Running Head: A maize leaf model applied to nitrogen metabolism

Research Area: Biochemistry and Metabolism

Corresponding Author:
Costas D. Maranas
Department of Chemical Engineering, The Pennsylvania State University, University Park, PA
Phone: 814-863-9958
Fax: 814-865-7846
Email: costas@psu.edu
Assessing the Metabolic Impact of Nitrogen Availability using a Compartmentalized Maize Leaf Genome-Scale Model

Margaret Simons¹a, Rajib Saha¹a, Nardjis Amiour², Akhil Kumar³, Lénaïg Guillard², Gilles Clément², Martine Miquel², Zheni Li², Gregory Mouille², Peter J. Lea⁴, Bertrand Hirel², and Costas D. Maranas¹*

¹Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
²Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique (INRA), Centre de Versailles-Grignon, Unité Mixte de Recherche 1318 INRA-Agro-ParisTech, Equipe de Recherche Labellisée (ERL), Centre National de la Recherche Scientifique (CNRS) 3559, RD10, F-78026 Versailles Cedex, France
³Bioinformatics and Genomics, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA
⁴Lancaster Environment Centre. Lancaster University, Lancaster LA1 4YQ, U.K.
One-Sentence Summary
A cell-type and leaf tissue-specific model provides new insights into nitrogen metabolism in the maize leaf.

Footnotes:
This work was supported by the U.S. Department of Energy, grant no. DE-FG02-05ER25684 and by the Agence National pour la Recherche (ANR)-Genoplante program Maize and Yield (GNP05015G).

*These authors contributed equally to this work.

*Corresponding Author: Costas D. Maranas; email: costas@psu.edu
Abstract

Zea mays L. is an important C₄ plant due to its widespread use as a cereal and energy crop. A second-generation genome-scale metabolic model for the maize leaf was created to capture C₄ carbon fixation and investigate nitrogen (N) assimilation by modeling the interactions between the bundle sheath and mesophyll cells. The model contains gene-protein-reaction relationships, elemental and charge balanced reactions, and incorporates experimental evidence pertaining to the biomass composition, compartmentalization, and flux constraints. Condition-specific biomass descriptions were introduced that account for amino acids, fatty acids, soluble sugars, proteins, chlorophyll, lignocellulose, and nucleic acids as experimentally measured biomass constituents. Compartmentalization of the model is based on proteomic/transcriptomic data and literature evidence. With the incorporation of information from the MetaCrop and MaizeCyc databases, this updated model spans 5,824 genes, 8,525 reactions, and 9,153 metabolites, an increase of approximately four times the size of the earlier iRS1563 model. Transcriptomic and proteomic data has also been used to introduce regulatory constraints in the model to simulate a N limited condition and mutants deficient in glutamine synthetase, gln1-3 and gln1-4. In silico results achieved 90% accuracy when comparing the wild-type (WT) grown under an N complete condition (N⁺ WT) with the WT grown under an N deficient (N⁻ WT) condition.
Introduction

_Zea mays_ L., commonly known as maize or corn, is an essential dual-use food and energy crop. Maize production is increasing at the greatest rate among all cereals with a worldwide trend of 0.06 t ha\(^{-1}\) year\(^{-1}\) (tons hectare\(^{-1}\) year\(^{-1}\)) (Leveau et al., 2011), and a record 877 million tons produced in the 2011-2012 fiscal year (International Grains Council, 2013). With the recent completion of the maize genome in 2009 along with the creation and curation of databases such as MaizeGDB in 2011 (Schaeffer et al., 2011), MaizeCyc in 2013 (Monaco et al., 2013), and MetaCrop 2.0 in 2012 (Schreiber et al., 2012), there is a need for an updated genome-scale metabolic (GSM) model (Saha et al., 2011) that will integrate all newly available information from diverse sources. The integration of this information with experimental transcriptomic data, proteomic data, and biomass composition measurements obtained with wild-type plants grown under optimal nitrogen (N\(^{+}\) WT) conditions and limited nitrogen (N\(^{-}\) WT) conditions (Amiour et al., 2012), as well as two glutamine synthetase (GS) mutants grown under optimal nitrogen (N), _gln1-3_ and _gln1-4_ mutants (Martin et al., 2006), has provided a more accurate assessment of N metabolism within the maize leaf. Moreover, since integration of transcriptome, proteome, and metabolome data appeared not to be straightforward (Amiour et al., 2012; 2014), the development of a model could help to identify putative candidate genes, proteins, and metabolic pathways contributing to plant growth and development.

Maize is a C\(_{4}\) plant that overcomes the inefficiencies of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), to capture O\(_{2}\) over the preferred CO\(_{2}\), by separating the photosynthetic carbon (C) fixation process into two cell types: the bundle sheath and mesophyll cells. In comparison to C\(_{3}\) plants, this separation allows C\(_{4}\) plants to have a lower rate of photorespiration, higher rate of photosynthesis at high light intensities (under standard air and temperature conditions), and a higher photosynthetic N use efficiency (NUE) (Christin and Osborne, 2013; Driever and Kromdijk, 2013; Peterhansel et al., 2013; Sage, 2014; Wang et al., 2014). A C\(_{4}\)-specific maize GSM could provide insight into N metabolism and provide cues for improving NUE (i.e. the vegetative...
biomass or grain yield produced per unit of N present in the soil). Since N is the major limiting factor in agricultural production among mineral fertilizers (Vitousek et al., 1997; Hirel et al., 2007; Andrews and Lea, 2013; Andrews et al., 2013) and NUE is estimated to be far below 50% in cereal grains (Raun and Johnson, 1999), improving NUE is essential for improving overall productivity in maize (Hirel and Gallais, 2011). Amiour et al. (2012) experimentally determined 150 gene transcripts, 40 proteins, and 89 metabolites that are significantly different between the N+ WT and N- WT conditions during the vegetative stage of growth. N utilization is strongly linked to the GS enzyme as all N, either in the form of nitrate or ammonium ions, is channeled through the reaction catalyzed by the GS enzyme (Martin et al., 2006; Cañas et al., 2010; Hirel and Gallais, 2011; Andrews et al., 2013). The mesophyll cell-specific GS1-3 isozyme is involved in synthesizing glutamine after nitrate reduction from the vegetative state until the plant reaches maturity. Leaf aging induces the synthesis of the bundle sheath-specific GS1-4 isozyme. Consequently, Martin et al. hypothesized that GS1-4 isoform is used in the reassimilation of ammonium during protein degradation in senescing leaves (Martin et al., 2006). During vegetative growth in the leaf tissue, DNA microarray data revealed that 243 gene transcripts, 46 proteins, and 48 metabolites exhibited significant differences in the _gln1-3_ mutants and 107 gene transcripts, 14 proteins, and 18 metabolites displayed substantial differences in the _gln1-4_ mutants (Amiour et al., 2014). In this second-generation maize model, we explore the effect of the computational knockout of genes encoding for GS1-3 and GS1-4 isozymes using flux balance analysis (FBA) to elucidate the role of GS in N metabolism.

FBA of GSMs is used to model organism-specific metabolism by simulating the internal flow of metabolites. The number of GSMs for plants has increased rapidly with models available for Arabidopsis (Poolman et al., 2009; de Oliveira Dal'Molin et al., 2010), barley seed (Grafahrend-Belau et al., 2009), maize (de Oliveira Dal'Molin et al., 2010; Saha et al., 2011), sorghum (de Oliveira Dal'Molin et al., 2010), sugarcane (de Oliveira Dal'Molin et al., 2010), rapeseed (Pilalis et al., 2011), and rice (Poolman et al., 2013). These models rely on annotation information to assemble comprehensive compilations of all reactions and metabolites known to occur within the organism. Currently, whole-
genome sequencing has been completed for approximately 40 vascular plants including *Arabidopsis thaliana* (The Arabidopsis Initiative, 2000), *Arabidopsis lyrata* (Hu et al., 2011), *Glycine max* (Schmutz et al., 2010), *Oryza sativa* (Goff et al., 2002; Yu et al., 2002), *Populus trichocarpa* (Tuskan et al., 2006), *Sorghum bicolor* (Paterson et al., 2009), *Theobroma cacao* (Tuskan et al., 2006), and *Zea mays* (Schnable et al., 2009). Gene annotations of the whole-genome sequences have been used to determine the reactions within an organism and therefore build a GSM. FBA calculates all reaction fluxes in a metabolic network based on the optimization of an objective function (typically the maximization of the biomass yield). A quasi-steady state is assumed and flux constraints are set based on the specific media or the reversibility of reactions derived from thermodynamics. Incorporation of “omics” data into GSMs is achieved through appropriate constraints on fluxes that restrict metabolic flows to only condition-relevant phenotypes.

During the last few years, multiple methods have been developed to integrate “omics” data into GSMs. Proteomic and transcriptomic data have been used to apply flux constraints on corresponding reactions determined by gene-protein-reaction (GPR) associations. The GIMME (Becker and Palsson, 2008), iMAT (Shlomi et al., 2008), and MADE (Jensen and Papin, 2011) algorithms use a “switch” approach to turn on/off reactions based on expression levels. The GIMME algorithm turns off reactions based on a user-specified threshold of the expression level. The iMAT algorithm turns on a minimal set of reactions associated with low expression data in order to achieve a user-specified metabolic function. The MADE algorithm incorporates related experimental datasets into the model to activate or repress reactions based on the progression of the experimental conditions. A different class of algorithms, known as the “valve” approach, was developed to incorporate proteomic and transcriptomic data by constraining the allowable flux ranges of reactions. The E-Flux method incorporates a user-specified function to convert gene expression data to flux constraints (Colijn et al., 2009). Finally, the PROM (Chandrasekaran and Price, 2010) algorithm uses multiple datasets to constrain flux bounds (i.e. allowable flux ranges) based on the probabilities associated with gene activity among all datasets. Lee et al. (2012) integrated gene expression data by
minimizing the difference between the predicted flux level and gene expression data over all reactions with corresponding expression levels. Using the Yeast 5 model (Heavner et al., 2012) for *Saccharomyces cerevisiae*, Lee et al. (2012) compared the predicted fluxes to experimentally determined exometabolome fluxes using the coefficient of determination $R^2$. The authors achieved an $R^2$ value of 0.87 and 0.96 under 75% and 85% of the maximal biomass level, respectively. In comparison, the authors generate a “best” FBA solution, which maximizes $R^2$ over all feasible solutions generated for FBA, and achieve an $R^2$ values of 0.20 and 0.58 under 75% and 85% of the maximal biomass level, respectively. These advancements pertaining to the integration of “omics” data with GSMs has enabled more accurate model predictions.

In this work, we describe the reconstruction of a second-generation maize leaf model and the incorporation of “omics” data into the model with the goal of improving the understanding of N metabolism. Both the primary and secondary metabolic pathways of maize are included, by combining information from MetaCrop (Schreiber et al., 2012), MaizeCyc (Monaco et al., 2013), and the earlier $i$RS1563 (Saha et al., 2011) model. In comparison to the $i$RS1563 model, this second-generation model spans an additional 4,261 genes and 6,540 reactions. The increased number of genes and reactions enables the inclusion of additional pathways such as fructan biosynthesis, siroheme biosynthesis, and ubiquinol-9 biosynthesis. The model accounts for the two major cell types in the leaf (i.e., the bundle sheath and mesophyll cells). The bundle sheath cell contains seven compartments: the cytosol, mitochondrion, peroxisome, chloroplast stroma, plasma membrane, thylakoid membrane, and vacuole. The mesophyll cell contains six compartments: the cytosol, mitochondrion, chloroplast stroma, plasma membrane, thylakoid membrane, and vacuole. Compartmentalization is based on maize-specific experimental proteomic and transcriptomic measurements (Majeran et al., 2005; Friso et al., 2010; Li et al., 2010; Chang et al., 2012) as opposed to the Arabidopsis-based compartmentalization adopted in the previous $i$RS1563 maize model (Saha et al., 2011). Light reactions have been expanded from an aggregate reaction (as described in the $i$RS1563 model) to multiple reactions for each complex with the inclusion of a thylakoid membrane compartment. In contrast to the C4GEM maize model (de Oliveira Dal'Molin
et al., 2010), which focuses exclusively on primary metabolism in maize, the developed model also spans secondary metabolism by including all reactions known to occur within the maize leaf tissue. The model includes as many as 763 secondary metabolism reactions (without including duplicate counting due to compartmentalization). Through the incorporation of “omics” data, regulatory restrictions are introduced in the model to switch-off/on reactions under the N+ WT, N- WT condition and two GS knockout mutants (gln1-3 and gln1-4) in the vegetative (V) stage during which the plant absorbs and assimilates N for root and leaf biomass production (Amiour et al., 2012; 2014). Reactions linked to genes or proteins with significantly different expression levels between N+ WT and N- WT condition, as well as the gln1-3 and gln1-4 mutants versus the N+ WT are conditionally turned on or off accordingly. The metabolite pool is simulated by maximizing the total flux through a metabolite (i.e. flux-sum) as a proxy for the metabolite turnover rate (Chung and Lee, 2009). The directional change of flux-sum levels between the N- WT condition and the N+ WT condition, as well as the GS mutant conditions and the N+ WT condition are qualitatively compared to the directional change in experimentally measured concentration levels. These analyses reveal similar trends as the recently developed Flux Imbalance Analysis (Reznik et al., 2013) that makes use of dual variable values associated with metabolite balances to infer the effect of concentration changes on the objective function value.

Results and Discussion

Effect of Nitrogen Conditions on Biomass Components

Biomass components were measured in the N+ WT condition, as well as for each N background (N- WT, gln1.3 and gln1.4 mutants). Table I and Figure 1 display the composition of the classes of biomass metabolites and Supplemental Table S1 indicates the specific biomass measurements in all modeled conditions. As expected, in the majority of cases, the N- WT condition produced a smaller concentration of biomass components than the N+ WT, the gln1-3, and the gln1-4 mutant conditions. However, the concentration of amino acids produced was about five times higher in the gln1-4 mutant than the gln1-3 mutant, resulting in comparable amino acid concentrations between the gln1-4 mutant and the N+ WT, as well as between the gln1-3 mutant and N- WT conditions.
conditions. The similar amino acid concentrations between the *gln1-4* mutant and N⁺ WT condition in the vegetative stage help to confirm that the GS1-4 isozyme is essential in plant maturity and has a smaller effect compared to the GS1-3 isozyme, at the vegetative stage. As expected, the concentration of starch was higher in the N⁺ WT condition than the N⁻ WT condition. Under N⁻ conditions, the breakdown of starch is limited by the amount of N available (Tercé-Laforgue et al., 2004; Amiour et al., 2012). Due to the limited N available, the starch is stored rather than broken down to produce other biomass components. The stained micrograph depicting the starch visible in the N⁺ WT, *gln1-3* mutant, and *gln1-4* mutant conditions are available in Supplemental Figure S1. The condition-specific biomass concentrations have been incorporated in the maize leaf model to more accurately represent metabolism under each condition.

**Development of the second-generation maize leaf model**

The second-generation maize leaf model was developed using a combination of gene, protein, and reaction information from the previously developed maize model *i*RS1563 (Saha et al., 2011), biological databases such as KEGG (Kanehisa et al., 2014), MaizeCyc (Monaco et al., 2013), and MetaCrop (Schreiber et al., 2012), as well as published literature sources. The model contains 5,824 genes and 8,525 reactions, a significant increase from the *i*RS1563 model, which contained 1,563 genes and 1,985 reactions. The second-generation maize model is split into two cell types, (i.e. the bundle sheath and mesophyll cells). The bundle sheath cell is further divided into seven compartments, while the mesophyll cell contains six compartments (Fig. 2). Of the 8,525 reactions in the model, 3,892 reactions are unique, as duplicated counts due to compartmentalization have been disregarded. Of these 3,892 unique reactions, 1,012 reactions were assigned localization information based on transcriptomic and proteomic data (Majeran et al., 2005; Friso et al., 2010; Li et al., 2010; Chang et al., 2012). Light reactions were adjusted to model the flow of protons across the thylakoid membrane to the chloroplast stroma, to represent the pH differential between compartments, and to describe the conversion of light to ATP (Nelson and Cox, 2009). The mitochondrial electron transport chain was similarly updated to include the proton exchange of ATP synthase between the intermembrane space and the mitochondrial matrix (Taiz, 2010).
Finally, 303 specific reactions were added to model glycerolipid synthesis as shown in Supplemental Figure S2 and Table S2 (Moore, 1982; Murata, 1983; Murata and Tasaka, 1997; Mekhedov et al., 2000; Bachlava et al., 2009; Li-Beisson et al., 2010; Rolland et al., 2012). To the best of our knowledge, this is the first plant model to include detailed glycerolipid synthesis. Aggregate reactions were included to link specific two-tailed glycerolipids to the experimentally measured single lipids (see Supplemental Table S2). Compiling transcriptomic and proteomic compartmentalization data with literature-based pathways yielded a model of 3,587 reactions, leaving 2,880 unique reactions, still with their localization unknown.

Once reactions were compartmentalized based on transcriptomic data, proteomic data, and published literature, the reactions were divided into two groups. The first group (core set) includes reactions with known localization, while the second group (non-core set) spans reactions known to occur within the maize leaf but with no localization evidence. Whenever possible, core reactions were unblocked by first adding reaction(s) from the non-core set to one or multiple compartment(s) and second by appending inter- or intra-cellular transporter(s) (see Materials and Methods section). By following this approach, 1,032 unique reactions with previously unknown localization were assigned to compartments and 729 transporters were added. The remaining 1,848 unique reactions were assigned to compartments based on available pathway information or assigned to the cytosol of both the bundle sheath and mesophyll cells.

With all the reactions assigned to specific compartments, thermodynamically infeasible cycles that were generated due to the overly permissive inclusion of reactions in the model, as well as lack of reaction directionality information, were subsequently identified and eliminated. By first restricting the directionality of reactions and second removing reactions, it was possible to eliminate all thermodynamically infeasible cycles in the model. By this process we restricted the directionality of 36 reactions and removed 2,055 reactions from the model (Table II). Upon the resolution of thermodynamically infeasible cycles, attempts were made to unblock the remaining blocked core reactions and biomass formation by adding reactions from similar organisms (Krumholz et al., 2012) and model
organisms (i.e. *Oryza sativa* L. ssp. *japonica*, *Brachypodium distachyon*, *Sorghum bicolor*, and *Arabidopsis thaliana*). By adding five reactions from similar organisms, the flux through three additional reactions known to be in maize was resolved. These reactions were all involved in the formation of glutamate from histidine through urocanic acid. The model is provided in a Microsoft Excel format in Supplemental Table S3 and Systems Biology Markup Language (SBML) format in S4.

**Incorporation of transcriptomic and proteomic data in the model**

In order to more accurately model the N+ WT, N- WT, and GS mutant conditions in maize, GPR associations mapped the gene transcripts and proteins that were statistically expressed at a low level to reactions that were turned-off in the model. However, no essential reactions to the model, which are required for biomass formation, were altered. For example, the δ-aminolevulinate acid dehydratase reaction was experimentally determined to be higher in the N+ WT condition, suggesting it should be restricted in the N- WT condition. However, when the flux through the δ-aminolevulinate acid dehydratase reaction is restricted to zero, biomass cannot be formed, as this reaction produces porphobilinogen, a precursor to chlorophyll (Gupta et al., 2013). Due to the incomplete information available in published literature/databases regarding possible alternate routes of the production and degradation of a specific metabolite, blocking of reactions that are essential to the model will restrict biomass synthesis. By restricting the fluxes based on experimental evidence, the fluxes through 83 reactions in the N+ WT condition, 20 in the N- WT condition, 100 in the *gln1-3* mutant, and nine reactions in the *gln1-4* mutant were restricted. The reactions restricted in the N+ WT mainly correspond to reactions known to only occur under stress and are expressed at a low level in comparison to the N- WT and mutant conditions. Reactions that have been down-regulated based on “omics” data are indicated in the model file (Supplemental Table S3). N perturbations within the leaf tissue were modeled by combining the incorporation of transcriptomic and proteomic data with the unique biomass compositions for each condition.

The minimal set of reactions, whose elimination causes a decrease in biomass yield, was determined for the N+ WT, N- WT, *gln1-3* mutant, and *gln1-4* mutant conditions. There
are six reactions across the conditions that encompass the minimal set of reactions, as summarized in Table III. Of the 83 reactions with restricted flux in the N⁺ WT condition, only two reactions were identified to affect biomass yield. These two reactions are the conversion of ethanol to acetaldehyde through either ethanol oxidoreductase involving NAD⁺ or a hydrogen peroxide-dependent oxidation of ethanol catalyzed by catalase (Boamfa et al., 2005). These two reactions have a very slight effect on biomass formation as biomass yield drops by less than 1%. As expected, we find that many of the reactions that correspond to genes that are significantly down-regulated in the N⁺ WT condition do not hinder biomass formation. In the N⁻ WT condition, none of the reactions have an effect on the biomass yield suggesting, as expected, that the decreased amount of N is the main limiting factor in biomass yield. In the gln1-3 mutant condition, three of the 100 reactions, which are switched off based on “omics” data, affect the biomass yield. These three reactions include the glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase, and the fructose bisphosphatase reactions. The capacity of glyceraldehyde-3-phosphate dehydrogenase to form a multi-enzyme complex in the chloroplasts for a range of plants is regulated by environmental conditions such as the light/dark transitions (Howard et al., 2011). Glyceraldehyde-3-phosphate is synthesized during carbon fixation in photosynthesis and 1,3-bisphospho-D-glycerate (i.e. 3-phospho-D-glyceroyl phosphate) can be synthesized from 3-phospho-D-glycerate. ATP is required for the conversion of 3-phospho-D-glycerate to 1,3-bisphospho-D-glycerate catalyzed by 3-phospho-D-glycerate kinase in the bundle sheath chloroplast. This reaction is an important energy requiring reaction in the Calvin-Benson cycle, as it is essential that the enzyme immediately metabolizes 3-phospho-D-glycerate, the product of the Rubisco reaction. This conclusion is also consistent with the findings that 3-phospho-D-glycerate 1-phosphotransferase is sensitive to changes in energy state (Nakamoto and Edwards, 1987). The fructose-bisphosphate aldolase reaction, which is involved in the Calvin-Benson-Bassham cycle and glycolysis pathway, can be bypassed using the sedoheptulose 1,7-bisphosphate/D-glyceraldehyde-3-phosphate-lyase reaction which catalyzes the synthesis of glyceraldehyde 3-phosphate and D-erythrose 4-phosphate using sedoheptulose 1,7-bisphosphate (Lakshmanan et al., 2013). The decreased expression of cytosolic fructose bisphosphatase reaction has been shown to decrease the ATP/ADP ratio, lead to the
switch from sucrose to starch synthesis, and inhibit photosynthesis at high CO2 levels in Arabidopsis resulting in the inhibition of plant growth (Strand et al., 2000). Finally, the regulatory restrictions for the gln1-4 mutant involve only nine reactions of which one affected the biomass drain (i.e., ribose-5-phosphate isomerase reaction). The lack of ribose-5-phosphate isomerase reaction has been experimentally shown to cause premature death and affect cellulose synthesis in Arabidopsis (Howles et al., 2006; Xiong et al., 2009). A comparison of the number of reactions that affect the GS mutants suggests that at the V stage, the impact of the gln1-4 mutation is less severe than that occurring in the gln1-3 mutant. Such a finding is not surprising, since it has been shown that the gene encoding the GS1-3 isozyme is constitutively expressed irrespective of the leaf development stage and that the expression of the gene encoding the GS1-4 isozyme is much lower and only enhanced at later stages of leaf development (Hirel et al., 2005). Although only a subset of reactions affect the biomass production in the N+ WT, gln1-3 mutant, and gln1-4 mutant conditions, the additional regulation will have an effect on the flux predictions within the model.

**Flux range variations among conditions**

The flux range of each reaction was determined in the N+ WT, N WT, gln1-3 mutant, and gln1-4 mutant conditions under the assumption that the biomass is being maximized. The flux range of a reaction in the N WT, gln1-3 mutant, and gln1-4 mutant conditions was compared to the flux range in the N+ WT reference condition to determine reactions whose flux ranges must deviate from the N+ WT flux range. This indicates that the flux through the reaction must change as a result of the limited N or mutation. Overall, the flux through 202 reactions in the N WT condition are not contained within the flux range of the N+ WT condition, 765 reaction fluxes in the gln1-3 mutant diverge from the N+ WT flux range, and 678 reaction fluxes in the gln1-4 mutant must change from the N+ WT flux range (Supplemental Table S5). In all three N backgrounds (i.e. the N WT condition, gln1-3 mutant, and gln1-4 mutant conditions) the flux compared to the N+ WT reference condition decreases under maximum biomass through the chlorophyll cycle, chlorophyllide a biosynthesis, farnesyl diphosphate biosynthesis, methylerythritol phosphate pathway, and tetrapyrrole biosynthesis. Tetrapyrrole biosynthesis,
chlorophyllide $a$ biosynthesis, and the chlorophyll cycle link the production of chlorophyll from glutamate (Kim et al., 2013). The methylerythritol phosphate pathway and farnesyl diphosphate biosynthesis lead to a reactant required for the production of chlorophyll $a$ from chlorophyllide $a$ (Lange and Ghassemian, 2003). In both of the GS mutant conditions, the flux through chorismate biosynthesis (Tzin and Galili, 2010), serine biosynthesis (Ho and Saito, 2001), and the urea cycle (Merigout et al., 2008) must decrease compared to the N$^+$ WT condition. Choline biosynthesis (McNeil et al., 2001) is decreased in the N$^-$ WT condition, increased in the $gln1-3$ mutant, and decreased in the $gln1-4$ mutant condition. Flux through the isoleucine and leucine biosynthesis (McCourt and Duggleby, 2006) is lower in the N$^-$ WT condition, higher in the $gln1-3$ mutant condition, and lower in the $gln1-4$ mutant condition compared to the N$^+$ WT condition, as expected by the proportion of these biomass components in the various conditions. The flux through the glyoxylate cycle (Schnarrenberger and Martin, 2002), stearate biosynthesis (Li-Beisson et al., 2010), and urate degradation (Ramazzina et al., 2006) is higher in the $gln1-3$ mutant condition compared to the N$^+$ WT condition. Valine biosynthesis (McCourt and Duggleby, 2006) is lower in the $gln1-3$ mutant condition compared to the N$^+$ WT condition. Flux through the glutathione biosynthesis/degradation, tryptophan biosynthesis (Tzin and Galili, 2010), uracil degradation (Zrenner et al., 2006), and xylose degradation (Penna et al., 2002) is higher in the $gln1-4$ mutant compared to the N$^+$ WT condition. Glutamate is converted to glutathione through two ATP dependent steps requiring the addition of cysteine and then glycine. Glutathione is a vitally essential protectant against oxidative stress, heavy metals and xenobiotics (Noctor et al., 2012; Rahantaniaina et al., 2013). Several routes of glutathione breakdown have been proposed including the formation of cysteine and glycine through cysteinylglycine. The cysteine is then degraded to form pyruvate helping to alleviate the $gln1-4$ mutation. The increased fluxes associated with xylose (from 1,4 β-D-xylan) and uracil degradation generate a larger pool of xylulose-5-phosphate and β-alanine, respectively. Finally, phenylpropanoid biosynthesis (Vogt, 2010) is lower in the $gln1-4$ mutant condition compared to the N$^+$ WT condition. The majority of the changes in these pathways are directly related to the differences in the proportion of the biomass components between the modeled conditions.
Comparison of model predictions to metabolomic data

The metabolomic data were compared to flux predictions within the model in each of the various N background conditions. The increasing or decreasing trend of the metabolite concentration, displayed in Figure 3, was qualitatively compared to the change in the flux-sum range determined by the model, as displayed in Figure 4. The flux-sum is a measure of the amount of flow through the reactions associated with either the production or consumption of the metabolite. A variability analysis of flux-sum was performed and flux-sum ranges, normalized by the biomass rate, that do not overlap between the N background condition and the N⁺ WT condition were analyzed. An increase/decrease in the flux-sum (i.e., used as a proxy for the metabolite pool) of a metabolite between the N⁻ WT condition and the N⁺ WT condition and between the two GS mutants and the N⁺ WT condition was compared with the metabolite concentration changes. Figure 4 demonstrates the importance of restricting fluxes based on transcriptomic and proteomic data. In the N⁻ WT condition, the accuracy changes from 13% to 90% when the flux constraints based on “omics” data are incorporated. Without the incorporation of these constraints, all flux-sum ranges normalized by the biomass rate are predicted higher in the N⁻ WT condition. The identified flux-sum levels are included in Supplemental Table S6. The flux-sum variability approach is able to predict the change in metabolite pool sizes more accurately when the flux ranges are similar to the wild-type condition, as in the N⁻ WT condition. Between the N⁻ WT and N⁺ WT condition, only approximately 7% of the reactions active in either condition have flux ranges at the maximum biomass that do not overlap between the two conditions. In the gln1-3 and gln1-4 mutant conditions, the fluxes are significantly perturbed with 49% and 45% of the active reactions at maximum biomass resulting in non-overlapping ranges compared to the N⁺ WT condition, respectively. The accuracy of flux-sum in the gln1-3 mutant and gln1-4 mutant condition with “omics”-based constraints incorporated reach 53% and 25% accuracy with 8 of 15 metabolites predicted correctly and 1 of 4 metabolites predicted correctly in the gln1-3 and gln1-4 mutant conditions, respectively. This level of prediction accuracy is far below what was seen for the N⁻ WT alluding to a tenuous connection between concentration changes and gene expression levels when the genetic background changes.
We explored the efficacy of the flux-sum method under different genetic backgrounds for a much more well studied and data-rich organism (i.e., *E. coli*) to explore whether the dissonance between gene expression levels and concentrations was maize-specific or applied broadly. We applied flux-sum variability to the Ishii et al. (2007) fluxomic and metabolomic data using the *i*AF1260 (*Feist et al., 2007*) *E. coli* model. Two single-gene knockout mutants (i.e. *ppsA* and *glk*) were compared to the wild-type condition and predicting the directional change of the metabolite pool size was met with less than 50% accuracy in each condition. This implies that changes in the genetic background seem to cause concentration changes that are not predictable by gene expression changes alone. In contrast, changes in nutrient availability as in the N WT condition can be captured with 90%.

We also decided to explore whether the dissonance between gene expression levels and concentration ranges was caused due to a deficiency in the proposed flux-sum method. As an alternative, we used flux imbalance analysis (Reznik et al., 2013) that measures the effect of the deviation of a metabolite’s concentration from steady state on the maximum biomass by applying the concept of duality. Flux imbalance analysis, examines how the model responds to a deviation from steady state by measuring the effect on biomass when a metabolite is allowed to accumulate or deplete. By determining the change in biomass formation due to the accumulation or depletion of the metabolite, a prediction can be made as to the change in metabolite levels. Flux imbalance analysis was applied to the model and the deviation in the maximum biomass was qualitatively compared to the experimental data for the metabolite in each compartment. Only non-overlapping ranges of the marginal value associated with each compartment-specific metabolite were analyzed. If all compartment-specific metabolites have marginal values that indicate the same trend compared to the N WT condition, a prediction was made for the tissue-specific metabolite. The flux imbalance analysis is 66%, 33%, and 78% accurate in the N WT, *gln*1-3 mutant, and *gln*1-4 mutant conditions, respectively, as compared to the N WT condition. While flux imbalance analysis makes a prediction for every metabolite in the model, the flux-sum analysis only predicts a direction of change for metabolites whose associated reactions can carry flux. Flux imbalance analysis allows for the
prediction of compartment-specific metabolites whose associated reactions do not carry flux under maximum biomass formation. Comparable results between the flux imbalance analysis and flux-sum analysis in the N− WT condition provide independent backing regarding the validity of the flux-sum concept.

**Concluding remarks**

We have introduced a second-generation model that is specific for the leaf tissue of maize and differentiates between the bundle sheath and mesophyll cell types. By incorporating transcriptomic and proteomic data into the model, we were able to reproduce the metabolomic data with up to 90% accuracy when comparing the N− WT and N+ WT conditions. Ethanol oxidoreductase/catalase, glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase, fructose bisphosphatase, and ribose-5-phosphate isomerase were shown to be important genes in suppressing biomass formation in the modeled conditions. In order to study the impact of these genes on plant biomass production when optimal N is provided, their functional validation can then be undertaken using transgenic technologies, mutagenesis, or by association genetics, either at the single gene or genome-wide level (Simons et al., 2014). The model also predicted a modification of the flux of metabolites formed during glutathione catabolism in the gln1-4 mutant conditions compared to the N+ WT condition. This modification is predicted to compensate for the lack of GS1-4 by using the glutamate and pyruvate derived from glutathione to produce alanine. Thus, it will be interesting to determine whether the increase in alanine is related to the importance of the enzyme alanine aminotransferase in the improvement of plant productivity in general and NUE in particular (Good and Beatty, 2011; McAllister et al., 2013). In all N background conditions (i.e. N− WT, gln1-3 mutant, and gln1-4 mutant conditions), we find that the flux through chlorophyll biosynthesis, and those pathways directly related to chlorophyll biosynthesis, decrease confirming the important link between N metabolism and chlorophyll synthesis through the use of its precursor glutamate (Forde and Lea, 2007). The leaf model, with the addition of other maize tissue-specific models, can be integrated into a whole-plant genome-scale model for maize. By determining a required metabolic function that is specific to each tissue, tissue-specific models can be created ensuring that only relevant
reactions are included in each tissue. Future efforts will focus on tissue-specific models for the kernel, stalk, tassel, and root tissues. These tissue-specific models will follow community (Zomorrodi and Maranas, 2012; Zomorrodi et al., 2014) and multi-tissue human models (Duarte et al., 2007; Bordbar et al., 2011; Thiele et al., 2013) reconstruction principles. The tissues can be linked using inter-tissue transport reactions with the stalk tissue acting as the central transporter among the various tissues and particularly to developing ear (Cañas et al., 2012). A whole-plant genome-scale model of maize will help to elucidate the flow of N from the root to the other tissues in the plant, from the shoot to the ear, and within the developing ear (Cañas et al., 2010). By modeling the entire plant, non-intuitive bottlenecks in N metabolism can be determined, which then can be used to suggest genetic interventions through mutagenesis, transgenic technology, or maker-assisted selection to increase the NUE in maize. In addition, the flow of sugars to the kernel tissue can be analyzed to guide the increase of carbohydrate/sugar content of maize kernel by breaking the inverse relationship existing between carbohydrates and proteins (Feil et al., 1990). Apart from its crucial role as a food crop, maize is also used for cellulosic biofuels. To this end, the amount and composition of cell wall polymers is important in developing cellulosic maize. Lignin not only provides rigidity to the maize plant (Sticklen, 2008; Vanholme et al., 2008), but also makes digestion of cellulosic and hemicellulosic sugars difficult during delignification (Li et al., 2008). Recent research endeavors have focused on altering lignin content, since plant viability and fitness are affected by lignin reductions (Li et al., 2008; Bonawitz et al., 2014). Therefore, by utilizing the whole-plant genome-scale model a system-wide implication of these genetic disruptions can be quantitatively assessed, thus facilitating new strategies for reducing lignin content without affecting the mechanical integrity of the maize plant.

**Materials and Methods**

**Plant Material**

Maize wild-type (WT) plants, (*Zea mays* L., genotype B73) and *gln1.3* and *gln1.4* mutant seeds in the B73 background (see Martin et al., 2006, for the production, selection and characterization of the mutants) were grown as described by Amiour et al. (2012) in a glasshouse at the Institut National de la Recherche Agronomique, Versailles, France from...
May to September 2004. Three individual plants of similar size and of similar developmental stage were selected corresponding to the three replicates used for the “omics” experiments. The three youngest fully expanded leaves at the 10 to 11 leaf stage without the midrib were harvested and pooled for the vegetative stage (V) samples to obtain enough homogenous plant material representative of this plant development stage.

Plants were watered daily with a complete nutrient solution containing 10 mM KNO$_3$ as the sole N source in the N$^+$ WT, gln1-3 mutant, and gln1-4 mutant conditions (Coïc and Lesaint, 1971). The N- WT condition was supplied 0.01 mM KNO$_3$. The complete nutrient solution also contained 1.25 mM K$^+$, 0.25 mM Ca$^{2+}$, 0.25 mM Mg$^{2+}$, 1.25 mM H$_2$PO$_4^-$, 0.75 mM SO$_4^{2-}$, 21.5 μM Fe$^{2+}$ (Sequestrene; Ciba-Geigy, Basel, Switzerland), 23 μM B$^{3+}$, 9 μM Mn$^{2+}$, 0.3 μM Mo$^{2+}$, 0.95 μM Cu$^{2+}$, and 3.5 μM Zn$^{2+}$.

**Yield Components Analysis**

Kernel yield, its components, and the N content of different parts of the plant at stages of development from silking to maturity were determined according to the method described by Martin et al. (2005) and corresponded to the data described in Martin et al. (2006) and Amiour et al. (2012).

**RNA and DNA Preparation**

Total RNA was extracted as described by Verwoerd et al. (1989) from leaves that had been stored at -80°C. Total RNAs (50 μg) for transcriptome and quantitative Real Time-Polymerase Chain Reaction (qRT-PCR) studies were treated and prepared as previously described by Amiour et al. (2012). Reverse transcription reactions and quantitative first strands were synthesized according to Amiour et al. (2012). Primers for qRT-PCR and RT-PCR cloning were designed from Bacterial Artificial Chromosomes (BAC) sequences found in the public maize genome databases (Maizesequence.org, PlantGDB, Genbank). The sequences of the primers used in RT-PCR and qRT-PCR are presented in Table S1.

**Gene Expression Profiles using Maize cDNA Microarrays**
Whole genome leaf transcript profiling was performed using the maize 46K arrays obtained from the maize oligonucleotide array project (http://www.maizedcna.org/outreach/resources.html) essentially as described previously by Amiour et al. (2012). Transcript abundance in each of the three replicates for V leaves was determined using a mixture of all the samples (18 in total, each with the same mRNA concentration) as a reference. Statistical significance for differentially expressed genes essentially was evaluated through statistical group comparisons performed using multiple testing procedures as described by Amiour et al. (2012). Transcriptomic data were validated by qRT-PCR analysis performed on a selected number of up- or down-regulated gene transcripts.

**Statistical Analysis of Maize cDNA Microarray Data**

Statistical significance for differentially expressed genes was evaluated with statistical group comparisons using multiple testing procedures. The following two gene selection approaches were applied: the Significance Analysis of Microarrays (SAM; Tusher et al., 2001) permutation algorithm, and a \( p \)-value ranking strategy using both \( z \)-statistics in ArrayStat 1.0 software (Imaging Research Inc.) and moderated \( t \)-statistics using a moderated \( t \)-test available in MAnGO tools [http://bioinfome.cgm.cnrs-gif.fr] and BRBArrayTools v3.2.3 packages (Korn et al., 2002). For multiple testing corrections, the false discovery rate (FDR) procedure was used (Benjamini and Hochberg, 1995). Statistical tests were computed and combined for each probe set using the log-transformed data. A significant probe set indicates an adjusted \( p \)-value was less than the effective \( \alpha \)-level (\( \alpha=0.05 \)) in at least one of the two gene selection tests. A filtering procedure was used to exclude data points with low signal intensities (Amean < 7.0) that are considered biologically unreliable.

**Total Protein Extraction, Solubilization, and Quantification**

A TCA/acetone protein precipitation was performed as described by Méchin et al. (2007), from the leaves of \( N^+ \) and \( N^- \) plants harvested at the V stage of development. The frozen leaf powder was resuspended in acetone with 0.07% (v/v) 2-mercaptoethanol and 10% (w/v) TCA. Proteins were allowed to precipitate for 1 h at -20°C. The pellet was then...
washed overnight with acetone containing 0.07% (v/v) 2-mercaptoethanol. The
supernatant was discarded and the pellet dried under vacuum. Protein resolubilization
was performed according to Méchin et al. (2007) using 60 μL/mg of R2D2 buffer (5 M
urea, 2 M thiourea, 2% CHAPS, 2% SB3-10, 20 mM dithiothreitol, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride, 0.75% carrier ampholytes). After resolubilization,
samples were centrifuged and the supernatant was transferred to an Eppendorf tube prior
to protein quantification. Total protein content of each sample was evaluated using the 2-
D Quant kit (Amersham Biosciences).

Two-dimensional Electrophoresis, Gel Staining, and Image Analysis
Total protein extraction, solubilization, and quantification were performed as
described by Méchin et al. (2007). Solubilized proteins (300μg) were subjected to two-
dimensional gel (2-D gel) electrophoresis and identified by liquid chromatography-mass
spectrometry as described by Amiour et al. (2012)

Protein Identification by LC-MS/MS
Spot digestion and LC-MS/MS were performed as described by Martin et al. (2006). In-
gel digestion was performed with the Progest system (Genomic Solution). Gel pieces
were washed twice by successive separate baths of 10% acetic acid, 40% ethanol, and
acetonitrile (ACN). The pieces were then washed twice with successive baths of 25 mM
NH₄CO₃ and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of
modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH₄CO₃. The
peptides were extracted successively with 2% trifluoroacetic acid (TFA) and 50% ACN
and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in
20 mL of 0.05% TFA, 0.05% formic acid, and 2% ACN. HPLC was performed on an
Ultimate LC system combined with a Famos Autosampler, and a Switchos II
microcolumn switch system (Dionex). Trypsin digestion was declared with one possible
cleavage. Cys carboxyamidomethylation and Met oxidation were set to static and variable
modifications, respectively. A multiple-threshold filter was applied at the peptide level:
Xcorr magnitude were up to 1.7, 2.2, 3.3, and 4.3 for peptides with one, two, three, and
four isotopic charges, respectively; peptide probability lower than 0.05, ΔCn > 0.1 with a minimum of two different peptides for an identified protein. A database search was performed with BioWorks 3.3.1 (Thermo Electron). The TIGR maize gene index database v 16, 72047*6 EST sequences (http://compbio.dfci.harvard.edu/tgi/plant.html) was used.

**Metabolite Extraction and Analyses**

Lyophilized leaf material was used for metabolite extraction. Approximately 20 mg of the powder was extracted in 1 ml of 80% ethanol/20% distilled water for an hour at 4°C. During extraction, the samples were continuously agitated and then centrifuged for 5 min at 15,000 rpm. The supernatant was removed and the pellet was subjected to a further extraction in 60% ethanol and finally in water at 4°C, as described above. All supernatants were combined to form the aqueous alcoholic extract.

Nitrate was determined by the method of Cataldo et al. (1975). Total soluble amino acids were determined by the Rosen colorimetric method with leucine as a standard (Rosen, 1957). Chlorophyll was estimated using 10 mg of fresh leaf material (Arnon, 1949). The total N content of 2 mg of lyophilized material was determined in a N elemental analyzer using the combustion method of Dumas (Flash 2000, Thermo Scientific, Cergy-Pontoise, France). Starch content was determined as described by Ferrario-Méry et al. (1998).

Total lipids were extracted from frozen leaf material according to Miquel and Browse (1992). Individual lipids were purified from the extracts by one-dimensional thin layer chromatography on silica gel 60 plates (Ohnishi and Yamada, 1980; Lepage, 1967), which were obtained from Merck (Merck-Millipore, Molsheim, France). Lipids were located by spraying the plates with solution of 0.001 % primuline (Sigma, Saint-Quentin Fallavier, France) in 80 % acetone, followed by visualization under ultraviolet light. To determine the fatty acid composition and relative amounts of individual lipids, the silica gel for each lipid was transferred to a screw-capped tube with 1 ml of 2.5 % (v/v) H2SO4 in methanol and an appropriate amount of C17:0 fatty acid (Sigma, Saint-Quentin Fallavier, France), as an internal standard. After heating for 90 minutes at 80 °C, 1 ml of
hexane and 1.5 ml of 0.9% NaCl$_2$ were added. Fatty acids were extracted in the upper organic phase by shaking and low-speed centrifugation. Samples (1 µl) of the organic phase were separated by gas chromatography on a 30-m x 0.53-mm EC$^\text{TM}$-WAX column (Alltech Associates Inc., Deerfield, USA) and quantified using a flame ionization detector. The gas chromatograph was programmed for an initial temperature of 160 °C for 1 min, followed by an increase of 20 °C min$^{-1}$ to 190 °C and a ramp of 4 °C min$^{-1}$ to 230 °C, with a 9-min hold of the final temperature.

The monosaccharide composition and linkage analysis of polysaccharides were determined as follows: 100 mg (FW) of ground leaf were washed twice in 4 volumes of absolute ethanol for 15 min, then rinsed twice in 4 volumes of acetone at room temperature for 10 min and left to dry under a fume hood overnight at room temperature. The neutral monosaccharide composition was measured on 5 mg of dried alcohol insoluble material after hydrolysis in 2.5 M trifluoroacetic acid for 1.5 h at 100 °C as described by Harholt et al. (2006). To determine the cellulose content, the residual pellet obtained after the monosaccharide analysis was rinsed twice with ten volumes of water and hydrolysed with H$_2$SO$_4$ as described by Updegraff (1969). The released glucose was diluted 500 times and then quantified using an HPAEC-PAD chromatography as described by Harholt et al. (2006).

For lignin quantification, 100 mg (FW) of ground leaf were washed twice in 4 volumes of absolute ethanol for 15 min, twice with 4 volumes of water at room temperature then rinsed twice in 4 volumes of acetone at room temperature for 10 min and left to dry under a fume hood overnight at room temperature. The following protocol is adapted from Fukushima and Hatfield (2001). Lignins from the prepared cell wall residue were solubilized in 1 mL of acetyl bromide solution [acetyl bromide/acetic acid (1/3, V/V)] in a glass vial at 55°C for 2.5 h under shaking. Samples were then allowed to cool down to room temperature and 1.2 mL of NaOH 2M/Acetic acid (9/50 V/V) were added in the vial. A hundred µL of this sample were transferred in 300 µL of 0.5M hydroxylamine chlorhydrate and mixed with 1.4 mL of acetic acid. The A280 absorbance of the samples
was measured. The lignin content was calculated using the following formula:

\[ \% \text{lignin} = \frac{100 \times (A_{280} \times V_{\text{reaction}} \times V_{\text{dilution}})}{20 \times V_{\text{sample solution}} \times m_{\text{sample mg}}} \]

**Metabolome Analysis**
All steps were adapted from the original protocol described by Fiehn (2006) following the procedure described by Amiour et al. (2012).

**Model Development and Curation**
Figure 5 outlines the workflow used for model development. Our previously developed maize model, iRS1563 (Saha et al., 2011) and biological databases such as MetaCrop (downloaded in December 2012; Schreiber et al., 2012) and MaizeCyc (version 2.0.2; Monaco et al., 2013) provided information pertaining to the genes, proteins, reactions, and metabolites used to reconstruct the second-generation maize leaf genome-scale model. In addition, available proteomic and transcriptomic data, maize-specific biological databases, namely MetaCrop and MaizeCyc and published literature were used to assign cellular (i.e., bundle sheath or mesophyll) and intra-cellular organelle specificity to the curated reactions.

When the gene expression level was reported in reads per kilobase per million mapped reads (RPKM) (Li et al., 2010; Chang et al., 2012) the cell specificity of any gene \( i \) can be calculated as:

\[ R_i = \frac{|m_i - b_i|}{\max(m_i, b_i)} \]

Here, \( m_i \) and \( b_i \) are the RPKM abundance of gene \( i \) in the mesophyll and bundle sheath cells, respectively (Chang et al., 2012). A gene that is only expressed in one cell type will have a \( R_i \) of 1, while a gene that is equally expressed in both cell types will have a \( R_i \) of 0. As suggested by Chang et al., a threshold of 0.8 or a five-fold abundance difference is adopted to assign gene cell type specificity. In the absence of RPKM information, an adjusted spectral count (adjSPC) along with the fold change difference between the
mesophyll and bundle sheath cells was used to determine gene cell type specificity (Friso
et al., 2010). adjSPC is the number of mass spectra identified for a protein normalized by
the number of unique spectral counts. Since low counts are not statistically informative, a
cutoff of 10 was used for adjSPC (Zybailov et al., 2008; Kim et al., 2009). Similar to the
threshold used for RPKM data, a five-fold difference between the mesophyll and bundle
sheath cell type normalized spectral abundance factor (nSAF) was used to determine
 cellular specificity of any gene (Friso et al., 2010). nSAF is a weighted adjSPC based on
the number of theoretical tryptic peptides with a relevant length (Ehleringer et al., 1997;
Friso et al., 2010). Additional intracellular compartmentalization was carried out based
on the MetaCrop database (Schreiber et al., 2012), MaizeCyc database (Monaco et al.,
2013), and primary literature sources (Chang et al., 2012; Zhao et al., 2013).

The intracellular compartmentalization was determined based first on the MetaCrop
database (Schreiber et al., 2012), literature sources (Friso et al., 2010; Chang et al., 2012),
compartmentalization information in the MaizeCyc database, and finally the Plant
Proteome Database (PPDB) (Sun et al., 2009). An original set of intercellular and
intracellular transporters was determined based on literature evidence (Alberte and
Thornber, 1977; Leegood, 1985; Stitt and Heldt, 1985; Furman et al., 1989; Weiner and
Heldt, 1992; Doulis et al., 1997; Burgener et al., 1998; Taniguchi et al., 2004; Sowiński
et al., 2008; Friso et al., 2010). In the subsequent standardization step, the MetRxn
knowledgebase (Kumar et al., 2012) as well as manual curation was used to standardize
the description of metabolites and reactions such as fixing stoichiometric errors (i.e.,
elemental or charge imbalances) and incomplete atomistic detail (e.g. absence of stereo-
specificity, and presence of R-group(s)). Reactions and metabolites were given KEGG
identifiers where available or were otherwise given new identifiers (in the form of MR or
MC, respectively). Reaction directionality was adopted from the manually curated
MetaCrop database, as available, and from the MaizeCyc database for the remaining
reactions.

In the next step of model development, all reactions (including metabolic, intra- and
extra-cellular transport reactions) were divided into two categories based on the evidence
of their inter- and intra-cellular compartmental specificity. The core set contains all metabolic reactions with experimental or literature-backed evidence of intracellular or intercellular compartmentalization, as well as known intracellular and intercellular transporters. The non-core set contains reactions with partial or completely absent localization information. Barring any conflicting evidence, these reactions were provisionally placed in all compartments. An optimization formulation (as shown below) was developed by imposing flow though the maximal number of core reactions while including minimal intra- and inter-cellular transporters and minimal participation of non-core reactions in various compartments. A parsimony criterion was used to apportion non-core functions, so that core functions could be restored. Furthermore, in order to restore a core function, the resolution strategy was prioritized in the following order: (i) apportion non-core reaction(s) in one/multiple compartment(s), (ii) add intra-cellular transporter(s) and (iii) add inter-cellular transporter(s). To this end, an objective function was formulated by taking the weighted sum of number of non-core reactions, intra- and inter-cellular transporters by providing weights of 1, $10^4$ and $10^6$, respectively for these three groups of reactions. However, it is important to ensure that any resolution strategy does not cause thermodynamically infeasible cycles. Therefore, each of these solutions was further checked and those reactions that result in the formation of a cycle were rejected. For each core reaction, multiple solutions were determined and the solution that fixes the largest number of core reactions was accepted. When required, manual curation was used to delineate between multiple solutions. This approach is analogous to the one proposed by Mintz-Oron et al. (2012) but does not rely on a complicated scoring system. It is also computationally less taxing, as it activates one core reaction at a time. Furthermore, in contrast to the Mintz-Oron et al. approach, the method proposed here allows for the minimal number of transporters added, rather than potentially minimizing the flux through many transporters. The process of minimally adding the number of reactions and transporters to the model is similar to that used by Model SEED (Henry et al., 2010). In order to allow flux through all reactions in the core set $C=\{1,\ldots,c\}$, we minimized the addition of reactions from the non-core set $NC=\{c+1,\ldots,g\}$, intracellular transporter set $T=\{g+1,\ldots,t\}$, and intercellular transporter set $IC=\{t+1,\ldots,m\}$. This
encompasses an overall set of reactions \( M = \{1, \ldots, m\} \) and a set of metabolites \( N = \{1, \ldots, n\} \). In addition, binary variable \( y_j \) is defined as:

\[
y_j = \begin{cases} 
1 & \text{if the reaction is added to the model from NC, I, or IC sets} \\
0 & \text{otherwise}
\end{cases}
\]

The task of identifying the minimal set of additional reactions that enable flux through a core reaction \( j^* \) is posed as the following mixed integer linear programming problem.

\[
\text{Minimize } c_1 \sum_{j \in NC} y_j + c_2 \sum_{j \in I} y_j + c_3 \sum_{j \in IC} y_j \quad \forall j \in C \quad (1)
\]

Subject to:

\[
\sum_{j=1}^{m} S_{ij} v_j = 0 \quad \forall i \in 1, \ldots, n \quad (2)
\]

\[
v_j \geq \varepsilon \quad \forall j \in C \quad (3)
\]

\[
v_{j,\text{max}} \geq v_j \geq v_{j,\text{min}} \quad \forall j \in C \quad (4)
\]

\[
v_{j,\text{max}} y_j \geq v_j \geq v_{j,\text{min}} y_j \quad \forall j \in \text{NC or T or I or IC} \quad (5)
\]

Here, \( S_{ij} \) is the stoichiometric coefficient of metabolite \( i \) in reaction \( j \) and \( v_j \) is the flux value of reaction \( j \). Parameters \( v_{j,\text{min}} \) and \( v_{j,\text{max}} \) denote the minimum and maximum allowable fluxes for reaction \( j \), respectively. \( v_{j^*} \) represents the core reaction flux that is currently being unblocked and \( \varepsilon \) is a small value to ensure a threshold amount of flux through each core reaction. \( c_1, c_2, \) and \( c_3 \) represent weights associated with each set of reactions (i.e., non-core set, intracellular transporters set, and intercellular transporters set, respectively). In this formulation, the objective function (1) minimizes the number of added reactions (from three reaction sets as mentioned earlier) so as to restore flux flow through reaction \( j^* \). We chose values of 1, \( 10^4 \), and \( 10^6 \) for \( c_1, c_2, \) and \( c_3 \), respectively, so metabolic reactions without experimental or literature evidence for compartmental specificity are added to specific compartment(s) before including additional transport reactions with no literature evidence. Constraint set (2) represents the pseudo-steady state assumption, while constraint (3) determines the threshold amount of flux necessary through \( j^* \). Bounds on core reaction fluxes are imposed by constraint set (4), while constraint set (5) ensures that only reactions from those three sets having non-zero flow are added to the model. This algorithm is repeated for each core reaction \( j^* \) to ensure flux
and, hence, provides compartmentalization assignments for 431 metabolic reactions by assigning them to at least one compartment, adding 1,032 total metabolic reactions to the model as shown in Table III.

The reactions identified by the above-mentioned algorithm plus the reactions from the core set constituted two new sets, a set of reactions with resolved compartmental information and a set whose location still needs resolution as shown in Figure 5. Reactions from the latter set that are known to occur within the maize leaf tissue, but were not in the initial model were added to intra/inter-cellular compartments manually based on pathway localization or simply added to cytosol of bundle sheath and/or mesophyll cells. Thermodynamically infeasible cycles were resolved by changing the minimum number of reaction directionalities as possible and eliminating the smallest number of reactions from the model (Schellenberger et al., 2011) while conserving biomass formation. An optimization procedure was iteratively run for each reaction in a thermodynamically infeasible cycle to determine the minimum number of directionality changes or removal of reactions required to fix the cycle. These results were then compared for each reaction to determine the changes that resolve the largest number of reactions participating in thermodynamically infeasible cycles. The solutions found were manually inspected before the changes were applied to the model. The application of this optimization procedure led to restricting the directionality for 507 reactions that prevented 889 reactions from carrying unbounded fluxes thus eliminating the corresponding thermodynamically infeasible cycles.

In the final step, as shown in Figure 5, the GapFind/GapFill (Kumar et al., 2007) procedure was applied to identify blocked/dead-end metabolites and subsequently restore their connectivity. A gapfilling database of reactions was created by combining reactions from phylogenetically close/model plant species (i.e. *Oryza sativa* L. ssp. *japonica*, *Brachypodium distachyon*, *Sorghum bicolor*, and *Arabidopsis thaliana*), non-core reactions without compartmental specificity (not identified by our aforementioned algorithm), and all possible intra/inter-cellular transporters. The gapfilling procedure was modified by prioritizing the addition of reactions from closely related/model plant species
or non-core reactions over transporters to unblock the flow through metabolites while ensuring no new thermodynamically infeasible cycles are created. After completing this step, we added five reactions from closely related/model plant species, changed the directionality of 14 reactions and added eight intracellular transporters.

**Incorporation of Transcriptomic, Proteomic and Metabolomic Data**

Significantly different gene transcripts and proteins were incorporated into the model by switching off corresponding reactions under the N+ WT condition, N- WT condition (Amiour et al., 2012), gln1-3 mutant, and gln1-4 mutant (Martin et al., 2006) conditions. The number of proteins, gene transcripts, and metabolites with abundances that are statistically differentially expressed in the various conditions are listed in Table IV. Reactions with GPRs associated with significantly lowered transcriptomic and proteomic expression are switched off under the corresponding conditions. Metabolite turnover rates were determined based on the flux-sum analysis method (Chung and Lee, 2009) and compared to the metabolomic data. The range of the flux-sum or the flow through each metabolite with experimental measurements was maximized/minimized as follows:

\[
\begin{align*}
\text{Max/Min} & \quad 0.5 \sum_{j=1}^{m} S_j v_j \\
\text{Subject to:} & \\
\sum_{j=1}^{m} S_j v_j &= 0, \forall i \in 1,...,n \\
v_{j,\text{min}} & \leq v_j \leq v_{j,\text{max}} \\
v_j &= 0, j \in \text{LE} \\
v_{\text{biomass}} &= v_{\text{biomass}}^{\text{max}}
\end{align*}
\]

Here set \( E \) represents the set of metabolites with experimental measurements and set \( \text{LE} \) represents reactions with statistically lower expression of gene transcripts and/or proteins. The formulation was run in an iterative manner for each metabolite with experimental measurements. The formulation was also repeated for each individual condition ensuring the proper nutrients and simulated knockouts were considered. By linearizing the objective function the resulting formulation is a mixed integer linear programming problem similar to the description by Chung and Lee (2009). Therefore, the basic idea is to determine the range of the flux-sum of a metabolite (for which metabolomic data is
available) under a given condition by switching off reaction fluxes corresponding to gene transcripts and/or proteins with lower expression levels (i.e. constraint 7). The flux-sum ranges were determined at the maximum biomass for the condition as displayed in constraint 8. Predictions were only made when the flux-sum ranges did not overlap between the N background condition and the N+ WT condition and when the direction of change in all compartments was consistent. In this way, the compartment-specific predictions of the flux-sum ranges were compared to tissue-specific experimental measurements. The flux-sum levels in the N WT, \( gln1-3 \) mutant, and \( gln1-4 \) mutant conditions were compared to the reference N+ WT condition to find the qualitative trend in the change of metabolite pool size between the conditions.

Flux variability analysis (FVA) was used to determine the flux range of each reaction under maximum biomass by subsequently maximizing and minimizing the flux through each reaction. The flux range of each reaction for the N WT, \( gln1-3 \) mutant, and \( gln1-4 \) mutant conditions was compared to the reference N+ WT condition. Flux ranges that did not overlap between one of the N background conditions and the reference condition were further analyzed. These are reactions that must change in response to the limited amount of N or the mutant conditions. Finally, for each condition, the minimum number of reactions which, when not regulated, will restore the biomass to the yield obtained when no “omics” based regulation is applied were determined. This was done by identifying the minimal set of reactions, which are included in the “omics” based regulation, that when active would allow for a biomass yield equivalent to the yield under no “omics” based regulation. This set of reactions represent the reactions whose restriction affects the biomass yield.

The CPLEX solver (version 12.3 IBM ILOG) was used in the GAMS (version 23.3.3, GAMS Development Corporation) environment to solve the optimization problems. The Python programming language is also used during model development (mainly for scripting and data analysis). All computations are carried out on Intel Xeon X5675 Six-Core 3.06 GHz processors constituting the lionxf cluster, which was built and operated by the Research Computing and Cyberinfrastructure Group of The Pennsylvania State University.
Acknowledgements
We are indebted to Isabelle Quilleré, François Gosse and Michel Lebrusq for their technical help.

Supplemental Material
Supplemental Figure S1. Stained micrograph of starch in the N+ WT, gln1-3 mutant and gln1-4 mutant. Thin leaf sections (1 µm) were stained by the periodic acid Schiff's-naphthol blue-black (PAS-NBB) method as described by Sangwan et al. (1992).
Supplemental Figure S2. Glycerolipid synthesis in maize
Supplemental Table S1. Experimental biomass measurements
Supplemental Table S2. Glycerolipid synthesis reactions included in the model
Supplemental Table S3. Second-generation leaf model in excel format
Supplemental Table S4. Second-generation leaf model in SBML format.
Supplemental Table S5. Metabolic reactions with fluxes that must change in a N background condition compared to the N+ WT condition
Supplemental Table S6. Flux-sum levels in each condition
References


highlights the importance of asparagine for nitrogen translocation within sink organs. Plant Biotechnol J 8: 966-978


Driever SM, Kromdijk J (2013) Will C₃ crops enhanced with the C₄ CO₂-concentrating mechanism live up to their full potential (yield)? J Exp Bot 64: 3925-3935


Howard TP, Lloyd JC, Raines CA (2011) Inter-specific variation of the oligomeric state of the higher plant Calvin cycle enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase. J Exp Bot 62: 3799-3805


Kumar VS, Dasika MS, Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics 8: 212


Lepage H (1967) Identification and composition of turnip root lipids. Lipids 2: 244-250


Stitt M, Heldt HW (1985) Generation and maintenance of concentration gradients between the mesophyll and bundle sheath in maize leaves. Biochimica Et Biophysica Acta 808: 400-414


FIGURE LEGENDS

Figure 1. The weight percentage of biomass components. The weight percentage for each class of metabolites experimentally measured contributing to biomass synthesis is displayed. The composition is displayed for the: (A) N+ WT, (B) N- WT, (C) gln1-3 mutant, and (D) gln1-4 mutant conditions. The measurements for specific components within each class of metabolites are shown in Supplemental Table S1.

Figure 2. Number of metabolic and transport reactions distributed between compartments in the bundle sheath and mesophyll cell types. The number of metabolic and transport reactions are shown for each compartment. Integral membrane proteins are counted for the compartment in which the main biotransformation occurs. For example, the ATP synthase associated with the mitochondrial electron transport chain is counted as a metabolic reaction in the mitochondrion, not the inner mitochondrial membrane (IMM).

Figure 3. Number of metabolites in each condition that statistically varied from the N+ WT condition at the vegetative stage. The number of metabolites that experimentally significantly increased (↑) or decreased (↓) in comparison to the N+ WT condition are displayed for each of the N conditions tested (i.e. N- WT, gln1-3 mutant, and gln1-4 mutant conditions). The metabolites are shaded based on whether they are involved in C, N, or other metabolism.

Figure 4. Effect of “omics” based regulation on the flux-sum prediction compared to the experimental trend in metabolite concentration. The accuracy in predicting the increasing (↑) or decreasing (↓) trend in metabolite change between the N background condition and the N+ WT condition is displayed. By restricting the reaction flux based on the transcriptomic and proteomic data, the accuracy of the qualitative trend in metabolite pool size between the N- WT to N+ WT increases. Before adding “omics” based constraints, the model was able to correctly predict the direction of change in 13% of the metabolites measured in the N- WT compared to the N+ WT condition. The accuracy increases to 90% when “omics” based constraints are included. The flux-sum method is not able to accurately represent the gln1-3 and gln1-4 mutant conditions suggesting that
the genetic background affects the ability of the flux-sum method to predict metabolite changes.

**Figure 5.** Model development and curation schematic. The workflow for the second-generation genome-scale metabolic model of the maize leaf is displayed. The data sources give three types of retrieved data (i.e. the raw reaction data, reaction directionality, and compartmentalization) that are then manipulated as shown to create the final model.
**TABLES**

**Table I. Experimental content of classes of metabolites in different conditions**

The biomass components were determined experimentally for each of the conditions (N+ WT, N- WT, *gln1*-3 mutant, and *gln1*-4 mutant). Values are the mean of three replicates unless indicated by *, representing that two replicate measurements were taken. Biomass measurements for the specific metabolites within each class are displayed in Supplemental Table S1.

<table>
<thead>
<tr>
<th>Biomass components</th>
<th>N+ WT</th>
<th>N- WT</th>
<th><em>gln1</em>-3 mutant</th>
<th><em>gln1</em>-4 mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass Yield (g of DW)</td>
<td>61 ± 3.5</td>
<td>15 ± 2</td>
<td>64 ± 4.5</td>
<td>65 ± 1.5</td>
</tr>
<tr>
<td>Soluble Amino Acid Content (mmol g⁻¹ DW⁻¹)</td>
<td>0.0732 ± 0.0170</td>
<td>0.0261 ± 0.0040</td>
<td>0.02124 ± 0.00100</td>
<td>0.09303 ± 0.00640</td>
</tr>
<tr>
<td>Protein Content (mg g⁻¹ DW⁻¹)</td>
<td>132.6 ± 0.7</td>
<td>58 ± 3</td>
<td>125.25 ± 1.7</td>
<td>140.39 ± 5.72</td>
</tr>
<tr>
<td>Fatty Acid Content (mg g⁻¹ DW⁻¹)</td>
<td>43.3 ± 4.4</td>
<td>16.6 ± 2.2</td>
<td>45.1*</td>
<td>16.3 ± 1.1</td>
</tr>
<tr>
<td>Starch Content (mmol g⁻¹ DW⁻¹)</td>
<td>0.152 ± 0.005</td>
<td>0.199 ± 0.007</td>
<td>0.085 ± 0.011</td>
<td>0.107 ± 0.006</td>
</tr>
<tr>
<td>RNA Content (mg g⁻¹ DW⁻¹)</td>
<td>3.78 ± 0.19</td>
<td>0.92 ± 0.10</td>
<td>1.05 ± 0.09</td>
<td>1.77 ± 0.11</td>
</tr>
<tr>
<td>DNA Content (mg g⁻¹ DW⁻¹)</td>
<td>8.315 ± 0.270</td>
<td>2.53 ± 0.10</td>
<td>9.62 ± 0.22</td>
<td>5.48 ± 0.13</td>
</tr>
<tr>
<td>Soluble Carbohydrate Content (mmol g⁻¹ DW⁻¹)</td>
<td>0.235 ± 0.012</td>
<td>0.112 ± 0.013</td>
<td>0.198 ± 0.041</td>
<td>0.193 ± 0.023</td>
</tr>
<tr>
<td>Cell Wall Carbohydrate Content (mg g⁻¹ DW⁻¹)</td>
<td>0.32 ± 0.03</td>
<td>0.26 ± 0.09</td>
<td>0.187 ± 0.017</td>
<td>0.29 ± 0.07</td>
</tr>
<tr>
<td>Chlorophyll Content (mg g⁻¹ DW⁻¹)</td>
<td>1.87 ± 0.16</td>
<td>0.69 ± 0.06</td>
<td>1.71 ± 0.14</td>
<td>1.85 ± 0.08</td>
</tr>
<tr>
<td>Total Nitrogen (% of gDW)</td>
<td>4.36 ± 0.08</td>
<td>1.80 ± 0.15</td>
<td>4.28 ± 0.10</td>
<td>4.31 ± 0.12</td>
</tr>
</tbody>
</table>
Table II. Number of reactions after each model creation and curation step

The original two data sets are the core set and gapfill set which combine to form the final model statistics. The total number of metabolic, transport, exchange, and biomass reactions are displayed after each process during model curation. Metabolic reaction totals include duplication from compartmentalization.

<table>
<thead>
<tr>
<th>Data Processing</th>
<th>Metabolic Reactions</th>
<th>Transport Reactions</th>
<th>Exchange Reactions</th>
<th>Biomass Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Set</td>
<td>3002</td>
<td>418</td>
<td>82</td>
<td>85</td>
</tr>
<tr>
<td>Core Set + Manually Created Pathways</td>
<td>3264</td>
<td>469</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Non-Core Set</td>
<td>18951</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Processes Performed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Gapfill</td>
<td>3971</td>
<td>1198</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Compartmentalization</td>
<td>9005</td>
<td>1198</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Manually Determined</td>
<td>7033</td>
<td>1115</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Compartmentalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thermodynamically</td>
<td>7040</td>
<td>1115</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Infeasible cycles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similar Organism</td>
<td>7040</td>
<td>1115</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Gapfill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Model</td>
<td>7040</td>
<td>1115</td>
<td>285</td>
<td>85</td>
</tr>
<tr>
<td>Second-generation GSM model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table III. Summary of reactions that affect biomass synthesis.

The minimum set of reactions that are down-regulated as a result of the inclusion of proteomic and transcriptomic data and affect biomass synthesis are displayed. The corresponding condition is displayed for each reaction as well as the role of the reaction.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Condition(s) affected</th>
<th>Role of the reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol oxidoreductase/Catalase</td>
<td>N⁺ WT</td>
<td>Produces acetaldehyde alleviating flux through pyruvate decarboxylase</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>gln1-3 mutant</td>
<td>Participates in glycolysis and carbon fixation, but not required as 3-phospho-D-glycerate kinase can restore flux to 1,3-bisphospho-D-glycerate</td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase</td>
<td>gln1-3 mutant</td>
<td>Participates in the Calvin-Benson-Bassham cycle, but can be bypassed through the sedoheptulose 1,7-bisphosphate/D-glyceraldehyde-3-phosphate lyase reaction</td>
</tr>
<tr>
<td>Fructose bisphosphatase</td>
<td>gln1-3 mutant</td>
<td>Decreases ATP/ADP ratio, switches metabolism from sucrose to starch synthesis, and inhibits photosynthesis at high CO₂ levels in Arabidopsis</td>
</tr>
<tr>
<td>Ribose-5-phosphate isomerase</td>
<td>gln1-4 mutant</td>
<td>Affects cellulose synthesis in Arabidopsis</td>
</tr>
</tbody>
</table>
Table IV.

Number of gene transcripts, proteins, and metabolites that significantly vary

The WT condition for each study was combined to create one uniform WT condition. The number of gene transcripts, proteins, and metabolites that statistically vary are displayed below.

<table>
<thead>
<tr>
<th>Type Of Data</th>
<th>WT Condition</th>
<th>N-Condition</th>
<th>gln1-3 Mutant</th>
<th>gln1-4 Mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcriptomic</td>
<td>256</td>
<td>76</td>
<td>102</td>
<td>53</td>
</tr>
<tr>
<td>Proteomic</td>
<td>38</td>
<td>14</td>
<td>29</td>
<td>-</td>
</tr>
<tr>
<td>Metabolomic</td>
<td>83</td>
<td>20</td>
<td>31</td>
<td>13</td>
</tr>
</tbody>
</table>
Figure 1. The weight percentage of biomass components. The weight percentage for each class of metabolites experimentally measured contributing to biomass synthesis is displayed. The composition is displayed for the: (A) N\textsuperscript{+} WT, (B) N\textsuperscript{-} WT, (C) gln1-3 mutant and, (D) gln1-4 mutant conditions. The measurements for specific components within each class of metabolites are shown in Supplemental Table S1.
Figure 2. Number of metabolic and transport reactions distributed between compartments in the bundle-sheath and mesophyll cell types. The number of metabolic and transport reactions are shown for each compartment. Integral membrane proteins are counted for the compartment in which the main biotransformation occurs. For example, the ATP synthase associated with the mitochondrial electron transport chain is counted as a metabolic reaction in the mitochondrion, not the inner mitochondrial membrane (IMM).
Figure 3. Number of metabolites in each condition that statistically varied from the N⁺ WT condition at the vegetative stage. The number of metabolites that experimentally significantly increased (↑) or decreased (↓) in comparison to the N⁺ WT condition are displayed for each of the N conditions tested (i.e. N⁻ WT, gln1-3 mutant, and gln1-4 mutant conditions). The metabolites are shaded based on whether they are involved in C, N, or other metabolism.
Figure 4. Effect of “omics” based regulation on the flux-sum prediction compared to the experimental trend in metabolite concentration. The accuracy in predicting the increasing (↑) or decreasing (↓) trend in metabolite change between the N background condition and the N WT condition is displayed. By restricting the reaction flux based on the transcriptomic and proteomic data, the accuracy of the qualitative trend in metabolite pool size between the N WT to N WT increases. Before adding “omics” based constraints, the model was able to correctly predict the direction of change in 13% of the metabolites measured in the N WT compared to the N WT condition. The accuracy increases to 90% when “omics” based constraints are included. The flux-sum method is not able to accurately represent the gln1-3 and gln1-4 mutant conditions suggesting that the genetic background affects the ability of the flux-sum method to predict metabolite changes.
**Figure 5.** Model development and curation schematic. The workflow for the second-generation genome-scale metabolic model of the maize leaf is displayed. The data sources give three types of retrieved data (i.e. the raw reaction data, reaction directionality, and compartmentalization information) that are then manipulated as shown to create the final model.