PLoS one

Zea mays iRS1563: A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism

Rajib Saha, Patrick F. Suthers, Costas D. Maranas*

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, United States of America

Abstract

The scope and breadth of genome-scale metabolic reconstructions have continued to expand over the last decade. Herein, we introduce a genome-scale model for a plant with direct applications to food and bioenergy production (i.e., maize). Maize annotation is still underway, which introduces significant challenges in the association of metabolic functions to genes. The developed model is designed to meet rigorous standards on gene-protein-reaction (GPR) associations, elementally and charged balanced reactions and a biomass reaction abstracting the relative contribution of all biomass constituents. The metabolic network contains 1,563 genes and 1,825 metabolites involved in 1,985 reactions from primary and secondary maize metabolism. For approximately 42% of the reactions direct literature evidence for the participation of the reaction in maize was found. As many as 445 reactions and 369 metabolites are unique to the maize model compared to the AraGEM model for *A. thaliana*. 674 metabolites and 893 reactions are present in *Zea mays* iRS1563 that are not accounted for in maize C4GEM. All reactions are elementally and charged balanced and localized into six different compartments (i.e., cytoplasm, mitochondrion, plastid, peroxisome, vacuole and extracellular). GPR associations are also established based on the functional annotation information and homology prediction accounting for monofunctional, multifunctional and multimeric proteins, isozymes and protein complexes. We describe results from performing flux balance analysis under different physiological conditions, (i.e., photosynthesis, photorespiration and respiration) of a C4 plant and also explore model predictions against experimental observations for two naturally occurring mutants (i.e., *bm1* and *bm3*). The developed model corresponds to the largest and more complete to-date effort at cataloguing metabolism for a plant species.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: costas@psu.edu

Introduction

Zea mays, commonly known as maize or corn, is a plant organism of paramount importance as a food crop, biofuel production platform and a model for studying plant genetics [1]. Maize accounts for 31% of the world production of cereals occupying almost one-fifth of the worldwide land dedicated for cereal production [2]. Maize cultivation led to 12 billion bushels of grain in the USA alone in 2008 worth $47 billion [3]. Maize is the second largest crop, after soybean, used for biotech applications [2]. In addition to its importance as a food crop, 3.4 billion gallons of ethanol was produced from maize in 2004 [3]. Maize derived ethanol accounts for 99% of all biofuels produced in the United States [3]. However, currently nearly all of this bioethanol is produced from corn seed [4]. Ongoing efforts are focused on developing and commercializing technologies that will allow for the efficient utilization of plant fiber or cellulosic materials (e.g. maize stover and cereal straws) for biofuel production. Maize is the most studied species among all grasses with respect to cell wall lignification and digestibility, which are critical for the efficient production of cellulosic biofuels [5]. A thorough evaluation of the metabolic capabilities of maize would be an important resource to address challenges associated with its dual role as a food (e.g., starch storage) and biofuel crop (e.g., cell wall deconstruction).

This decade we witnessed significant advancements towards mapping plant genes to metabolic functions culminating with the complete genome sequencing and partial annotation of a number of plant species, namely, *Arabidopsis thaliana* [6], *Oryza Sativa* [7,8], *Sorghum bicolor* [9], *Zea mays* [10] and *Theobroma cacao* [11]. Nevertheless, attempts to engineer plant metabolism for desired overproductions have been met with only limited success [12]. Genetic modifications seldom bring about the expected/desired effect in plant metabolism primarily due to the built-in metabolic redundancy circumventing the imposed genetic changes [13,14]. This necessitates the development of genome-wide comprehensive metabolic reconstructions capable of taking account of the complete inventory of metabolic transformations of a given plant organism.

Genome-scale metabolic reconstructions are available for an increasing number of organisms [15,16]. At least 40 bacterial, 2 archaeal and 15 eukaryotic reconstructions are available to-date [12,15,17,18] while many others are under development. Recently Poolman et al (2009) and Dal'Molin et al (2010) independently constructed the first two genome-scale metabolic reconstructions for a plant organism (i.e., *Arabidopsis thaliana*). The model by Dal'Molin *et al* identifies the set of essential reactions, accounts for the classical photorespiratory cycle and highlights the significant differences between photosynthetic and non-photosynthetic me-

tabolism. The model by Poolman *et al* includes ATP demand constraints for biomass production and maintenance and suggests strategies for the construction of metabolic modules as a consequence of variation in ATP requirement. Both models make a significant step forward towards assessing the metabolic capabilities of plants establishing production routes for key biomass precursors and major pathways of Arabidopsis primary metabolism. In addition, two recent efforts involved the reconstruction of plant models with an emphasis on specific physiological conditions or tissue types [19,20]. Model C4GEM [20] focused on C4 plants such as maize, sugarcane and sorghum and investigated flux distributions in mesophyll and bundle sheath cells during C4 photosynthesis. Grafahrend-Belau *et al* developed a metabolic network of only primary metabolism in barley seeds and studied grain yield and metabolic fluxes under a variety of oxygen availability scenarios and genetic manipulations [19]. Pilalis *et al.* reconstructed a multi-compartmental model of the central metabolism of *Brassica napus* (Rapeseed) and simulated seed growth during the stage of oil accumulation and subsequently studied network properties of seed metabolism via Flux Balance Analysis, Principal Component Analysis and reaction deletion studies [21].

In this paper, we describe the construction of a genome-scale in silico model of maize metabolism (i.e., *Zea mays* iRS1563). This is, to the best of our knowledge, the first attempt of globally characterizing the metabolic capabilities (both primary and secondary metabolism) using a compartmentalized photosynthetic model of an important crop and energy plant species. The development of a genome-scale model for maize is a significant challenge due to its genome size which is 14 times larger [10] than that of *Arabidopsis thaliana* (157 million base pairs) [22]. The constructed model contains 1,563 genes and 1,825 metabolites participating in 1,985 reactions from both primary and secondary metabolism of maize. For 42% of the reaction entries direct literature evidence in addition to homology criteria for their inclusion to the model was identified. We found that as many as 676 reactions and 441 metabolites are unique to *Zea mays* iRS1563 in comparison to the AraGEM model by Dal'Molin *et al*. We chose the AraGEM model as a basis of comparisons as at the onset of this study it was the most comprehensive genome-scale compartmentalized model of a plant species capable of recapitulating basic plant physiological states. In order to deduce the genuine differences between maize and Arabidopsis irrespective of annotation chronology we also reconstructed an up-to-date model of Arabidopsis, *A. thaliana* iRS1597. *A. thaliana* iRS1597 contains 1597 genes, 1798 reactions and 1820 metabolites. In comparison to *A. thaliana* iRS1597, *Zea mays* iRS1563 has 445 new reactions and 369 new metabolites. Notably, 893 reactions and 674 metabolites are included in *Zea mays* iRS1563 that are absent from the maize C4GEM model. All reactions present in Zea mays iRS1563 are elementally and charged balanced and localized into six compartments including cytoplasm, mitochondrion, plastid, peroxisome, vacuole and extracellular space. Provisions for accounting that photosynthesis in maize (i.e., a C4 plant) occurs in two separate cell types (i.e., mesophyll cell and bundle sheath cell) are included in the model. GPR associations are delineated from the available functional annotation information and homology prediction accounting for monofunctional, multifunctional and multimeric proteins, isozymes and protein complexes. A biomass equation is established that quantifies the relative abundance of different constituents of dry plant cell biomass. Biomass production under three different physiological states (i.e., photosynthesis, photorespiration and respiration) is demonstrated and the model is tested against experimental data for two naturally occurring maize mutants (i.e., bm1 and bm3).

Results

The metabolic model reconstruction process follows three major steps: (1) Reconstruction of draft model via automated homology searches for the identification of native biotransformations; (2) Generation of a computations-ready model after defining biomass equation and system boundary and establishing GPR; (3) Model refinement via GapFind and GapFill [23] to unblock biomass precursors as well as reconnect unreachable metabolites. Upon construction of the model, key features such as physiological constraints, network connectivity, light reactions, carbon fixation and secondary metabolism and uniqueness compared to AraGEM and maize C4GEM are described. In addition, model predictions are contrasted against experimental observations.

Construction of Auto and Draft models

The B73 maize genome [10] has 32,540 genes and 53,764 transcripts in the Filtered Gene Set (FGS). Out of 32,540 genes, 30,599 (93%) are evidence-based [24], while the remaining 2,141 (7%) are predicted by the Fgenesh program [25]. 13,726 genes (42% of total) do not have any functional annotation information or are identified as proteins with no or hypothetical/putative functions. Of the remainder, 1,361 (7%) genes encode proteins that do not participate in specific metabolic transformations but rather are involved in transcription, signal transduction, DNA repair, DNA binding, DNA/RNA polymerization, protein folding and adhesion. Because the B73 maize genome is not completely annotated we first established Gene-Protein-Reaction (GPR) mappings for the AraGEM genome-scale model of *A. thaliana* [12] to be used as a proxy. Using these GPRs as a point of comparison we next identified Arabidopsis gene orthologs in maize and transferred the corresponding GPRs via the AUTO-GRAPH method [26]. This step was followed by annotation of the remainder maize genes by bidirectional protein BLAST (i.e., BLASTp) searches against the NCBI non-redundant (nr) database. Out of a total of 1,567 metabolic or transport reactions of AraGEM, GPRs were established for 1,254 reactions via 1,467 genes and 653 enzymes by making use of information from several online databases such as AraCyc, KEGG, Uniprot and Brenda (see File S1). Bidirectional BLASTp searches for each one of the 1,467 genes included in AraGEM model were carried out against the B73 maize genome using a stringent cutoff value of 10^{-30}. This fully automated process generated an initial model, termed as 'Automodel', containing 946 genes and 1,365 unique metabolites participating in 1,186 reactions (see Table 1 and File S2) exclusively derived from AraGEM. Out of 1,186 reactions, 32 are inter-organelle transport reactions for which homologs were found in maize.

Genes not included in the automodel were scrutinized further by comparing them against the NCBI non-redundant protein database using the same BLASTp cut-off. This increased the model size to 1,485 genes and 1,703 unique metabolites involved in 1,667 reactions by pulling functionalities absent in AraGEM. This is referred to as the 'Draft model' (see Table 1 and Files S2 and S3). As described in Table 2, orthologous genes were found in *Oryza Sativa* (Rice), *Arabidopsis thaliana* (Arabidopsis), *Sorghum bicolor* (Sorghum) and less frequently in other plant species such as wheat, tobacco, spinach, soya bean, etc. (See File S3). Notably, 802 orthologous genes from *A. thaliana* were added in the model *Zea mays* iRS1563 that were absent from AraGEM primarily due to recent annotation updates. Reactions associated with these genes were subsequently extracted from on-line databases such as KEGG and BRENDA. Table 2 shows the total number of reactions as well as the number of new reactions included in the

Table 1. Model size after each reconstruction step.

	Auto model	Draft model	Functional model	Final model
Included genes	946	1,485	1,552	1,563
Proteins	472	714	774	876
Single functional proteins	178	322	381	463
Multifunctional proteins	92	150	153	170
Protein complexes	0	4	4	4
Isozymes	21	36	36	36
Multimeric proteins	87	140	148	148
Others[a]	94	62	62	55
Reactions	1,186	1,667	1,821	1,985
Metabolic reactions	1,154	1,635	1,739	1,900
Transport reactions	32	32	67	70
GPR associations				
Gene associated (metabolic/transport)	1,100	1,581	1,635	1,668
Nonenzyme associated (metabolic/transport)	86	86	86	86
Spontaneous[b]	0	0	7	41
Nongene associated (metabolic/transport)	0	0	78	175
Exchange reactions	0	0	15	15
Metabolites[c]	1,365	1,703	1,769	1,825
Cytoplasmic	1,309	1,643	1,689	1,744
Plastidic	91	102	114	115
Peroxisomic	67	69	92	93
Mitochondrial	60	82	86	86
Vaccuolic	5	5	5	5
Extracellular	0	0	15	15

[a]Others include proteins involve in complex relationships, e.g. multiple proteins act as protein complex which is one of the isozymes for any specific reaction.
[b]Spontaneous reactions are those without any enzyme as well as gene association.
[c]Unique metabolites irrespective of their compartmental location.
doi:10.1371/journal.pone.0021784.t001

draft model. Seven reactions having KEGG reaction IDs R00379, R00381, R06023, R06049, R06082, R06138 and R06209 were excluded since they involve generic groups and were not elementally fully defined. Figure 1 shows the distribution of the newly added reactions in the draft model based on their orthologous gene of origin.
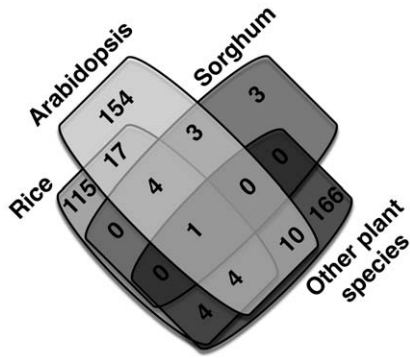
Generation of computations-ready model

A computations-ready model requires a fully characterized biomass equation, assignment of metabolites to reactions, establishment of GPR associations, localization of reactions in compartment(s), and inclusion of intra- and extracellular transport reactions [27].

(i) Establishing a fully characterized biomass equation. A biomass equation that drains all necessary precursors present in maize was derived (see File S4 and Table 3). We used the biomass composition of young and vegetative maize plants as measured by Penningd *et al.* and expressed on a dry weight basis [28]. The amino acid and lignin composition were derived based on the data from [29,30]. The composition of hemicellulose was approximated using data for Orchard Grass [31], another monocot grass species, as no corresponding information was found for maize. Based on these compositions we also defined aggregate reactions such as 'Amino acid synthesis', 'Protein synthesis', 'Carbohydrate synthesis', 'Hemicellulose synthesis', 'Lignin synthesis', 'Lipid synthesis',

Table 2. Maize gene annotation via bidirectional BLASTp homology searches against NCBI non-redundant protein database.

Species	Number of orthologs	Number of associated reactions	Number of newly added reactions in draft model
Oryza Sativa (Rice)	4,109	312	145
Other plant species	833	214	185
Arabidopsis Thaliana (Arabidopsis)	802	258	193
Sorghum Bicolor (sorghum)	47	20	11

doi:10.1371/journal.pone.0021784.t002

Figure 1. Species origin of newly added reactions in the Draft model.
doi:10.1371/journal.pone.0021784.g001

'Material synthesis', 'Nitrogenous compound synthesis', 'Nucleic acid synthesis' and 'Organic acid synthesis' to produce necessary biomass precursors (i.e., amino acids, protein, carbohydrates, hemicellulose, lignin, lipids, materials, nitrogenous compounds, nucleic acids and organic acids respectively). The biomass equation also contains a non-growth associated ATP maintenance as in the latest Arabidopsis model AraGEM [12].

(ii) Assignments of genes, reactions, metabolites and compartments. All metabolic and inter-organelle transport reactions in the draft model have full gene associations. During this step all reactions were elementally balanced and metabolites were assigned appropriate protonation states corresponding to a physiological pH of 7.2. We included an additional 86 reactions to the model without enzyme association information based on direct literature evidence [12]. For example, reactions with KEGG IDs R08053, R08054 and R08055 involved in chlorophyll metabolism are included in the model. Reaction localization information for maize can in some cases be found in database PPDB (a plant proteome database of maize and Arabidopsis) [32]. Because only limited reaction localization information exists for maize, we adopted the compartment or organelle reaction location of the corresponding orthologous gene/enzyme in Arabidopsis using the Arabidopsis Subcellular Database, SUBA [33] and also PPDB [32]. As in AraGEM, reactions for which no such information is available we assumed that they are present only in the cytoplasm.

(iii) Identification of system boundary. The entire reaction network (i.e., system boundary) was distributed across five different intracellular organelles enveloped by the cytoplasmic membrane. Exchange reactions were added in the model to ensure that gaseous metabolites (i.e., carbon dioxide and oxygen), inorganic nutrient metabolites (i.e., nitrate, ammonia, hydrogen sulfide, sulfate, phosphate, potassium and chloride), sugar metabolites (i.e., glucose, fructose, maltose and sucrose), water and photons could enter and leave the system whenever necessary depending on the physiological state. As shown in Table 4, constraints on these exchange reactions as well as reactions involved with enzyme RuBisCO (Ribulose-1, 5-bisphosphate carboxylase oxygenase) were established to define three different physiological states (i.e., photosynthesis, photorespiration and respiration) by allowing the selective uptake/release of certain metabolites. Even though photorespiration is limited in C4 plants (i.e., maize, sorghum, etc.), literature evidence [34,35,36] alludes that it is still present. Therefore, we made sure that the model is capable of simulating this condition.

The stoichiometric matrix of the draft model (see Table 1) contains 1,901 rows (i.e., total metabolites after taking account of their compartmental appearance) and 1,682 columns (i.e., metabolic reactions, inter-organelle transport reactions and exchange reactions). 970 reactions have one-to-one GPR associations whereas 712 map to more than one gene. 532 reactions map to both isozymes and protein complexes while 4 of them map to only protein complexes, 36 to only isozymes, and 140 to only multimeric proteins.

Table 3. Biomass component list in *i*RS1563.

Major components	Protein	Carbohydrates	Lipids	Ions
Nitrogenous compounds	L-alanine	ribose	glyceroltripalmitate	potassium
Carbohydrates	L-arginine	glucose	gleceroltristearate	chloride
Lipids	L-aspartic acid	fructose	glyceroltrioleate	
Lignin	L-cystine	mannose	glyceroltrilinolate	**RNA**
Organic acids	L-glutamic acid	galactose	glyceroltrilinoleate	ATP
Ions	L-glycine	sucrose		GTP
	L-histidine	cellulose	**Lignin**	CTP
	L-isoleucine	hemicellulose	4-coumaryl alcohol	UTP
Nitrogenous compounds	L-leucine	pectin	coniferyl alcohol	
amino acids	L-lysine		sinapyl alcohol	**DNA**
protein	L-methionine			dATP
nucleic acids	L-phenylalanine	**Hemicellulose**	**Organic acids**	dGTP
	L-proline	arabinose	oxalic acid	dCTP
	L-serine	xylose	glyoxalic acid	dUTP
	L-threonine	mannose	Oxalo-acetic acid	
	L-tryptophan	galactose	Malic acid	
	L-tyrosine	glucose	Citric acid	
	L-valine	uronic acids	aconitic acid	

doi:10.1371/journal.pone.0021784.t003

Table 4. Definition of three different physiological states.

Constraints	Photosynthesis (PS)	Photorespiration (PR)	Respiration (R)
CO_2 transport	Uptake	Uptake	Release
Sucrose transport	Disabled	Disabled	Uptake
Photon transport	Uptake	Uptake	Disabled
H_2O transport	Uptake	Uptake	Uptake
Inorganic nutrient transport	Uptake	Uptake	Uptake
O_2 transport	Release	Unconstrained	Uptake
RUBISCO: EC 4.1.1.39	Carboxylation	Carboxylation:Oxygenation = 3:1	Both disabled

doi:10.1371/journal.pone.0021784.t004

Network connectivity analysis and restoration

The draft metabolic model inherently contained gaps, unreachable metabolites, omitted transport mechanisms and missing biomass components. We used the procedures termed GapFind and GapFill [37] to correct for these pathologies. We first concentrated on resolving problems with the participation of components in the biomass equation followed by network connectivity.

We found that 723 out of the 1,683 reactions in the draft model could not carry any flux (i.e., blocked reactions) under any of the relevant three physiological states (e.g. photosynthesis (PS), photorespiration (PR) and respiration (R)). As a result, these blocked reactions prevented the formation of some of biomass precursors. GapFind [37] revealed that only 21 out of 64 biomass components could be synthesized using the draft model. GapFill [37] was applied for bridging the gaps through the addition of metabolic and inter-organelle transport reactions and the relaxing of irreversible of existing reactions in the model. GapFill suggested the addition of 94 metabolic and 35 inter-organelle transport reactions in the model to unblock the production of all 64 biomass components. These putative additions to the model were tested by performing an additional round of BLASTp searches for the corresponding genes against the maize genome. We found that 54 (out of 93) metabolic reactions could be assigned to maize gene(s) if the expectation value cut-off for BLASTp was lowered to 10^{-5}. In light of the critical need of restoring biomass formation the less stringent cut-off for inclusion was accepted for these genes. Addition of these reactions ensured the production of biomass under all relevant physiological states validating the use of the term 'Functional' for the updated model (see Table 1).

Upon ensuring biomass formation GapFind was also applied to assess network connectivity and 715 blocked metabolites were

Table 5. Restoration of network connectivity using GapFill [36].

Number of metabolites	Number of blocked metabolites: before applying GapFill	Number of blocked metabolites: after applying GapFill
Cytosolic (1744)	680	382
Plastidic (115)	28	11
Peroxisomic (93)	5	0
Mitochondrial (86)	2	0

doi:10.1371/journal.pone.0021784.t005

found in the functional model. By applying GapFill connectivity of 322 (45%) blocked metabolites was restored through the addition of 159 metabolic and 3 inter-organelle transport reactions. Table 5 shows the distribution of blocked metabolites into four intracellular organelles before and after applying GapFill. BLASTp searches allowed us to assign 31 (20% of GapFill suggestions) metabolic reactions with specific maize genes (File S2). Biological evidence of the occurrence of such additional reactions in maize or other plant species was sought whenever possible. For example, as shown in Figure 2 phenylacetaldehyde appears to be a "no-consumption" [37] metabolite in the functional model as no reaction can consume it. Using GapFill we found a homolog in maize (i.e., BLASTp score of 10^{-24}) and also literature evidence [38] that *Arabidopsis thaliana* has a aldehyde dehydrogenase activity that catalyzes the conversion of phenylacetaldehyde to phenylacetic acid. Hence, by adding this chemical transformation to *Zea mays* iRS1563 a consumption pathway for phenylacetaldehyde is established. After adding these reactions to the functional model and following charge and elemental balancing and GPR association checking the 'Final' *Zea mays* iRS1563 model (see Table 1) is derived.
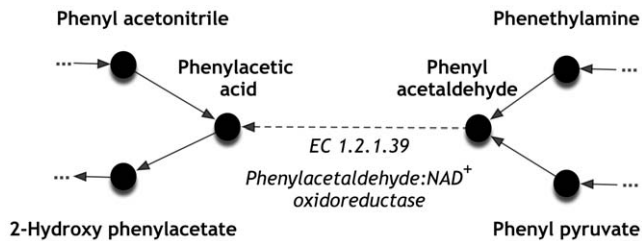
Zea mays iRS1563 model

The *Zea mays* iRS1563 metabolic reconstruction contains 1,825 unique metabolites and 1,985 reactions associated with 1,563 genes and 876 proteins. Of these reactions 1,898 are metabolic reactions, 70 are inter-organelle transport reactions and 15 are exchange reactions between intra- and extracellular environments. GPR associations are established for all entries (see Table 1). Notably, we identified that the fraction of multifunctional proteins (19% of the total number of proteins) in *Zea mays* iRS1563 is similar to the ratio found in *E. coli* [39]. *Zea mays* iRS1563 accounts for the metabolic functions for all three physiological states. Photosynthetic as well as photorespiration metabolism was modelled by including light mediated ATP and NADPH production via separate charged balanced reactions in the electron transfer system of the thylakoid membrane [40]. Furthermore, the ratio of fluxes for the carboxylation and oxidation reactions associated with enzyme RuBisCO was kept at 1:0 thus ensuring complete carbon fixation during photosynthesis. This ratio was shifted to 3:1 during photorespiration to model simultaneous carbon fixation and oxidation [41]. Because sucrose is the main growth substrate during respiration for higher plants [42], the aforementioned reactions were inactivated and the exchange reaction for sucrose uptake was activated. Under all these three conditions, inorganic nutrients required for plant growth, e.g. sulfate, nitrate, ammonia, hydrogen sulfide, phosphate, potassium

Figure 2. Example of connectivity restoration for phenylace-taldehyde.
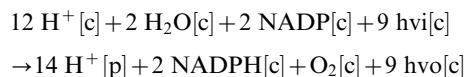doi:10.1371/journal.pone.0021784.g002

and chloride, were allowed to be freely taken up from the environment via extracellular exchange reactions.

The participation of *Zea mays* iRS1563 metabolites across different compartments is shown in Figure 3. The five intracellular organelles differ notably in terms of mutual connectivity, metabolite uniqueness and number of metabolites. As shown in Figure 3a, approximately 90% of these metabolites are unique to cytoplasm. In addition, cytoplasm contains all metabolites shared between any two organelles because any metabolite needs to be transported through cytoplasm in order to be exchanged between organelles. Among the remaining metabolites, cytoplasm shares the highest number with the plastid (i.e., 63) where photosynthesis and photorespiration occur. It also shares a significant number of metabolites with mitochondrion (i.e., 27) and peroxisome (i.e., 22) that are involved in energy production and fatty acid biosynthesis, respectively. Figure 3b shows the distribution of other non-cytoplasmic *Zea mays* iRS1563 metabolites in terms of how many organelles they participate.

Light reactions, carbon fixation and secondary metabolism

In plants photosynthesis reactions include light dependent and light independent or carbon fixation reactions [43]. *Zea mays* iRS1563 includes charged balanced light reactions culled from a number of literature sources [40,44,45,46]. The overall photosynthesis reaction cascade produces two NADPH, three ATP and one O_2 whenever nine photons are absorbed and fourteen H^+ are transferred via the electron-transport system. This defines the following overall balance equations:

$$12 \ H^+[c] + 2 \ H_2O[c] + 2 \ NADP[c] + 9 \ hvi[c]$$
$$\rightarrow 14 \ H^+[p] + 2 \ NADPH[c] + O_2[c] + 9 \ hvo[c]$$

$$3 \ ADP[c] + 14 \ H^+[p] + 3 \ Pi[c] \rightarrow 3 \ ATP[c] + 14 \ H^+[c]$$

Here, [c] and [p] represent cytoplasm and plastid and hvi and hvo signify input and output photons respectively. Carbon fixation in maize (C_4 plant) is more complex compared to Arabidopsis or other C_3 plants [43]. *Zea mays* iRS1563 captures these differences by accounting for (i) direct carboxylation of phosphoenol pyruvate and CO_2 fixation to form C_4 acids such as oxaloacetic acid [ATP: oxaloacetate carboxy-lyase *(ocl)*] and malic acid [Oxaloacetate: NADPH hydrogenase *(oha)*] in mesophyll cells, (ii) transport of malic acid from mesophyll cell to bundle-sheath cells, (iii) decarboxylation of malic acid [Malate:NADP+ oxidoreductase *(mor)*] in bundle-sheath cells to produce pyruvic acid and CO_2, which enters the Calvin cycle, (iv) transport of pyruvic acid from bundle-sheath cells to mesophyll cells, and (v) production of
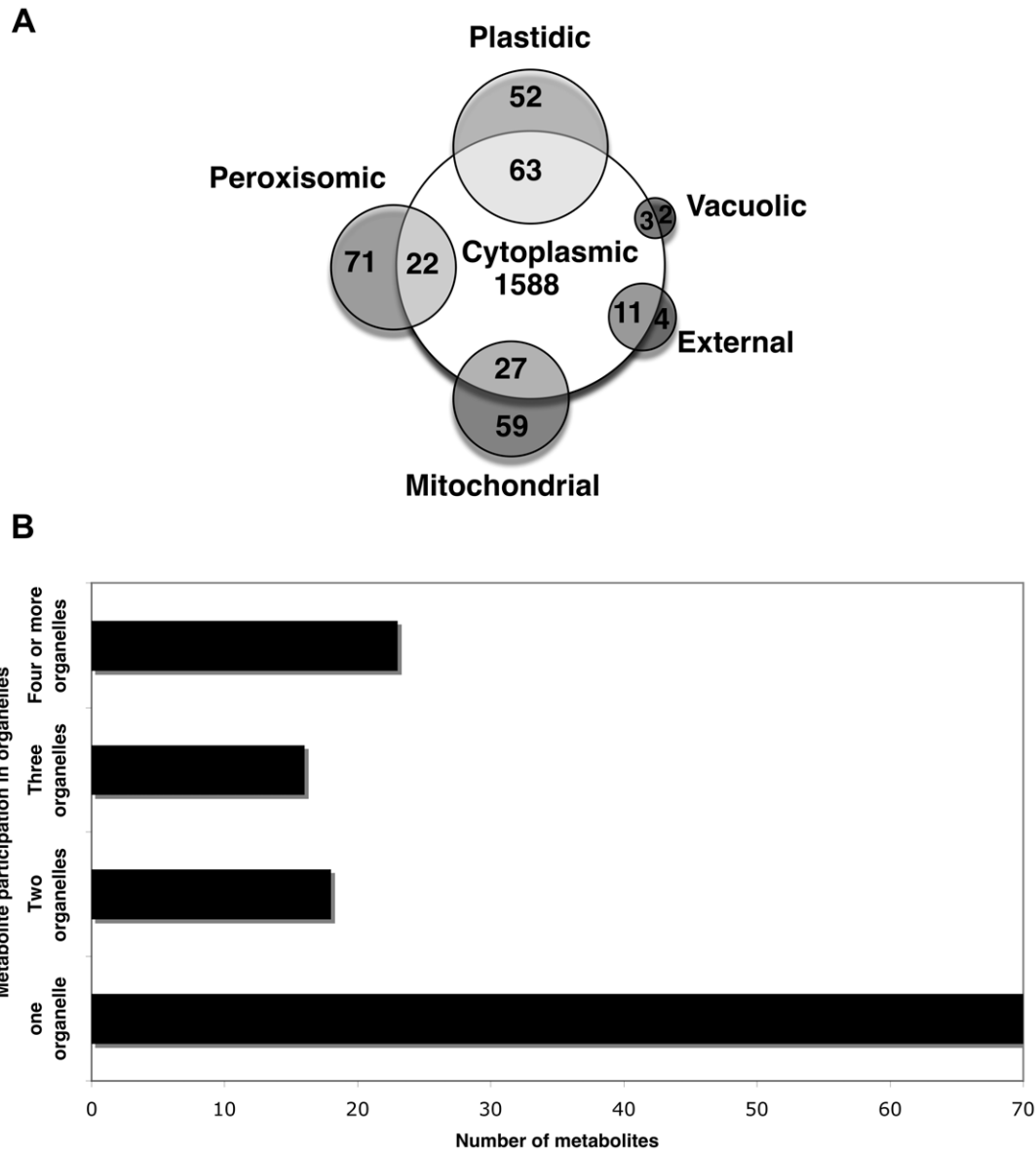
phosphoenol pyruvic (i.e., C_3) acid [ATP:pyruvate,phosphate phosphotransferase *(ppt)*] from pyruvic acid [43]. Figure 4, pictorially shows the localization of reactions and organelles between mesophyll and bundle sheath cells. In addition, to differences in carbon fixation reactions, the peroxisome activity is primarily present in bundle-sheath cells and largely absent from mesophyll cells [47]. Based on this localization information a standalone metabolic model can be developed for the photosynthetic tissue of maize. Because RuBisCO that operates in the Calvin cycle cannot come in direct contact with atmospheric oxygen during day time (see Figure 4), photorespiration is restricted providing an advantage for survival in hot and arid environments for maize and other C_4 plants. This comes at the expense of higher (ATP) requirements as C_4 carbon fixation involves additional steps [43].

In addition to photosynthesis, secondary metabolism plays a key role in the physiology of maize. For example, phenylpropanoid metabolism produces monolignols (i.e., *p*-coumaroyl alcohol, coniferyl alcohol and sinapyl alcohol) that are used in the generation of three major lignin subunits H-lignin, G-lignin and S-lignin, respectively [48]. Many of these enzymes such as hydroxycinnamoyl transferase (HCT), ferulate 5-hydroxylase (F5H) and caffeic acid 3-*O*-methyltranferase (COMT) along with their associated reactions are unique to C_4 plants and are not present in the lignin biosynthesis pathways of *A. thaliana* [48]. HCT is involved in the early stages of lignin biosynthesis by controlling the flux from *p*-coumaroyl-CoA towards caffeoyl-CoA while F5H and COMT regulate fluxes from coniferaldehyde and coniferyl alcohol to sinapaldehyde and sinapyl alcohol, respectively [48]. *Zea mays* iRS1563 contains all these enzymes and associated reactions thus providing a comprehensive lignin biosynthesis pathway for a C_4 plant.

In addition to phenylpropanoid metabolism, *Zea mays* iRS1563 provides a detailed description of flavonoid biosynthesis pathways. Flavonoids are pigments occurring in plant as secondary metabolites and mostly function in the recruitment of pollinators and/or seed dispersers [49]. For example, maize is known to produce 3-deoxyanthocyanins, which are a specialized class of flavonoids [50,51]. *Zea mays* iRS1563 contains the dihydroflavonol 4-reductase (DFR) enzyme that catalyzes the reaction for flavan-4-ols biosynthesis that channels flux towards 3-deoxyanthocyanins production [51]. The model also accounts for isoflavone 7-O-glucosyltransferase (IF7GT) and associated reactions that are involved in the production of necessary intermediates for pterocarpin phytoalexin conjugates such as medicarpin 3-O-glucoside-6′-O-malonate (MeGM) and maackain 3-O-glucoside-6′-O-malonate (MaGM) involved in plant defense against fungal elicitation [52].

Comparing *Zea mays* iRS1563 with *Arabidopsis thaliana* and maize C4GEM models

Figure 5a compares the total number of genes, reactions and metabolites between *Zea mays* iRS1563 and the *A. thaliana* AraGEM genome-scale-models [12]. Approximately, only 61% of genes in *Zea mays* iRS1563 are present in AraGEM. This yields a surprisingly low degree of matching between these two models of 64% and 76%, respectively in terms of reactions and metabolites. In the interest of elucidating the true differences between maize and Arabidopsis irrespective of annotation chronology we constructed a more up-to-date genome-scale model for Arabidopsis by appending onto AraGEM newly annotated genes as well as full GPR annotations. We refer to this updated model containing 1,597 genes, 1,798 reactions and 1,820 metabolites as *A. thaliana* iRS1597 (see File S1). The newly added 228 reactions

Figure 3. Distribution of metabolites based on their number of appearance in different organelles. (a) cytoplasmic *Zea mays* iRS1563 metabolites in cytoplasm and other organelles, and, (b) non-cytoplasmic *Zea mays* iRS1563 metabolite-organelle participation.
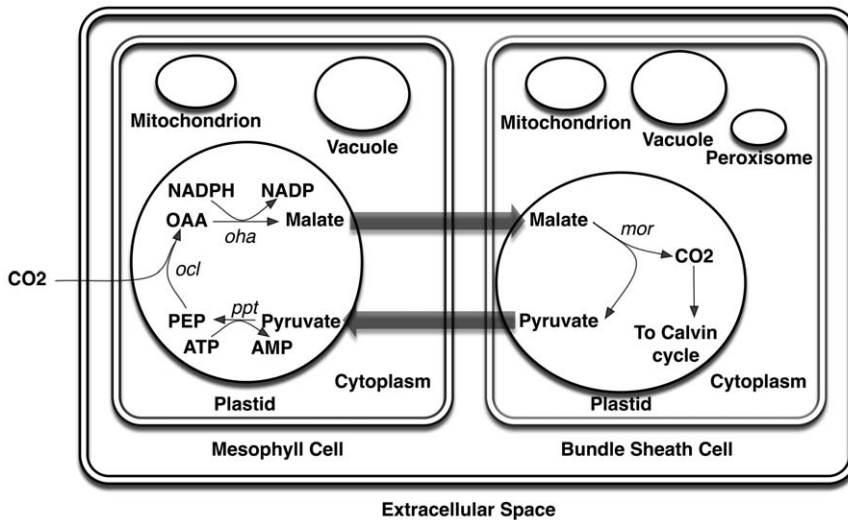doi:10.1371/journal.pone.0021784.g003

(absent from AraGEM) are involved in various pathways in primary (i.e., glycolysis, TCA, fatty acid and amino acid biosynthesis, starch and sucrose metabolism) and secondary (i.e., biosynthesis of steroid, ubiquionone, streptomycin, thiamin, riboflavin, terpenoid, brassinosteroid, phenylpropanoid, etc.) metabolism of Arabidopsis.

A direct comparison of *Zea mays* iRS1563 with *A. thaliana* iRS1597 reveals, as expected, an increased degree of matching of 72%, 76% and 80% in terms of genes, reactions and metabolites, respectively (see Figure 5b). We find that 445 reactions are unique to maize with no counterpart in *A. thaliana*. Secondary plant metabolism including flavonoid, mono- and diterpenoid, brassinosteroid, phenylpropanoid, anthocyanin, zeatin biosynthesis, riboflavin and caffeine metabolism account for 185 of the maize-specific reactions. In addition, a variety of primary metabolism reactions dispersed throughout central metabolism, photosynthesis, amino acid and fatty acid biosynthesis account for the

remaining 260 reactions. This comparison implies that about one third of the differences between *Zea mays* iRS1563 and AraGEM are caused by the incompleteness of AraGEM model especially in terms of secondary metabolism while the remaining two third reflect genuine differences between C_3 (i.e., Arabidopsis) and C_4 (i.e., maize) plant metabolism.

Figure 5c shows a similar comparison between *Zea mays* iRS1563 and maize C4GEM genome-scale-models. Degrees of matching between these two models are 39%, 53% and 63% in terms of genes, reactions and metabolites, respectively. This surprisingly low degree of matching is caused primarily due to the fact that maize C4GEM includes only metabolites and reactions in leaves during photosynthesis. Therefore, there are 893 reactions in *Zea mays* iRS1563 absent from maize C4GEM. 343 of these reactions describe secondary plant metabolism such as brassinosteroid, phenylpropanoid, carotenoid, flavonoid, mono- and diterpenoid, and glucosinolate metabolism. The remaining 550

Figure 4. Compartment and localization information for *Zea mays iRS* 1563. Mitochondrion and vacuole compartments are present in both cell types whereas peroxisome is only present in bundle-sheath cell [40]. Plastidic reactions are distributed between mesophyll and bundle-sheath cells.

doi:10.1371/journal.pone.0021784.g004

reactions are found in a wide range of primary metabolism pathways such as central metabolism, photosynthesis, benzoate degradtion, starch and sucrose metabolism, lipid metabolism, nitrogen metabolism amino acid and fatty acid biosynthesis. Conversely, 116 (out of 149) new reactions in maize C4GEM have untraceable EC numbers and gene loci.
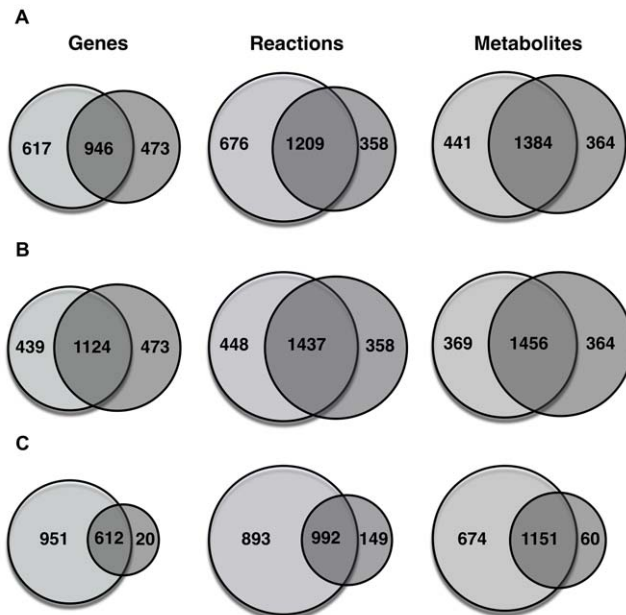
*Zea mays i*RS 1563 model testing

*Zea mays i*RS1563 allows for the production of biomass under all three different physiological states (see Files S5 and S6 for detailed

Figure 5. Venn diagram for genes, reactions and metabolites. (a) between *Zea mays i*RS1563 and AraGEM, (b) between *Zea mays i*RS1563 and *Arabidopsis thaliana i*RS1597, and (c) between *Zea mays i*RS1563 and maize C4GEM.

doi:10.1371/journal.pone.0021784.g005

information of the model). Due to limited photorespiration C4 plants usually have higher photosynthetic efficiency [43]. Under higher light intensity and photosynthetic condition, *Zea mays i*RS1563 produces 0.0008 mole biomass/mole CO_2 whereas *A. thaliana i*RS1597 yields 0.0006 mole biomass/mole CO_2. Thus, the model predictions match with findings reported in literature [43]. We also investigated the model's ability to predict the effect of suppressing genes in the lignin biosynthesis pathway observed in naturally occurring *brown midrib* (*bm*) maize mutants (i.e., *bm1*, *bm2*, *bm3* and *bm4*) [48,53,54,55]. These maize mutants are Mendelian recessives that are characterized by brown vascular tissue in leaves and stems due to a changed lignin content and/or composition [56]. The specific genetic background for two of these mutants (*bm1* and *bm3*) was elucidated based on the analysis of cell wall composition [55]. Mutants *bm1* and *bm3* were found to have disrupted enzymatic activity for cinnamyl alcohol dehydrogenase (CAD) and caffeic acid 3-*O*-methyltranferase (COMT). Both of these enzymes are involved in the last stages of the monolignol pathway [55] that controls lignin synthesis and composition (i.e., the ratio of three major subunits, H-lignin, G-lignin and S-lignin) [57].

We simulated mutants *bm1* and *bm3* using *Zea mays i*RS1563 under photosynthetic conditions by restricting the flux of the reactions catalyzed by enzymes CAD and COMT to 10% of the wild-type values. It is expected that the disruption of the activity for these genes will directly affect lignin content and composition (see File S7 to find literature data used for *bm1* and *bm3* mutants). We were interested to see whether the *Zea mays i*RS1563 metabolic model will be able to correctly propagate this disruption across the metabolic pathways and correctly predict the effect on other key metabolites. Table 6 contrasts experimental results by (Marita et al (2003), Vanholme et al (2008) and Sattler et al (2010)) with *in silico* predictions for the maximum theoretical yield of lignins, sugars and crude protein in terms of whether they increased, decreased, or remained the same in the mutant strains. Out of 21 compared components *Zea mays i*RS1563 correctly predicted the direction (or absence) of change for 17 cases.

In Figure 6 we highlight two cases that describe the availability of glucose and galactose to cell wall for mutants *bm1* and *bm3*,

respectively. 'Carbohydrate synthesis' and 'Hemicellulose synthesis' are aggregate reactions that describe the utilization ratios of sugar molecules such as arabinose, fructose, galactose, glucose ribose, mannose, sucrose, and xylose for the production of carbohydrate and hemicellulose present in the plant cell wall. For simplicity, we have simulated the model under the photosynthetic condition where CO_2 can be uptaken with a maximum allowable rate of 1000 mM/gDW-h along with photons in excess. In Figure 6a, wild-type and *bm1* mutant flux values for reactions involving glucose as reactant including 'Carbohydrate synthesis', 'Hemicellulose synthesis', 'Alpha,alpha-trehalose glucohydrolase' [R00010], 'Sucrose glucohydrolase' [R00801], 'Sn-Glycerol-3-phosphate: D-glucose 6-phosphotransferase' [R00850] and 'Cellobiose glucohydrolase' [R00306], are highlighted. For the wild-type case, the maximum theoretical yield of glucose is predicted to be 1.66 moles/mole of CO_2 but it is reduced to 0.93 moles/moles of CO_2 for the *bm1* mutant. The reduced capability of the *bm1* mutant to direct flux towards 'Carbohydrate synthesis' and 'Hemicellulose synthesis' implies that less glucose is available for the formation of cell wall components which is consistent with the experimental finding of Table 6.

Figure 6b contrasts the wild-type and *bm3* mutant maximum theoretical yields for all reactions involving galactose including 'Hemicellulose synthesis', 'ATP: D-galactose 1-phosphotransferase' [R01092] and 'Galactosylglycerol galactohydrolase' [R01104], '3-O-alpha-D-Galactosyl-1D-myo-inositol galactohydrolase' [R01194] and 'alpha-galactosidase' [R03634]. A reduction of the maximum theoretical yield of galactose from 0.81 to 0.65 moles/mole of CO_2 for the *bm3* mutant is observed. In addition, the maximum theoretical yield for reaction 'Hemicellulose synthesis' decreases by 4-fold compared to wild-type in line with the experimental finding. However, the experimentally observed increase of glucose availability in mutant *bm3* and xylose availability for both *bm1* and *bm3* mutants are in contrast with the model predictions (see Table 6). As reported by Guillaumie et al (2007) several gene expression levels were changed during *bm1* and

bm3 mutations implying that additional regulatory constraints may be needed to capture these changes.

Discussion

Maize, apart from its central role a food crop, is also a promising plant biomass target for cellulosic biofuels production. Plant cell wall cellulose, hemicellulose and lignin polymers are major contributors of plant biomass [48,58]. Therefore, controlling the amount and composition of cell wall polymers is important in developing cellulosic maize for biofuel production. In cell wall, lignin provides rigidity by forming a matrix where cellulose and hemicellulose are imbedded via cross-linking bonds [53,59]. This makes digestion of cellulose and hemicellulose by microbial enzymes (i.e., cellulases) difficult during dilignification, one of the critical steps in cellulosic biofuel production [60]. Many genetic modification strategies have been explored to improve maize food crop and/or biofuel characteristics. For example, cellulosic biomass yield improvements have been pursued before by altering the lignin content and composition [61,62], genetically manipulating the cellulose biosynthetic pathway [63] and over-expressing the gene encoding phosphoenolpyruvate carboxylase (PEPC) to improve CO_2 fixation rate [64]. At the same time, grain yield enhancements have been attempted by up-regulating ADP-glucose pyrophosphorylase (AGP) that catalyzes the rate limiting step in starch synthesis [65].

Unfortunately, existing genetic engineering strategies to reduce lignin content are problematic as lignin reductions are usually achieved at the expense of plant viability and fitness [60]. It is becoming widely accepted that focusing on a single pathway at a time without quantitatively assessing the system-wide implications of the genetic disruptions may be responsible for not preserving the agronomic properties of the plant. By accounting for both primary and some secondary metabolism pathways of maize, *Zea mays* iRS1563 can be used to explore *in silico* the effect of genetic modifications aimed at plant cell wall modification and/or starch storage on the overall metabolic state of the plant (e.g., biomass precursor availability, cofactor balancing, redox state, etc.). Moving a step further, the use of computational strain optimization techniques [66,67] can be customized for engineering plant metabolism. By taking full inventory of plant metabolism optimal gene modifications could be pursued for a variety of targets in coordination with experimental techniques. These may include (i) increase cellulose and hemicellulose production, (ii) starch yield, (iii) tolerance against biotic stress (e.g., fungal elicitation), or (iv) disruption of the production of lignin subunits (H/G/S) while enhancing the production of easily digestible lignin precursor (e.g., rosmarinic acid, conferyl ferulate, tyramine conjugates, etc).

In this paper, we introduced the first comprehensive genome-scale metabolic model (*Zea mays* iRS1563) for maize metabolism. The model meets (or exceeds) the quality and completeness criteria set out [68,69] for genome-scale reconstructions. In analogy to the human genome-scale model Recon 1 [70], *Zea mays* iRS1563 can be viewed as a mathematically structured database enabling systematic studies of maize metabolism.185 of unique to maize reactions accounting for a fraction of secondary metabolism were delineated. As a by product of this effort a more up-to-date version of AraGEM [12] was constructed including GPR associations. Comparisons between *Zea mays* iRS1563 and maize C4GEM also revealed the detail in description of primary and secondary metabolism. Model predictions of *Zea mays* iRS1563 for two widely occurring maize Mendelian mutants were tested against experimental observations with very good agreement in the direction of changes. By making use of high throughput enzymatic assays,
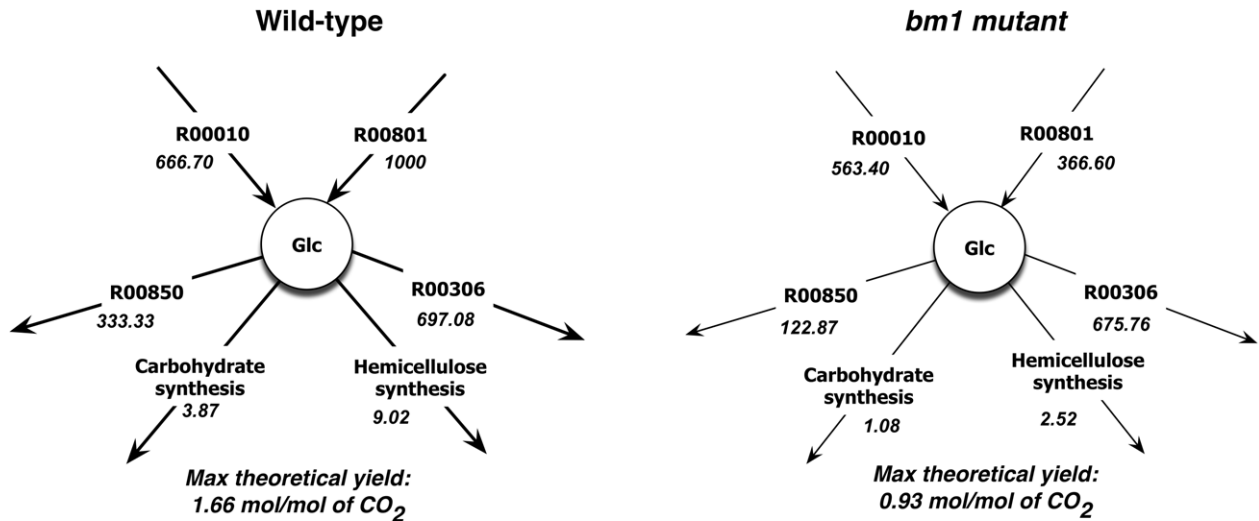
Table 6. Change in content of cell wall components in *bm1* and *bm3* Maize mutants.

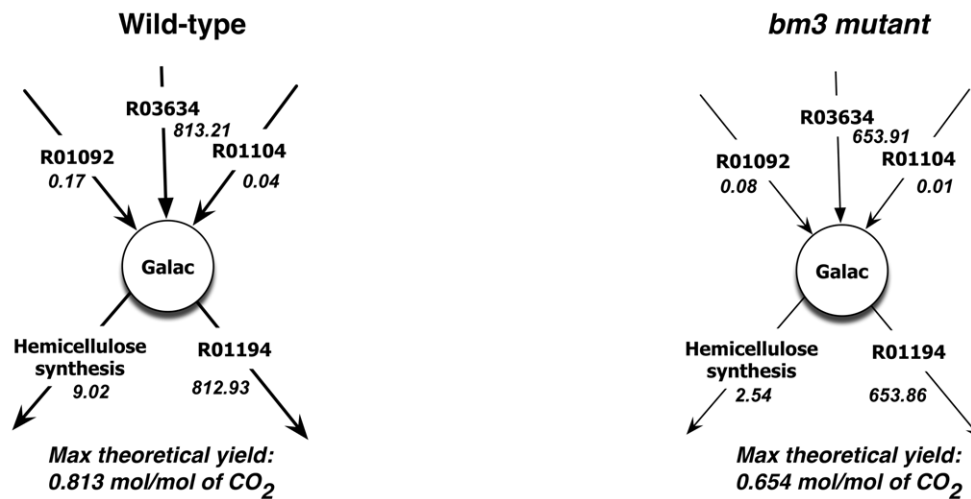|	Model findings vs Experimental observations	
	bm1 mutant	*bm3* mutant
H-lignin	↓ / =	↓ / ↓
G-lignin	↓ / ↓	↓ / ↓
S-lignin	↓ / ↓	↓ / ↓
Total lignin	↓ / ↓	↓ / ↓
S-lignin/G-lignin ratio	= / =	= / =
Glucose	↓ / ↓	↓ / ↑
Mannose	↓ / ↓	↓ / ↓
Arabinose	↓ / ↓	↓ / ↓
Galactose	↓ / ↓	↓ / ↓
Xylose	↓ / ↑	↓ / ↑
Crude protein	-	↓ / ↓

Cell wall components include lignin subunits, total lignin, S-lignin/G-lignin ratio, sugars, starch and protein. List of used symbols include ' ↓ ': decrease in quantity; ' ↑ ': increase in quantity; ' = ': no change in quantity, with respect to wild Maize plant; '/': comparison of model findings with actual observations, and '-': no experimental observation found.
doi:10.1371/journal.pone.0021784.t006

A

B

Figure 6. Maximum theoretical yields of (a) glucose and (b) galactose for wild-type vs *bm1* mutant and wild-type vs *bm3* mutant, respectively. Here the numeric values represent reaction fluxes and have the unit of mM/gDW-h.
doi:10.1371/journal.pone.0021784.g006

proteomic and transcriptomic data across different parts of the maize plant, *Zea mays* iRS1563 could serve as the starting point for the development of tissue-specific maize models [20,71,72]. Furthermore, *Zea mays* iRS1563 could also serve as the stepping stone for the development of genome-scale models for other important C$_4$ plants such as Sorghum and switch grass.

Materials and Methods

A number of recent publications [15,27,68] have outlined the general steps necessary for the metabolic reconstruction process. In the following section, we highlight the specific methods used in the reconstruction of *Zea mays* iRS1563 and subsequent model simulations in more detail.

Model reconstruction

The maizesequence database [10] provided the filtered gene set (FGS) which has been generated from the working gene set upon removing pseudogenes and low confidence hypothetical models.

The FGS of B73 maize genome (release 4a.53) was downloaded from maizesequence database on February 17, 2010. Once maize genes were obtained, we used sequence comparison tools [73] such as stand-alone BLAST (version 2.2.22, NIH) and BLAST+ (version 2.2.22, NIH) for performing homology comparisons. Marvin (version 5.3.3, ChemAxon Kft) was used to calculate the average micro-species charge to determine the net charge of individual metabolites at pH 7.2 assumed for all organelles. In the final step of the model reconstruction, we implemented GapFind and GapFill [37] for analyzing and subsequently restoring metabolic network connectivity.

Model simulations

Flux balance analysis (FBA) [74] was employed both in model validation and model testing phases. *Zea mays* iRS1563 was evaluated in terms of biomass production under three standard physiological scenarios: photosynthesis, photorespiration, and respiration. Flux distributions for each one of these states were approximated using FBA:

$$\text{Maximize } v_{Biomass}$$

Subject to

$$\sum_{j=1}^{m} S_{ij} v_j = 0 \ \forall i \in 1,\ldots, n \qquad (1)$$

$$v_{j,\min} \le v_j \le v_{j,\max} \ \forall j \in 1,\ldots,m \qquad (2)$$

Here, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j and v_j is the flux value of reaction j. Parameters $v_{j,min}$ and $v_{j,max}$ denote the minimum and maximum allowable fluxes for reaction j, respectively. As mentioned in Table 4, the three physiological states were represented via modifying the relevant minimum or maximum allowable fluxes and the following constraints:

$$v_{oxi} = 0 \qquad (3)$$

$$v_{carboxi} \ge 3 v_{oxi} \qquad (4)$$

$$v_{carboxi} = 0 \qquad (5)$$

where $v_{Biomass}$ is the flux of biomass reaction and v_{oxi} and $v_{carboxi}$ are the fluxes of carboxylation and oxidation reactions associated with enzyme RUBISCO. For photosynthesis and photorespiration, constraints (3) and (4) were respectively included in the linear model, whereas for respiration both constraints (3) and (5) were included.

Once the model was validated, it was further tested for two maize mutants (i.e., *bm1* and *bm3*) under the photosynthetic condition. The following two constraints were included individually in the linear model to represent the mutants:

$$v_{bm1} \le w \times WF_{bm1} \qquad (6)$$

$$v_{bm3} \le w \times WF_{bm3} \qquad (7)$$

Here, w represents the percent of residual activity of 10%. v_{bm1} and v_{bm3} are the fluxes of reactions catalyzed by CAD and COMT,

respectively and WF_{bm1} and WF_{bm3} are the corresponding wild-type flux values under the photosynthetic condition.

CPLEX solver (version 12.1, IBM ILOG) was used in the GAMS (version 23.3.3, GAMS Development Corporation) environment for implementing GapFind and GapFill [37] and solving the aforementioned optimization models. All computations were carried out on Intel Xeon E5450 Quad-Core 3.0 GH and Intel Xeon E5472 Quad-Core 3.0 GH processors that are the part of the lionxj cluster (Intel Xeon E type processors and 96 GB memory) of High Performance Computing Group of The Pennsylvania State University.

Supporting Information

File S1 *A. thaliana* iRS1597 model along with established GPR for AraGEM.
(XLS)

File S2 Sequential stages of *Zea mays* iRS1563 model development.
(XLS)

File S3 Results from BLASTp searches against NCBI nr database.
(XLS)

File S4 Biomass equation, added reactions for restoring biomass production under all three physiological conditions and unique reactions in *Zea mays* iRS1563 model.
(XLS)

File S5 *Zea mays* iRS1563 model.
(XLS)

File S6 SBML file of *Zea mays* iRS1563 model.
(XML)

File S7 Literature data used for testing *Zea mays* iRS1563 model.
(XLS)

Author Contributions

Conceived and designed the experiments: RS PFS CDM. Performed the experiments: RS. Analyzed the data: PFS. Wrote the paper: RS PFS CDM.

References

1. Bennetzen JL, Hake S, SpringerLink (Online service) (2009) Handbook of Maize Genetics and Genomics. New YorkNY: Springer New York. 812 p.
2. Sanchez OJ, Cardona CA (2008) Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresource Technology 99: 5270–5295.
3. Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, et al. (2006) Ethanol can contribute to energy and environmental goals. Science 311: 506–508.
4. Stewart CN, Jr. (2007) Biofuels and biocontainment. Nat Biotechnol 25: 283–284.
5. Mechin V, Argillier O, Rocher F, Hebert Y, Mila I, et al. (2005) In search of a maize ideotype for cell wall enzymatic degradability using histological and biochemical lignin characterization. J Agric Food Chem 53: 5872–5881.
6. Dennis C, Surridge C (2000) A. thaliana genome. Nature 408: 791–791.
7. Yu J, Hu SN, Wang J, Wong GKS, Li SG, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science 296: 79–92.
8. Goff SA, Ricke D, Lan TH, Presting G, Wang RL, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science 296: 92–100.
9. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457: 551–556.
10. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.
11. Xavier A, Salse J, Aury JM, Droc G, Gouzy J, et al. (2010) Deciphering the genome structure and paleohistory of Theobroma cacao. Nature Proceedings; doi:10.1038/npre.2010.4908.1.
12. Dal'Molin CGD, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. Plant Physiology 152: 579–589.
13. Sweetlove LJ, Last RL, Fernie AR (2003) Predictive metabolic engineering: A goal for systems biology. Plant Physiology 132: 420–425.
14. Gutierrez RA, Shasha DE, Coruzzi GM (2005) Systems biology for the virtual plant. Plant Physiology 138: 550–554.
15. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. Nature Reviews Microbiology 7: 129–143.
16. Park JM, Kim TY, Lee SY (2009) Constraints-based genome-scale metabolic simulation for systems metabolic engineering. Biotechnology Advances 27: 979–988.
17. Milne CB, Kim PJ, Eddy JA, Price ND (2009) Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. Biotechnol J 4: 1653–1670.

18. Poolman MG, Miguel L, Sweetlove LJ, Fell DA (2009) A Genome-Scale Metabolic Model of Arabidopsis and Some of Its Properties. Plant Physiology 151: 1570–1581.

19. Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH (2009) Flux Balance Analysis of Barley Seeds: A Computational Approach to Study Systemic Properties of Central Metabolism. Plant Physiology 149: 585–598.

20. Dal'Molin CGD, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) C4GEM, a Genome-Scale Metabolic Model to Study C-4 Plant Metabolism. Plant Physiology 154: 1871–1885.

21. 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): 1673–82. doi: 10.1002/bit.23107.

22. Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003) Comparisons with Caenorhabditis (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately 25% larger than the Arabidopsis genome initiative estimate of approximately 125 Mb. Ann Bot 91: 547–557.

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

24. Liang C, Mao L, Ware D, Stein L (2009) Evidence-based gene predictions in plant genomes. Genome Res 19: 1912–1923.

25. Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10: 516–522.

26. Notebaart RA, van Enckevort FH, Francke C, Siezen RJ, Teusink B (2006) Accelerating the reconstruction of genome-scale metabolic networks. BMC Bioinformatics 7: 296.

27. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, et al. (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. PLoS Comput Biol 5: e1000285.

28. Penningd FW, Brunstin AH, Vanlaar HH (1974) Products, Requirements and Efficiency of Biosynthesis - Quantitative Approach. Journal of Theoretical Biology 45: 339–377.

29. Spector WS (1956) Handbook of biological data. PhiladelphiaPA: Saunders. xxxvi, 584 p.

30. Muller F, Dijkhuis DJ, Heida YS (1970) On the relationship between chemical composition and digestibility in vivo of roughages. Agricultural Research Report 736: 1–27.

31. Wedig C, Jaster EH, Moore KJ (1987) Hemicellulose monosaccharide composition and in vitro disappearance of orchard grass and alfalfa hay. Journal of Agricultural and Food Chemistry 35: 23–27.

32. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, et al. (2009) PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Research 37: D969–D974.

33. Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the Arabidopsis Subcellular Database. Nucleic Acids Res 35: D213–218.

34. Volk RJ, Jackson WA (1972) Photorespiratory Phenomena in Maize - Oxygen-Uptake, Isotope Discrimination, and Carbon-Dioxide Efflux. Plant Physiology 49: 218–223.

35. Dai ZY, Ku MSB, Edwards GE (1995) C-4 Photosynthesis - the Effects of Leaf Development on the Co2-Concentrating Mechanism and Photorespiration in Maize. Plant Physiology 107: 815–825.

36. Jolivettournier P, Gerster R (1984) Incorporation of Oxygen into Glycolate, Glycine, and Serine during Photorespiration in Maize Leaves. Plant Physiology 74: 108–111.

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

38. Wei Y, Lin M, Oliver DJ, Schnable PS (2009) The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of Arabidopsis. BMC Biochem 10: 7.

39. Ouzounis CA, Karp PD (2000) Global properties of the metabolic map of Escherichia coli. Genome Res 10: 568–576.

40. Hong SJ, Lee CG (2007) Evaluation of central metabolism based on a genomic database of Synechocystis PCC6803. Biotechnology and Bioprocess Engineering 12: 165–173.

41. Wise RR, Hoober JK (2007) Synthesis, export and partitioning of end products of photosynthesis. *In* RR. Wise, JK. Hoober, eds. Structure and Function of Plastids, Vol 23. Dordrecht, The Netherlands: Springer. pp 274–288.

42. Dennis DT, Miernyk JA (1982) Compartmentation of Non-Photosynthetic Carbohydrate-Metabolism. Annual Review of Plant Physiology and Plant Molecular Biology 33: 27–50.

43. Taiz L, Zeiger E (2002) Plant Physiology. SunderlandMassachusetts: Sinauer Associates, Inc., Publishers.

44. Allen JF (2002) Photosynthesis of ATP - Electrons, proton pumps, rotors, and poise. Cell 110: 273–276.

45. Hervas M, Navarro JA, De La Rosa MA (2003) Electron transfer between membrane complexes and soluble proteins in photosynthesis. Accounts of Chemical Research 36: 798–805.

46. Gregory RF (1989) Biochemistry of Photosynthesis. ChichesterNY, , USA: John Wiley & Sons. 280 p.

47. Tsaftaris AS, Bosabalidis AM, Scandalios JG (1983) Cell-Type-Specific Gene-Expression and Acatalasemic Peroxisomes in a Null Cat2 Catalase Mutant of Maize. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80: 4455–4459.

48. Hisano H, Nandakumar R, Wang ZY (2009) Genetic modification of lignin biosynthesis for improved biofuel production. In Vitro Cellular & Developmental Biology-Plant 45: 306–313.

49. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology 126: 485–493.

50. Styles ED, Ceska O (1975) Genetic-Control of 3-Hydroxy-Flavonoids and 3-Deoxy-Flavonoids in Zea-Mays. Phytochemistry 14: 413–415.

51. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126: 485–493.

52. Weidemann C, Tenhaken R, Hohl U, Barz W (1991) Medicarpin and Maackiain 3-O-Glucoside-6′-O-Malonate Conjugates Are Constitutive Compounds in Chickpea (Cicer-Arietinum L) Cell-Cultures. Plant Cell Reports 10: 371–374.

53. Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Current Opinion in Plant Biology 11: 278–285.

54. Sattler SE, Funnell-Harris DL, Pedersen JF (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. Plant Science 178: 229–238.

55. Marita JM, Vermerris W, Ralph J, Hatfield RD (2003) Variations in the cell wall composition of maize brown midrib mutants. Journal of Agricultural and Food Chemistry 51: 1313–1321.

56. Kuc J, Nelson OE (1964) Abnormal Lignins Produced by Brown-Midrib Mutants of Maize .I. Brown-Midrib-1 Mutant. Archives of Biochemistry and Biophysics 105: 103-&.

57. Guillaumie S, Pichon M, Martinant JP, Bosio M, Goffner D, et al. (2007) Differential expression of phenylpropanoid and related genes in brown-midrib bm1, bm2, bm3, and bm4 young near-isogenic maize plants. Planta 226: 235–250.

58. Sticklen MB (2009) Expediting the biofuels agenda via genetic manipulations of cellulosic bioenergy crops. Biofuels Bioproducts & Biorefining-Biofpr 3: 448–455.

59. Sticklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat Rev Genet 9: 433–443.

60. Li X, Weng JK, Chapple C (2008) Improvement of biomass through lignin modification. Plant Journal 54: 569–581.

61. Vega-Sanchez ME, Ronald PC (2010) Genetic and biotechnological approaches for biofuel crop improvement. Current Opinion in Biotechnology 21: 218–224.

62. Grabber JH, Schatz PF, Kim H, Lu FC, Ralph J (2010) Identifying new lignin bioengineering targets: 1. Monolignol-substitute impacts on lignin formation and cell wall fermentability. Bmc Plant Biology 10: 114.

63. Abramson M, Shoseyov O, Shani Z (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability. Plant Science 178: 61–72.

64. Torney F, Moeller L, Scarpa A, Wang K (2007) Genetic engineering approaches to improve bioethanol production from maize. Current Opinion in Biotechnology 18: 193–199.

65. Smidansky ED, Martin JM, Hannah LC, Fischer AM, Giroux MJ (2003) Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. Planta 216: 656–664.

66. Kim J, Reed JL (2010) OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 4: 53.

67. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6: e1000744.

68. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93–121.

69. Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nature Biotechnology 26: 659–667.

70. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et al. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences of the United States of America 104: 1777–1782.

71. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. Nature Biotechnology 26: 1003–1010.

72. Jerby L, Shlomi T, Ruppin E (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Molecular Systems Biology 6: doi:10.1038/msb.2010.56.

73. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology 215: 403–410.

74. Varma A, Palsson BO (1994) Metabolic Flux Balancing - Basic Concepts, Scientific and Practical Use. Bio-Technology 12: 994–998.