of a robust metabolic network entails compartmentalizing metabolic components and ensuring network connectivity via gap filling, besides incorporating information on plant physiology and transporters. In this research, we studied the primary carbon metabolism in cassava to understand the metabolic processes relevant to biomass synthesis, i.e. carbohydrates, proteins, fibers, and lipid. We propose a compartmentalized metabolic model of cassava leaf cells, **ph**otosynthetic-***M**anihot **e**sculaenta* Metabolic Pathway **Recon**struction **(ph-MeRecon)**. The ph-MeRecon consists of 461 metabolites and 550 reactions associated with 1,037 metabolic genes. Cassava enzymatic genes were identified through comparative genomics approach, and were validated using existing mRNA expression data (Wilson et al., 2017). The constituent biochemical reactions were distributed into four intracellular compartments: cytosol, chloroplast, mitochondria, and peroxisome. In addition, our model accounts for the extensive knowledge available for carbon metabolism in cassava leaves, for which detailed metabolic reactions were included to represent the far-most understanding on the metabolic processes, involved in photosynthetic electron transfer, carboxylation and oxygenation of RuBisCO enzyme, role of phosphoenolpyruvate carboxylase (PEPC) enzyme and starch metabolism in leaves. Thus, ph-MeRecon marks a huge step forward towards discovering new strategies for improving cassava crop productivity via *in silico* metabolic modeling.

Results and discussion

A qualitative model of carbon metabolism in cassava leaves

Primary carbon metabolism is the backbone of metabolic processes essential for growth and maintenance of cellular functions. In plants, it begins with the carbon fixation pathway whereby atmospheric CO_2 is captured and converted to cellular carbon substrate, which is utilized for biomass synthesis through various pathways. The proposed model begins with the fixation of atmospheric carbon, and covers the biosynthesis of carbohydrates (starch, sucrose, fructose and glucose), fibers (cellulose and xylan), eighteen amino acids, and palmitic acid (Supplementary Table 1), biomass components of cassava leaves as reported by Montagnac et al. (2009). Thus, ph-MeRecon covers the photosynthetic light reaction, Calvin cycle, respiration (glycolysis, tricarboxylic acid (TCA) cycle, oxidative phosphorylation, and fermentation), pentose phosphate pathway (PPP), photorespiration, starch and sucrose metabolism, and anabolic pathways of amino acids, fatty acid, fibers, and nucleic acids. First, as outlined in Supplementary Figure 1, the network was reconstructed based on biochemical reactions derived from the genome-based pathway of carbon metabolism in cassava, MeRecon (Siriwat, 2012). Next, the 129 biochemical reactions (Figure 1) common in the rice (Lakshmanan et al., 2015; Lakshmanan et al., 2013) and Arabidopsis leaf models (Beckers et al., 2016; de Oliveira Dal'Molin et al., 2010a; Mintz-Oron et al., 2012; Saha, et al., 2011) were added, to reflect the metabolism in photosynthetic tissues. Additionally, 47 reactions overlapping two of the leaf models, at least, were added to fully connect the common reactions to MeRecon. The added reactions (common and partially overlapping) were confirmed by the existence of the corresponding enzymatic genes in cassava genome. The preliminary draft of carbon metabolism in cassava leaves (ph-MeRecon-d1) thus consists of 321 biochemical reactions and 1,029 metabolic genes of cassava, representing an additional 65 reactions and 198 genes (Figure 2A and Supplementary Table 2). To ensure that ph-MeRecon-d1 could represent the metabolism in cassava leaves, the expression of genes responsible for each biochemical reaction was investigated using transcriptome data from 3-months old cassava leaf tissues (Wilson et al., 2017). Only genes that were expressed in at least two of the three profile replicates were selected. In total, 95 percent of genes in the model (974 of 1,029 genes), corresponding to all metabolic reactions, were validated. Subsequently, the biochemical reactions in ph-MeRecon-d1 were compartmentalized into cytosol [c], chloroplast [p], mitochondria [m], and peroxisome [x], based on the signal peptide region of the enzymatic protein sequences and literature (Buchanan et al., 2015) (Figure 2B). The glycolysis pathway, pentose phosphate pathway, sucrose metabolism, and primary cell wall metabolism were located in cytosolic compartment; the photosynthetic light-dependent reactions, Calvin-Benson cycle, starch metabolism, and fatty acid metabolism were assigned to chloroplast; the TCA cycle, oxidative phosphorylation, and photorespiration were located in mitochondria; while metabolic reactions related to the photorespiration pathway were in peroxisome. Afterwards, transport reactions were assigned to relevant metabolites in order to enhance connectivity across subcellular compartments. In total, 106 intracellular transport reactions and 41 exchange reactions were assigned to convey 84 metabolites across subcellular and plasma membranes, respectively (Figure 2D).

Following the compartmentalization, ph-MeRecon-d2 was screened for metabolic gaps using GapFind and detectDeadEnds algorithms. Nine gap metabolites were found, of which three could not be produced by any reaction in the network (root no-production) and six could not be consumed by any reaction in the network (root no-

consumption). The gaps were filled using three enzymatic, two spontaneous, and two transport reactions. The genes of the gap reactions were found in the cassava genome, supporting the occurrence of the reactions in cassava metabolism.

Supplementary Figure 2 shows an example of the gap filling process for the gap metabolites, (S)-1-Pyrroline-5-carboxylate and L-glutamate 5-semialdehyde. (S)-1-Pyrroline-5-carboxylate was identified as 'root no-production' gap metabolite, whereas L-glutamate 5-semialdehyde was noted as 'root no-consumption' gap metabolite (Supplementary Figure 2A). These gap metabolites were connected by adding a spontaneous reaction (KEGG ID R03314) reported to catalyze L-Glutamate 5-semialdehyde to (S)-1-Pyrroline-5-carboxylate in the proline biosynthesis pathway in Arabidopsis (de Oliveira Dal'Molin et al., 2010a; Saha et al., 2011) and rice (Lakshmanan et al., 2015), as shown in Supplementary Figure 2B. Finally, the atomic mass of reactions was balanced, resulting in ph-MeRecon.

The ph-MeRecon consists of 550 reactions, 1,037 unique enzymatic genes, and 461 metabolites (Figure 2C, Supplementary Table 3 and available at http://bml.sbi.kmutt.ac.th/ph-MeRecon/). The reactions (550) comprise 399 enzymatic reactions, 2 spontaneous reactions, 108 intracellular transport reactions, and 41 exchange reactions and were assigned to cytosol (139), chloroplast (203), mitochondria (46), and peroxisome (11) (Figure 2D). All the biochemical reactions in the model were fully supported by the genome and expression information. However, the transport and exchange reactions were determined based on gene annotation and minimum number of metabolite transporters that are required to activate the enzymes with known localization in the compartments (de Oliveira Dal'Molin and Nielsen, 2013). Moreover, they were required by the network for the transfer of energy and metabolite precursors. These criteria were adopted in all compartmentalized models published to date (e.g. Poolman et al., 2009; Lakshmanan et al., 2013). Given the limited information on plant-specific transporters and the challenge of performing experimental validation, quantitative modeling of metabolic pathways could help reduce excessive prediction of transport reactions in compartmentalized models (Thiele and Palsson, 2010). The examples of the approach present in various model development studies, including Beckers et al., 2016; de Oliveira Dal'Molin et al., 2010a; Lakshmanan et al., 2013; Poolman et al., 2013; Poolman et al., 2009; Saha et al., 2011.

The ph-MeRecon: a representative model of primary carbon metabolism in cassava leaves

The ph-MeRecon model was reconstructed mainly based on cassava genome information, and the metabolic information on leaf tissues of template plant models (*see* Method). Here, the specificity of ph-MeRecon to represent primary carbon metabolism in cassava leaves was assessed. The ph-MeRecon was compared with the Arabidopsis genome-scale metabolic model —AraGEM (de Oliveira Dal'Molin et al., 2010a), Arabidopsis *i*RS1597 model (Saha et al., 2011), Arabidopsis leaf model (Beckers et al., 2016), rice leaf model (Lakshmanan et al., 2013), and tomato leaf model (Yuan et al., 2016), with respect to model characteristics.

The ph-MeRecon is unique, containing more biochemical reactions and metabolites than the rice leaf model but less compared to three Arabidopsis leaf models (Table 1). However, due to difference in model scope, the biochemical reactions and metabolites in the other three models (AraGEM, *i*RS1597, and tomato models) are considerably higher. Of the non-redundant reactions in all models under comparison, 14 unique reactions were found in ph-MeRecon — specifically, in the amino acid biosynthesis, starch and sucrose metabolism, and nucleotide biosynthesis pathways (Supplementary Table 4). The uniqueness of the reactions reflects the diversity of metabolic pathways, regarding precursor biosynthesis (Collakova et al., 2012; Dandekar et al., 1999) and could be indicative of model complexity. Notwithstanding, 75-90.2 percent of reactions in ph-MeRecon were found in the three Arabidopsis models (Figure 3 A-C), and about 53 percent were found in the rice model (Figure 3D), indicating a highly conserved metabolic network (Peregrín-Alvarez et al., 2009). Taken together, approximately 43.2 percent of reactions in ph-MeRecon were shared by all four leaf models, representing 30 of 67 reactions in the respiration pathway, 7 of 23 reactions in carbohydrate metabolism, and 71 of 98 reactions in amino acid metabolism found in ph-MeRecon. In addition, the photosynthesis and photorespiration pathways of the models were highly conserved. The higher percentage of shared reactions between ph-MeRecon and Arabidopsis models (75-90.2%), compared with the rice model (53%) is probably due to evolutionary relatedness; cassava and Arabidopsis are both eudicots, whereas rice is a monocot (Considine et al., 2002; Iwamoto et al., 1998).

Key features of ph-MeRecon to mimic carbon metabolism in cassava leaves

We proposed the first compartmentalized metabolic model of primary carbon metabolism in cassava leaves, ph-MeRecon. The metabolic information incorporated into the model enabled a qualitative representation of the entire primary carbon metabolism in cassava leaves, underlying physiological behavior. Specifically, the model represents metabolic processes reported for cassava, such as (1) photosynthesis light-dependent reaction in terms of photosynthetic electron transport and ATP synthesis, (2) carboxylation and oxygenation of RuBisCO enzyme, (3) role of phosphoenolpyruvate carboxylase (PEPC) enzyme in cassava metabolism, and (4) starch biosynthesis and degradation in leaf carbon metabolism (Figure 4).

Photosynthesis light-dependent reaction

Photosynthesis is one of the main metabolic processes in leaf tissues and is basically composed of two parts: light dependent and light independent reactions (carbon fixation process). Not all of the available metabolic models of leaves could represent the complete photosynthesis processes (e.g. metabolic model of Arabidopsis (de Oliveira Dal'Molin et al., 2010a) and rice leaf cell (Lakshmanan et al., 2013)), but ph-MeRecon does. The ph-MeRecon was reconstructed by taking into account that the light-dependent reactions are critical to photosynthesis.

Table 1. Comparison of model characteristics of ph-MeRecon and the existing metabolic models of plant leaves. Cellular compartments: cytosol (c), mitochondria (m), chloroplast (p), peroxisome (x), vacuole (v).

Model	Scope of network	Cellular compartments					No. of genes	No. of reactions	No. of metabolites	No. of transport and exchange reactions
		c	m	p	x	v				
ph-MeRecon	Primary metabolism	●	●	●	●		1,037	550	461	149
AraGEM (de Oliveira Dal'Molin et al., 2010a)	Primary metabolism	●	●	●	●	●	1,419	1,567	1,748	99
*Arabidopsis i*RS1597 model (Saha et al., 2011)	Primary metabolism	●	●	●	●	●	1,597	1,798	1,820	99
Arabidopsis leaf model (Beckers et al., 2016)	Primary metabolism	●	●	●	●		1,567	511	472	46
Rice leaf model (Lakshmanan et al., 2013)	Primary metabolism	●	●	●			1,096	326	371	91
Tomato leaf model (Yuan et al., 2016)	Primary and secondary metabolism	●	●	●	●	●	3,410	1,885	1,998	258

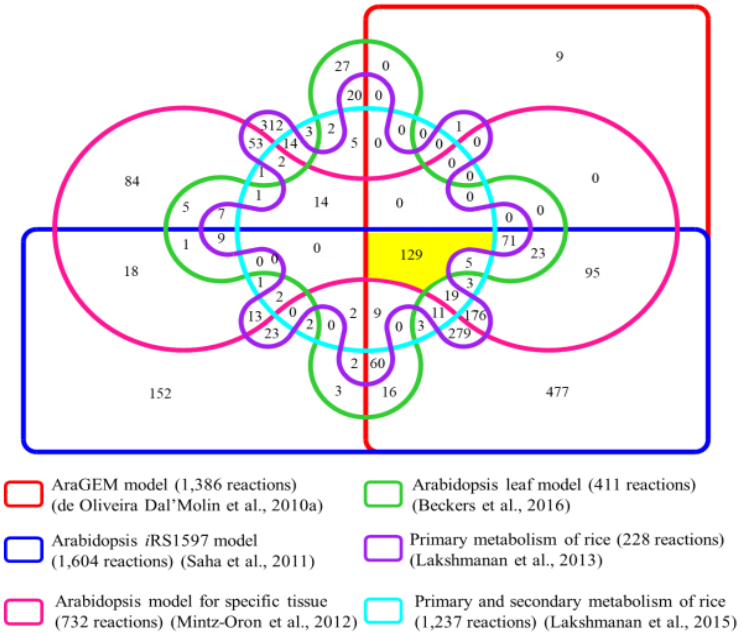
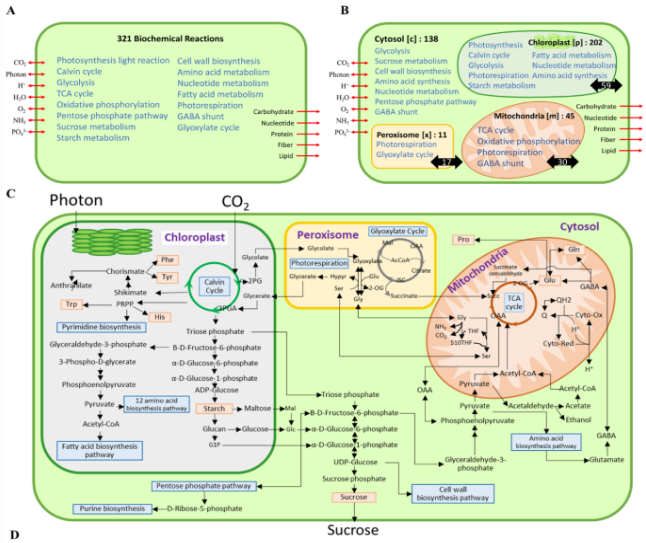


Fig 1. A Venn diagram illustrating the metabolic reactions found in the six plant-leaf models. The intersection of all six datasets (highlighted in yellow) denotes the common metabolic reactions (129) of the photosynthetic system.



Model	Number of entities	Enzymatic reactions					Spontaneous reactions					Transport reactions[a]					Total
		c	p	m	x	total	c	p	m	x	total	c	p	m	x	total	
ph-MeRecon-d1	Reactions					321					0						321
	Metabolites					282					0						282
ph-MeRecon-d2	Reactions	138	202	45	11	396					0	41	59	30	17	147	543
	Metabolites	157	218	64	22	461					0	41	59	30	17	147	461
ph-MeRecon	Reactions	139	203	46	11	399	1	1	0	0	2	41	61	30	17	149	550
	Metabolites	157	218	64	22	461	3	3	0	0	6	41	61	30	17	149	461

Fig 2. Schematics of the metabolic network in cassava leaves after each reconstruction step. (A) The major pathways of primary metabolism in first drafted metabolic network, ph-MeRecon-d1, are listed in blue. The red arrows represent the exchange reactions for nutrients and biomass components. (B) For ph-MeRecon-d2, the subcellular compartments and the number of constituent reactions are shown in bold. Numbers in black arrow show the transport reactions across membranes and cytosol. (C) For reconstructed ph-MeRecon, the sub-pathways in are represented in blue rectangles. Cell biomass components are highlighted in orange. (D) Characteristics of the ph-MeRecon reconstructive phase and the distribution of reactions and metabolites across compartments: (c) cytosol, (p) chloroplast, (m) mitochondria, and (x) peroxisome. [a]Transport across plasma membrane, chloroplast envelope membrane, mitochondrial membrane, and peroxisome membrane.

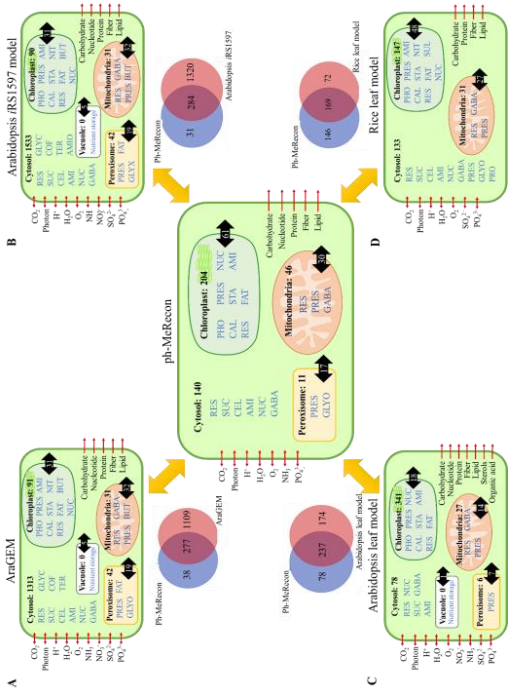


Fig 3. Differences in pathways and biochemical reactions between ph-MeRecon and published leaf models. The comparison was performed based on number of KEGG reaction IDs in the model structure between ph-MeRecon and: (A) AraGEM model (de Oliveira Dal'Molin et al., 2010a), (B) Arabidopsis *i*RS1597 model (Saha et al., 2011), (C) Arabidopsis leaf metabolic model (Beckers et al., 2016), and (D) rice leaf cell (Lakshmanan et al., 2013). Numbers in the Venn diagrams denote the biochemical reactions of ph-MeRecon and the other leaf models. Numbers in black arrow show the transport reactions across membranes and cytosol. The red arrows represent the exchange reactions. AMI: Amino acid metabolism, AMIO: Metabolism of other amino acids, BUT: Butanoate metabolism, CAL: Calvin cycle (carbon fixation), CEL: Cell wall metabolism, COF: Cofactors and vitamin metabolism, FAT: Fatty acid metabolism, GABA: GABA shunt pathway, GLYC: Glycan metabolism, GLYO: Glyoxylate and dicarboxylate metabolism, NIT: Nitrogen metabolism, NUC: Nucleotide metabolism, PHO: Photosynthesis light reactions, PRES: Photorespiration pathway, PRO: Propanoate metabolism, RES: Respiration pathway, SEC: Secondary metabolism, STA: Starch metabolism, SUC: Sucrose metabolism, SUL: Sulfur metabolism, TER: Metabolism of terpenoids and polyketides.

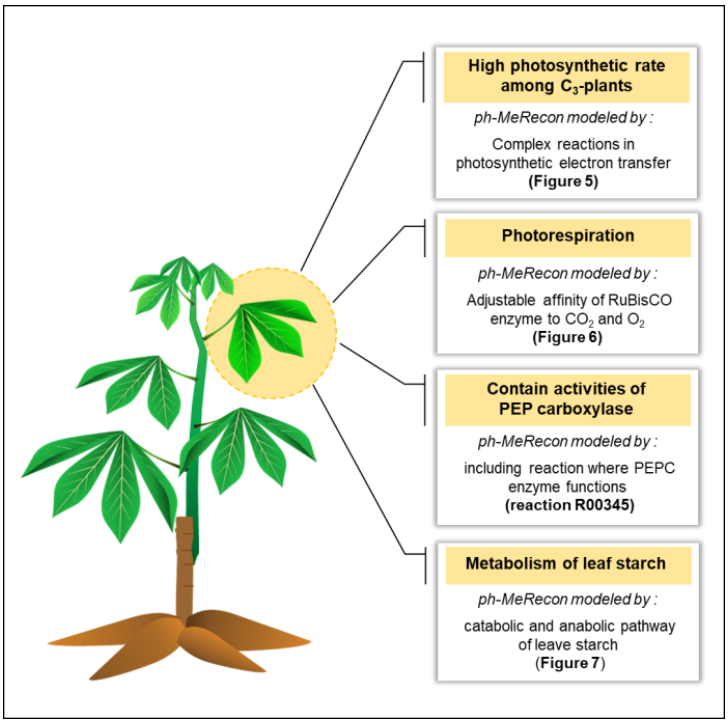
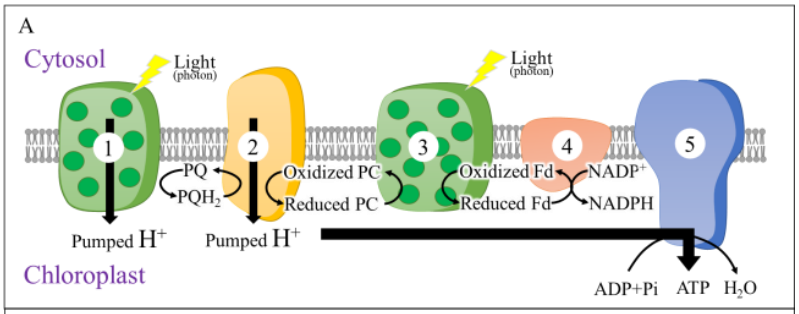


Fig 4. Key modelling features of the carbon metabolism in cassava leaves, ph-MeRecon.



B

(1) R09503, H_2O:plastoquinone reductase (PSII), EC 1.10.3.9;
$2\ H_2O + 2$ plastoquinone + 4 photon <=> $O_2 + 2$ plastoquinol + 4 pumped H^+

(2) R03817, plastoquinol:oxidized-plastocyanin oxidoreductase (Cyt b_6f), EC 1.10.9.1;
plastoquinol + 2 oxidized plastocyanin + 2 H^+ <=> plastoquinone + 2 reduced plastocyanin + 4 pumped H^+

(3) R09542, plastocyanin:ferredoxin oxidoreductase (PSI), EC 1.97.1.12;
reduced plastocyanin + oxidized ferredoxin + photon <=> oxidized plastocyanin + reduced ferredoxin

(4) R01195, Ferredoxin:NADP+ reductase (FNR), EC 1.18.1.2;
2 reduced ferredoxin + $NADP^+ + H^+$ <=> 2 oxidized ferredoxin + NADPH

(5) R00086, ATP synthase, EC 3.6.3.14;
ADP + Orthophosphate + 4 pumped H+ <=> ATP + H_2O + 3 H^+

(6) Reaction ID 9, ferredoxin-plastoquinone reductase (FQR):
2 reduced ferredoxin + plastoquinone -> 2 oxidized ferredoxin + plastoquinol

(7) Reaction ID 10, NAD(P)H dehydrogenase, EC 1.6.5.2 ;
plastoquinone + NADH + H^+ <=> plastoquinol + NAD^+

Fig 5. Photosynthetic light-dependent reactions as modelled by ph-MeRecon. (A) Scheme of linear photophosphorylation in the metabolic network reconstruction of cassava leaves. The numbers refer to individual protein complexes in (B). (B) List of model reactions involved in linear chain, cyclic photophosphorylation, and ATP synthase. PQ, Plastoquinone; PQH_2, Plastoquinol; PC, Plastocyanin; Fd, Ferredoxin.

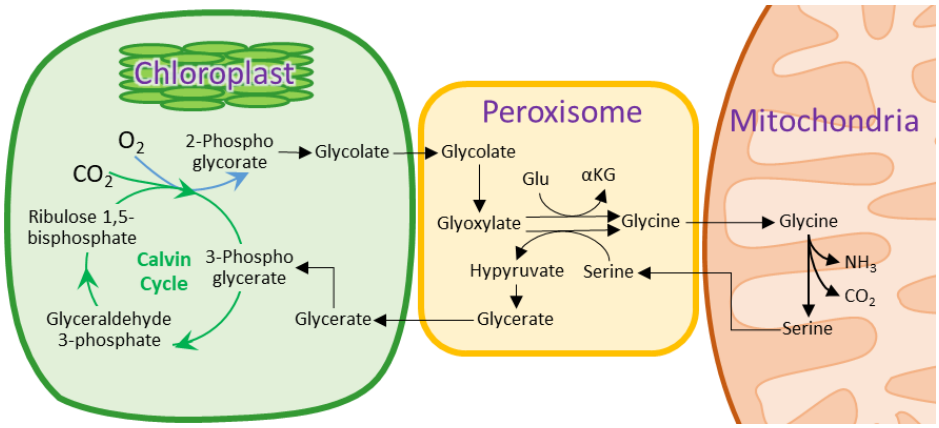
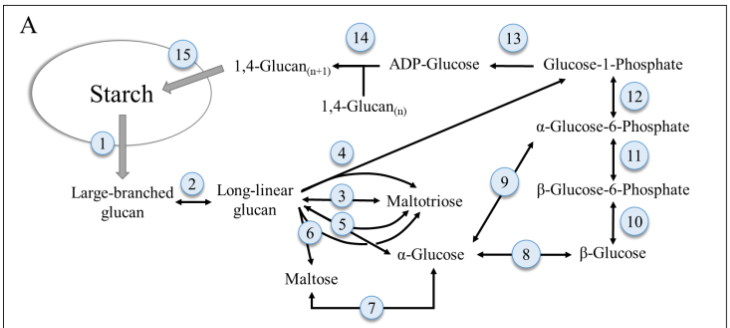


Fig 6. Photorespiration pathway in the ph-MeRecon. The oxygenation and carboxylation activities of RuBisCO is represented in blue and green lines, respectively. The pathway comprises three compartments: chloroplast, peroxisome, and mitochondria.



B

(1) α-amylase, EC 3.2.1.1;
Starch + H_2O <=> Large-branched glucan
(2) pullulanase, EC 3.2.1.41;
Large-branched glucan + 2 H_2O <=> 2 Long-linear glucan
(3) α-amylase, EC 3.2.1.1;
Long-linear glucan + H_2O <=> 2 Maltotriose
(4) glycogen phosphorylase, EC 2.4.1.1;
Long-linear glucan + Orthophosphate -> Maltotriose + Glucose-1-Phosphate
(5) 4-α-glucanotransferase, EC 2.4.1.25;
Long-linear glucan + alpha-Glucose <=> 2 Maltotriose
(6) β-amylase, EC 3.2.1.2;
Long-linear glucan + H_2O <=> Maltotriose + Maltose
(7) α-glucosidase, EC 3.2.1.20;
Maltose + H_2O <=> 2 alpha-Glucose
(8) aldose 1-epimerase, EC 5.1.3.3;
alpha-Glucose <=> beta-Glucose
(9) hexokinase, EC 2.7.1.1;
ATP + alpha-Glucose <=> ADP + alpha-Glucose-6-Phosphate
(10) glucokinase, EC 2.7.1.2;
ATP + beta-Glucose <=> ADP + beta-Glucose-6-Phosphate
(11) glucose-6-phosphate 1-epimerase, EC 5.1.3.15;
alpha-Glucose-6-Phosphate <=> beta-Glucose-6-Phosphate
(12) phosphoglucomutase, EC 5.4.2.2;
Glucose-1-Phosphate <=> alpha-Glucose-6-Phosphate
(13) glucose-1-phosphate adenylyltransferase, EC 2.7.7.27;
ATP + Glucose-1-Phosphate -> Diphosphate + ADP-Glucose
(14) starch synthase (glycosyl-transferring), EC 2.4.1.21;
1,4-Glucan(n) + ADP-Glucose <=> 1,4-Glucan(n+1) + ADP
(15) 1,4-α-glucan branching enzyme, EC 2.4.1.18;
1,4-Glucan(n+1) <=> Starch

Fig 7. Starch metabolism in the ph-MeRecon model. (A) scheme of starch metabolism in the ph-MeRecon model. The numbers refer to consecutive reactions in (B). (B) List of biochemical reactions, enzyme names, and corresponding EC number.

The light-dependent reactions occur once pigment molecules in leaves absorb photon (hv) of visible light. The excited electrons are transferred through a series of integral membrane protein complexes, resulting in photosynthetic electron transport reactions that yield energy for carbon fixation process. In higher plants, photosynthetic electron transport involves both a linear electron flow (LEF) and a cyclic electron flow (CEF). LEF requires photosystem II (PSII), Cytochrome b_6f complex (Cyt b_6f), Photosystem I (PSI), and ferredoxin-NADP+ reductase (FNR). CEF proceeds either via

(1) the NAD(P)H-dehydrogenase (NDH)-dependent pathway, or (2) ferredoxin-dependent pathway, employing ferredoxin-plastoquinone reductase (FQR), involved in the PGR5-PGRL1 protein complex (Johnson, 2011; Suorsa et al., 2016). Although there is no extensive study about cyclic electron transfer system in cassava to-date, homologous sequences associated to cyclic electron transfer were identified in cassava genome by comparative analysis with PGR5/PGRL1 and NAD(P)H-dehydrogenase genes of Arabidopsis and rice (Suorsa et al., 2016; Yamori et al., 2016). The existence of the reactions and pathway in cassava leaves was also supported by the expression evidence of genes related to PGR5/PGR1 and NDH-dependent pathways. The light-dependent reactions employed in ph-MeRecon were based on relevant metabolic reactions obtained from published plant metabolic models (Beckers et al., 2016; Simons et al., 2014; Yuan et al., 2016) and metabolic omnibus (i.e. KEGG and PMN), as shown in Figure 5. To represent the electron transfer process in ph-MeRecon (Figure 5A), individual protein-complexes were assumed to be biochemical reactions (Figure 5B) that occur inside chloroplast. Protons (H^+) that are pumped from chloroplast stroma to thylakoid lumen during electron transportation were denoted as pumped H^+ molecules within chloroplast. ATP synthase complex reaction was assumed to generate ATP from the pumped H^+. Reactions and corresponding genes of the modelled photosynthetic electron transfer are listed in Supplementary Table 5.

Carboxylation and oxygenation of RuBisCO enzyme

The ph-MeRecon was reconstructed to reflect the substrate selection of RuBisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase; KEGG reaction ID R00024, EC number 4.1.1.39) in leaves of C_3 plants. Enzyme RuBisCO is able to catalyze RuBP through both the carboxylation (CO_2 fixation) and the oxygenation (O_2 fixation) processes (Taiz and Zeiger, 2002). The kinetic parameters of RuBisCO in cassava is reportedly similar to C_3 species (El-Sharkawy, 2003), but there is no information on the carboxylation-to-oxygenation ($V_C/V_{O)}$ ratio for cassava to date. To present the physiological function of RuBisCO under normal condition, the reaction was modeled using a V_C/V_O ratio of 3:1, based on the published rice model (Lakshmanan et al., 2013). The oxygenation and carboxylation functions of enzyme RuBisCO in the ph-MeRecon was represented as:

$$4\ RuBP\ +\ O_2\ +\ 3\ CO_2\ +\ 3\ H_2O\ \rightarrow\ 2PG\ +\ 7\ G3P$$

In addition, the ph-MeRecon contains the photorespiration pathway essential for recycling 2-Phosphoglycolate, a product of the oxygenation of RuBP by RuBisCO, to the Calvin cycle (Taiz and Zeiger, 2002). In general, the photorespiration process operates across three compartments: chloroplast, peroxisome, and mitochondria. The relevant genes for the photorespiratory process in cassava were identified and incorporated into the ph-MeRecon pathway **(**Figure 6).

Role of phosphoenolpyruvate carboxylase (PEPC) enzyme in cassava metabolism

The role of phosphoenolpyruvate carboxylase (PEPC) enzyme in cassava metabolism was also captured in ph-MeRecon. Phosphoenolpyruvate carboxylase enzyme (PEPC; E.C. 4.1.1.31) is recognized as a key enzyme in C_4 plant species. Unusually high PEP carboxylase activity was found in cassava (about 10-25% of C_4-PEPC), although it lacks the leaf Kranz anatomy typical for C_4 plants (El-Sharkawy, 2003). The high expression of the genes of PEPC enzyme in cassava, under different growth conditions, supports its function in cassava metabolism (El-Sharkawy et al., 2008). Accordingly, the R00345 reaction was added to the cytosolic pathway of ph-MeRecon to represent its C_3 and C_4 physiology.

$$Pi\ +\ OAA\ \leftrightarrow\ PEP\ +\ {HCO_3}^-$$

Starch biosynthesis and degradation in leaf carbon metabolism

The ph-MeRecon covered the metabolism of starch anabolism and catabolism that generally occur in leaf tissue. During day time, transitory starch is synthesized as a product of photosynthesis and is in turn degraded to sucrose, which is allocated to sink tissues at night (Kötting et al., 2010; Orzechowski, 2008). Starch metabolism in leaves have been extensively studied in Arabidopsis (Streb and Zeeman, 2012), but to a limited extent in cassava (Rongsirikul et al., 2010; Saithong et al., 2013). To incorporate the information on starch metabolism in cassava leaves, the ph-MeRecon was reconstructed using existing metabolic reactions and associated genes (Kötting et al., 2010; Rongsirikul et al., 2010; Saithong et al., 2013; Zeeman et al., 2010). The starch and sucrose metabolism in ph-MeRecon (Figure 7) represent a more realistic pathway of starch metabolism in cassava leaves; the reactions and corresponding genes are listed in Supplementary Table 6.

In conclusion, steps such as network compartmentalization, reaction stoichiometric balance, and network connectivity analysis are crucial to the reconstruction of a high-quality genome-scale metabolic network (Thiele and Palsson, 2010). The proposed metabolic pathway model of carbon metabolism in cassava leaves — ph-MeRecon, the first of its kind, is robust and better represents carbon metabolism in cassava than hitherto existing models. Compared with the metabolic pathways of cassava available in the PMN public database (Schlapfer et al., 2017), ph-MeRecon is compartmentalized, fully connected, and took into account information on the metabolic pathways in leaves. In addition, reactions in ph-MeRecon were validated using RNA expression data. Given its robustness, ph-MeRecon is a promising tool for the quantitative modeling of carbon metabolism in cassava leaves at different growth stage and environment. It offers a computation-ready platform for modeling the dynamics in plant metabolism.

Materials and methods

Data resources

The ph-MeRecon was constructed by integrating the genomic information and biochemical reactions of cassava, Arabidopsis, and rice into the genome-based pathway of primary metabolism in cassava, MeRecon (Siriwat, 2012). The genomic information of cassava and rice were taken from Phytozome database (*Manihot esculenta* protein annotation version 6.1 and *Oryza sativa* protein annotation version 7.0, https://phytozome.jgi.doe.gov/pz/ portal.html), and the genomic information of Arabidopsis was obtained

from The Arabidopsis Information Resource (TAIR version 10, https://www.arabidopsis.org/). Data on the biochemical reactions involved in photosynthetic carbon metabolism and associated genes were obtained from the six leaf models — four models of Arabidopsis (Beckers et al., 2016; de Oliveira Dal'Molin, et al., 2010a; Mintz-Oron et al., 2012; Saha et al., 2011) and two models of rice (Lakshmanan et al., 2015; Lakshmanan et al., 2013). The missing associated genes of enzymes were obtained from the metabolic databases, Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg), and Plant metabolic network (PMN, https://www.plantcyc.org/). The stoichiometry and reversibility potential of each reaction was inferred from KEGG, BRENDA (https://www.brenda-enzymes.org/), and MetaCyc (https://metacyc.org/) databases. For the transcriptomic data, RNA expression of 3-month-old leaf tissues of cultivar TME204 (NCBI Accession No. GSE82279) was obtained from Wilson et al. (2017).

Pathway reconstruction

The ph-MeRecon was initially reconstructed using genomic data, including cassava-enzymatic genes, Enzyme Commission (EC) numbers, and metabolic reactions proposed in the genome-based pathway of carbon metabolism in cassava, MeRecon (Siriwat, 2012) (Supplementary Figure 1). Then, the common metabolic reactions in the six leaf models were added. In addition, some overlapped reactions, in at least two leaf models, were added to fully connect the common reactions to MeRecon. Subsequently, the metabolic genes relevant to common metabolic reactions were used to identify the responsible homologous enzymatic genes in cassava genome, to ensure their availability in cassava metabolism. The enzymatic protein sequences were obtained from published metabolic models of Arabidopsis and rice, when available. Otherwise, the missing protein sequences were obtained from metabolic database omnibus; KEGG and PMN. To identify the homologous enzymatic genes of cassava, enzymatic protein sequences were compared against the cassava protein library through bidirectional BLAST (Bredeson et al., 2016) using the following criteria: (1) E-value $\leq 1x10^{-10}$, (2) identity percentage ≥ 60, and (3) coverage percentage ≥ 80. The common reactions in leaf metabolism were, then, integrated into the pre-existing genome-based pathway, i.e. MeRecon, to construct the first draft of ph-MeRecon — ph-MeRecon-d1. Afterwards, ph-MeRecon-d1 was verified using cassava transcriptome data (NCBI Accession No. GSE82279) (Wilson et al., 2017). The paired-end reads were mapped to the cassava reference genome version 6.1 provided on the Phytozome database using the STAR aligner (Dobin et al., 2013), and the expression level was calculated using the Cufflinks tool (Trapnell et al., 2010).

Pathway compartmentalization

Subcellular location(s) of enzymes was predicted based on the signal peptide region of the enzymatic protein sequences. The prediction was performed using the iLoc-Plant classifier (Wu et al., 2011). The biochemical reactions in ph-MeRecon-d1 were subsequently compartmentalized into cytosol [c], chloroplast [p], mitochondria [m] and peroxisome [x] resulting in the second draft network, ph-MeRecon-d2. All reactions assigned to subcellular compartments were curated with the published metabolic models of the representative plants (de Oliveira Dal'Molin et al., 2010a; Hay and Schwender, 2011; Lakshmanan et al., 2015; Lakshmanan et al., 2013; Mintz-Oron et al., 2012; Saha et al., 2011; Yuan et al., 2016) and the existing evidence in literature (Buchanan et al., 2015). Intracellular transport reactions were assigned to metabolites that are required to move between compartments, based on the metabolic networks of the representative plants (de Oliveira Dal'Molin et al., 2010a; Hay and Schwender, 2011; Lakshmanan et al., 2015; Lakshmanan et al., 2013; Mintz-Oron et al., 2012; Saha et al., 2011; Yuan et al., 2016) and literature (Buchanan et al., 2015).

Pathway curation

Following the compartmentalization, the reconstructed pathway, ph-MeRecon-d2, underwent connectivity analysis using detectDeadEnds and gapFind functions in the COBRA toolbox v2.0, MATLAB (version 7.13; MathWorks, Inc.) (Schellenberger et al., 2011). Gap metabolites that could not be produced or consumed in the metabolic network were identified and used to predict the gap reactions based upon metabolic networks of the published plant models (Beckers et al., 2016; de Oliveira Dal'Molin et al., 2010a; Lakshmanan et al., 2015; Lakshmanan et al., 2013; Mintz-Oron et al., 2012; Pilalis et al., 2011; Saha et al., 2011; Yuan et al., 2016) and available information in public databases (e.g. KEGG and PMN). The manual curation and network refinement were iteratively performed until no gap was found. In addition, each reaction in the compartmentalized metabolic network was manually curated for reversibility, stoichiometry, and mass balance. The reversibility potential of each reaction was defined primarily based on Buchanan et al. (2015) and metabolic network omnibus (e.g. KEGG, BRENDA, and PMN). For reaction stoichiometry, the data was inferred from KEGG and MetaCyc databases. Finally, the pathway atomic balance in relation to hydrogen, carbon, oxygen, phosphorus, sulfur, and nitrogen, was checked and corrected using checkMassChargeBalance function in COBRA toolbox v3.0 on MATLAB platform.

Conclusion

Metabolic network offers a simplified view of complex biological processes occurring at the system level. It also provides a highly mathematical, structured platform that enables the understanding of molecular mechanisms that underlie the physiology of organisms. The ph-MeRecon, a metabolic network of primary carbon metabolism in cassava photosynthetic tissues, was constructed by incorporating genomic data, physiological data and biochemical reactions of cassava, Arabidopsis, and rice into the genome-based pathway of primary metabolism in cassava. The reconstructed model consists of 461 metabolites and 550 reactions distributed in four cellular compartments: cytosol, chloroplast, mitochondria, and peroxisome. Through comparative genomics approach, 1,037 cassava genes

(3.14% of total protein-coding genes in the genome) were identified as enzymatic genes that function in primary carbon metabolism. More than 95 percent of entire enzymatic genes were validated using RNA-expression data of cassava leaf tissues. The reconstructed pathway well-illustrates the physiological function of carbon metabolism in cassava leaves and provides a platform for omics data integration and systematic analysis of properties and capabilities of metabolic networks. Ultimately, it may facilitate crop improvement by indicating strategies for enhancing the photosynthetic capacity of cassava crop.

Acknowledgments

We would like to thank Ms. Rungaroon Suksamran for her assistance on transcriptome data analysis and Ms. Porntip Cheiwchankaset for her valuable suggestion. NP's studentship was supported by the Petch Pra Jom Klao Master's Degree Research Scholarship from King Mongkut's University of Technology Thonburi (KMUTT). The authors acknowledge the financial support provided by National Science and Technology Development Agency (NSTDA) and National Research Council of Thailand (NRCT) (P-13-50395), KMUTT under the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission, Ministry of Education, Thailand, and KMUTT through the "KMUTT 55th Anniversary Commemorative Fund".

References

Beckers V, Dersch LM, Lotz K, Melzer G, Bläsing OE, Fuchs R, Ehrhardt T, Wittmann C (2016) In silico metabolic network analysis of Arabidopsis leaves. BMC Syst Biol. 10(1):102.

Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nat Biotechnol. 34(5):562.

Buchanan BB, Gruissem W, Jones RL (eds) (2015) Biochemistry and molecular biology of plants, 2nd ed. John Wiley & Sons, West Sussex, UK.

Collakova E, Yen JY, Senger RS (2012) Are we ready for genome-scale modeling in plants?. Plant Sci. 191:53-70.

Considine MJ, Holtzapffel RC, Day DA, Whelan J, Millar AH (2002) Molecular distinction between alternative oxidase from monocots and dicots. Plant Physiol. 129(3):949-953.

Dandekar T, Schuster S, Berend SN, Huynen M, Peer BO (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem J. 343(1):115-124.

de Oliveira Dal'Molin CG, Nielsen LK (2013) Plant genome-scale metabolic reconstruction and modelling. Curr Opin Biotech. 24(2):271-277.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010a) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol. 152(2):579-589.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010b) C4GEM-Genome-Scale Metabolic Model to study C4 plant metabolism. Plant Physiol. 154:1871-1885.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15-21.

El-Sharkawy MA (2003) Cassava biology and physiology. Plant Mol Biol. 53(5):621-641.

El-Sharkawy MA, Lopez Y, Bernal LM. (2008) Genotypic variations in activities of phosphoenolpyruvate carboxylase and correlations with leaf photosynthetic characteristics and crop productivity of cassava grown in low-land seasonally-dry tropics. Photosynthetica. 46(2):238.

Hay J, Schwender J (2011) Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (*Brassica napus* L.) embryos. Plant J. 67(3):526-541.

Howeler R, Lutaladio N, Thomas G (2013) Save and grow:cassava. A guide to sustainable production intensification. FAO, Rome.

Iwamoto M, Maekawa M, Saito A, Higo H, Higo K (1998) Evolutionary relationship of plant catalase genes inferred from exon-intron structures:isozyme divergence after the separation of monocots and dicots. Theor Appl Genet. 97(1-2):9-19.

Johnson GN (2011) Physiology of PSI cyclic electron transport in higher plants. BBA-Bioenergetics. 1807(3):384-389.

Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. Curr Opin Biotech. 23(4):617-623.

Kötting O, Kossmann J, Zeeman SC, Lloyd JR (2010) Regulation of starch metabolism: the age of enlightenment?. Curr Opin Plant Biol. 13(3):320-328.

Lakshmanan M, Cheung CY, Mohanty B, Lee DY (2016) Modeling rice metabolism: from elucidating environmental effects on cellular phenotype to guiding crop improvement. Front Plant Sci. 7:1795.

Lakshmanan M, Lim SH, Mohanty B, Kim JK, Ha SH, Lee DL (2015) Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multi-omics analysis. Plant Physiol. 169:3002–3020.

Lakshmanan M, Zhang Z, Mohanty B, Kwon JY, Choi HY, Nam HJ, Kim DI, Lee DY (2013) Elucidating the rice cells metabolism under flooding and drought stresses using flux-based modelling and analysis. Plant Physiol. 162(4):2140-2150.

Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. P Natl Acad Sci USA. 109(1):339-344.

Montagnac JA, Davis CR, Tanumihardjo SA (2009) Nutritional value of cassava for use as a staple food and recent advances for improvement. Compr Rev Food Sci F. 8(3):181-194.

Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol. 5(1):320.

Orzechowski S. (2008) Starch metabolism in leaves. Acta Biochim Pol. 55(3):435-445.

Peregrín-Alvarez JM, Sanford C, Parkinson J (2009) The conservation and evolutionary modularity of metabolism. Genome Biol. 10(6):R63.

Pilalis E, Chatziioannou A, Thomasset B, Kolisis F (2011) An *in silico* compartmentalized metabolic model of *Brassica napus*

enables the systemic study of regulatory aspects of plant central metabolism. Biotechnol Bioeng. 108(7):1673-1682.

Poolman MG, Kundu S, Shaw R, Fell D. (2013) Responses to light intensity in a genome-scale model of rice metabolism. Plant Physiol. 162:1060-1072.

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol. 151(3):1570-1581.

Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform. 10(4):435-449.

Rongsirikul O. et al. (2010) Reconstruction of starch biosynthesis pathway in cassava using comparative genomic approach. In: Chan J H, Ong YS, Cho SB. (eds) Computational Systems-Biology and Bioinformatics. CSBio 2010. Comm Com Inf Sc. vol 115. Springer, Berlin, Heidelberg. 118-129.

Rossi M, Bermudez L, Carrari F (2015) Crop yield: challenges from a metabolic perspective. Curr Opin Plant Biol. 25:79-89.

Saha R, Suthers PF, Maranas CD (2011) *Zea mays i*RS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PloS ONE. 6(7):e21784.

Saithong T, Rongsirikul O, Kalapanulak S, Chiewchankaset P, Siriwat W, Netrphan S, Suksangpanomrung M, Meechai A, Cheevadhanarak S (2013) Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. BMC Syst Biol. 7(1):75.

Sakurai T, Plata G, Rodríguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M (2007) Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. BMC Plant Biol. 7(1):66.

Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 6(9):1290.

Schlapfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, Dreher K, Chavali AK, Nilo-Poyanco R, Bernard T, Kahn D. (2017) Genome-wide prediction of metabolic enzymes, pathways and gene clusters in plants. Plant Physiol. 173(4):2041-2059.

Shaw R, Kundu S (2015) Flux balance analysis of genome-scale metabolic model of rice (*Oryza sativa*): Aiming to increase biomass. J Biosciences. 40(4):819-28.

Simons MN, Saha R, Amiour N, Kumar A, Guillard L, Clément G, Miquel M, Li Z, Mouille G, Lea PJ, Hirel B (2014) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol. 166(3):1659-1674.

Siriwat W (2012) Reconstruction of the key pathway in carbon assimilation in cassava roots. (Master of Science in Bioinformatics and Systems Biology program), King Mongkut's University of Technology Thonburi.

Streb S, Zeeman SC (2012) Starch metabolism in Arabidopsis. The Arabidopsis book, Aspp Symp P. 10:e0160.

Suorsa M, Rossi F, Tadini L, Colombo M, Jahns P, Kater MM, Leister D, Finazzi G, Aro EM, Barbato R, Pesaresi P (2016) PGR5-PGRL1-dependent cyclic electron transport modulates linear electron transport rate in *Arabidopsis thaliana*. Mol Plant. 9(2):271-288.

Taiz L, Zeiger E (2002) Plant Physiology, 3rd ed. Sinauer Associates, Massachusetts, USA.

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 5(1):93.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotech. 28(5):511.

Wilson MC, Mutka AM, Hummel AW, Berry J, Chauhan RD, Vijayaraghavan A, Taylor NJ, Voytas DF, Chitwood DH, Bart RS (2017) Gene expression atlas for the food security crop cassava. New Phytol. 213(4):1632-1641.

Wu ZC, Xiao X, Chou KC (2011) iLoc-Plant: a multi-label classifier for predicting the subcellular localization of plant proteins with both single and multiple sites. Mol Biosyst. 7(12):3287-3297.

Yamori W, Makino A, Shikanai T (2016) A physiological role of cyclic electron transport around photosystem I in sustaining photosynthesis under fluctuating light in rice. Sci Rep-UK. 6:20147.

Yuan H, Cheung CM, Poolman MG, Hilbers PA, van Riel NA (2016) A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. Plant J. 85(2):289-304.

Zeeman SC, Kossmann J, Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol. 61:209-234.