UNIVERSIDAD
NACIONAL
DE COLOMBIA

Modelo computacional de la interacción compatible entre *Phytophthora infestans* y *Solanum tuberosum* a través de la reconstrucción computacional del metabolismo de la planta y datos de expresión génica

Kelly Johana Botero Orozco

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2016

Modelo computacional de la interacción compatible entre *Phytophthora infestans* y *Solanum tuberosum* a través de la reconstrucción computacional del metabolismo de la planta y datos de expresión génica

Kelly Johana Botero Orozco

Tesis presentada como requisito parcial para optar al título de:

Magister en Bioinformática

Director:

Ph.D. Andres Mauricio Pinzón Velasco

Línea de Investigación:

Biología de Sistemas - Reconstrucción y Análisis de Redes Biológicas

Grupos de Investigación:

Bioinformática y Biología de Sistemas - Universidad Nacional de Colombia

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2016

Agradecimientos

Al profesor Andres Mauricio Pinzón, por su asesoría y apoyo durante el desarrollo de este trabajo.

A mi familia, por ser el gran motor de mi vida, y por todos los años de amor y enseñanzas, que han hecho de mí, una mujer que lucha con pasión por alcanzar sus metas.

A mi compañero de vida, por ser la persona que alegra cada uno de mis días y por darme su apoyo incondicional.

A Daniel Camilo Osorio, porque fue un excelente compañero de estudio.

Al grupo de investigación en bioinformática y biología de sistemas por acogerme y permitirme compartir diversas experiencias académicas y de vida.

Al grupo de micología y fitopatología de la Universidad de los Andes por los aportes a este trabajo.

A la profesora Liliana López Kleine por su asesoría para integrar a los flujos de reacción datos de expresión ausentes.

Al profesor Pedro Adolfo Jiménez por las sugerencias para incorporar las reacciones lumínicas de la fotosíntesis a la reconstrucción metabólica a escala genómica de papa.

Resumen

Solanum tuberosum, la papa común, es uno de los principales cultivos para el consumo humano; el patógeno más importante que afecta su producción es el oomicete *Phytophthora infestans,* causante de la enfermedad del tizón tardío. Análisis de la expresión génica en plantas enfermas sugieren disminución de la capacidad de fijación de CO_2. Este trabajo presenta un modelo metabólico computacional de la interacción compatible entre *S. tuberosum* y *P. infestans,* para analizar a nivel sistémico, los mecanismos metabólicos asociados a la fotosíntesis (reacciones lumínicas y fijación de CO_2) en plantas enfermas. Para este fin se reconstruyó la red metabólica de la planta a partir de anotaciones funcionales del genoma de *S. tuberosum* y bases de datos bioquímicas y metabólicas. La reconstrucción fue refinada manualmente y automáticamente usando el paquete "g2f", desarrollado para este trabajo en el lenguaje de programación R. La reconstrucción refinada se transformó en un modelo metabólico computacional, al cual se le integraron datos de expresión génica obtenidos durante la interacción compatible entre *S. tuberosum* y *P. infestans* en tres momentos (0, 2 y 3 días post-inoculación); para esto se desarrolló el paquete de R "exp2flux", que permite integrar restricciones de flujo a partir de datos ómicos a modelos metabólicos. El análisis de balance de flujo, mostró disminución de flujos metabólicos en la ruta de fijación de CO_2 y en las reacciones lumínicas de la fotosíntesis durante la infección. Este trabajo reafirma la disminución de la capacidad de fijación de CO_2 y sugiere, además, una supresión general de la capacidad fotosintética en la planta durante la interacción compatible con *P. infestans*, probablemente asociada a un mecanismo de defensa. Aquí se reporta la primer reconstrucción metabólica a escala genómica de *S. tuberosum* y el primer modelo metabólico a escala genómica de una interacción planta-patógeno.

Palabras clave: *Solanum tuberosum, Phytophthora infestans, interacción planta patógeno,* modelo metabólico a escala genómica, fotosíntesis.

Abstract

The potato plant (*Solanum tuberosum*) is one of the main crops for human consumption; the most important pathogen that affects its production is the oomycete *Phytophthora infestans*, which causes late blight disease. Gene expression analysis suggests decreased CO_2 fixation capacity in diseased plants. This work presents a computational metabolic model of the compatible interaction between *S. tuberosum* and *P. infestans* which allowed us to study the metabolic mechanisms associated with photosynthesis (light reactions and CO_2 fixation) at a systemic level in diseased plants. For this purpose, the reconstruction of the metabolic network of the plant was obtained from functional annotations of the *S. tuberosum* genome and biochemical and metabolic databases. The reconstruction was refined manually and automatically using the R package "g2f", developed for this work. Refined reconstruction was transformed into a computational metabolic model, and gene expression data during the compatible interaction between *S. tuberosum* and *P. infestans* at three times (0, 2 and 3 days post-inoculation) was integrated into the model. For this, the "exp2flux" R package was developed, which allows the integration of flux constraints from omics data to metabolic models. Flux balance analysis showed decreased metabolic fluxes in the CO_2 fixation pathway and in the light reactions of the photosynthesis during infection. This work reaffirms the decrease of the CO_2 fixation capacity and also suggests a general suppression of the photosynthetic capacity in the plant during the interaction compatible with *P. infestans*, likely associated to an oxidative defense mechanism. Here, the first genome-scale metabolic reconstruction of *S. tuberosum* and the first genome-scale metabolic model of a plant-pathogen interaction are reported.

Keywords: *Solanum tuberosum, Phytophthora infestans,* plant-pathogen interaction, genome-scale metabolic model, photosynthesis.

Contenido

Pág.

Lista de figuras

Lista de tablas

Objetivos

Objetivo general

Generar un modelo computacional del metabolismo de *Solanum tuberosum* que permita identificar los posibles mecanismos de inhibición de su capacidad fotosintética durante la interacción con *Phytophthora infestans*

Objetivo específicos

● Identificar el conjunto de reacciones bioquímicas y propiedades metabólicas de *S.tuberosum*

● Obtener una representación matemática y computacional del metabolismo *S.tuberosum*

● Identificar los posibles mecanismos moleculares relacionados con la inhibición de la capacidad fotosintética en *S. tuberosum* durante la interacción compatible con *P. infestans*

Introducción

Generalidades de la interacción entre *Phytophthora infestans* y *Solanum tuberosum*

Phytophthora infestans Mont. De Bary, es un Oomycete patógeno causante de la enfermedad del tizón tardío de la papa *Solanum toberosum* [1]. La propagación de la infección por este patógeno ocurre rápidamente en el follaje y en los tubérculos causando necrosis en la planta[2]. La infección de *P. infestans* en los tejidos de *S. tuberosum* se desarrolla durante el ciclo de vida hemibiotrófico del patógeno[3], el cual puede resumirse en cuatro etapas: (1) adhesión de zoosporas móviles a la superficie celular [4]; (2) penetración del patógeno a la célula mediante la formación de estructuras especializadas llamadas haustorios [5] y secreción de enzimas de degradación como fosfolipasas, glucanasas y endopoligalacturonasas [6]; (3) colonización a través de la liberación de proteínas efectoras del patógeno que manipulan procesos fisiológicos de la planta [4], [6], [7]. En esta etapa, cuando las proteínas efectoras no son reconocidas por el sistema inmune o cuando el patógeno genera estrategias para evadir respuestas de defensa del hospedero, la interacción planta-patógeno (patosistema) es compatible y la planta infectada se considera susceptible, de lo contario, la planta se considera resistente y la interacción incompatible [8], [9]; y (4) en plantas susceptibles el patógeno adquiere nutrientes y se propaga por los tejidos de la planta [4], [10].

Relevancia de la enfermedad de tizón tardío de la papa

La papa es el cuarto cultivo alimentario más importante (después del maíz, el arroz y el trigo) para el consumo humano [11]. Los tubérculos de papa suministran principalmente carbohidratos, pero también son fuente de proteínas, vitaminas, fibra dietética y minerales

[12]. Históricamente, el tizón tardío o gota de la papa es el principal factor que afecta la producción de la papa [13], [14]. Desde su primer brote de pandemia en la mitad del siglo XIX, *P. infestans* ha sido el patógeno más destructivo de las plantaciones de este cultivo, lo que resulta en pérdidas anuales que serían suficientes para alimentar a varios cientos de millones de personas [15]. Además, el valor económico de esta pérdida y el costo de protección del cultivo se estima en 6,7 mil millones de dólares por año [16]. A pesar de la importancia económica y social que implica el tizón tardío de la papa, los principales mecanismos metabólicos que subyacen la patogenicidad causada por *P. infestans* son aún poco entendidos [17].

Estudios moleculares sobre la interacción entre *S. tuberosum* y *P.infestans*

Múltiples investigaciones han intentado comprender aspectos moleculares que subyacen la interacción compatible e incompatible entre *P. infestans* y *S. toberosum* [5], [8], [9], [18]–[20]. Algunas han utilizado estrategias como expresión génica [17]–[22] y genómica [23], la mayoría de estos estudios se han enfocado en identificar factores que favorecen la interacción incompatible de las plantas resistentes a *P. infestans* [18], [21], [24]–[28], identificado así, genes y compuestos involucrados en la respuesta de defensa de la planta ante la invasión patogénica [21], [24], [25]. No obstante, también se han desarrollado investigaciones con microarreglos basados en clones ADNc[19], análisis de transcriptoma mediante DeepSAGE [29] y análisis de RNA-Seq [18] para intentar comprender la interacción compatible del patosistema. Particularmente, Restrepo *et al*., (2005) encontraron que 358 genes se reprimen y 241 genes se sobre expresan durante cinco puntos de tiempo del proceso de infección, estos resultados y otros análisis complementarios con rtq-PCR permitieron concluir que la interacción compatible suprime genes involucrados en la ruta metabólica de fijación de CO_2 y genes involucrados en la ruta metabólica del ácido jasmonico, la cual está asociada al sistema de defensa de plantas [19].

Por otra parte, en 2011 el consorcio para la secuenciación de la papa (PGSC, del inglés Potato Genome Sequencing Consortium) reportó la secuenciación del genoma de *S.*

tuberosum con 39,031 genes codificantes de proteínas, entre los que se identificaron homólogos altamente relacionados con genes de resistencia (R1, RB, R2, R3a, Rpi-blb2 and Rpi-vnt1.1) al tizón tardío de la papa [30]. De otro lado, el genoma de *P. infestans* también ha sido reportado con 17,797 genes codificantes de proteínas. La comparación de este genoma con las especies cercanas, *Phytophthora sojae* y *Phytophthora ramorum* mostró que el 70% de sus genes son ortólogos y permitió identificar genes que codifican proteínas efectoras que alteran la fisiología del hospedero y favorecen la patogénesis [16].

Los estudios descritos anteriormente, representan una importante fuente de conocimiento acerca de mecanismos y moléculas que subyacen la interacción planta-patógeno, sin embargo, dado que la respuesta biológica de las plantas al interactuar con un patógeno está regulada por complejas redes de interacción génica, proteica y metabólica [31], [32], estos estudios enmarcan una visión fragmentada de los procesos de evasión o acogida del patógeno por parte de la planta y evidencian la necesidad de contemplar la utilización de metodologías más integrativas.

En este caso, donde datos experimentales de la interacción, así como datos genómicos de *S. tuberosum* comienzan a acumularse, es factible y conveniente integrar toda esa información dentro de un modelo metabólico *in silico* de la planta. Esta aproximación permite obtener un modelo descriptivo de la interacción compatible entre *P. infestans* y *S. tuberosum*, el cual fortalecería la capacidad predictiva de futuras respuestas de la planta, bajo diferentes condiciones y mostraría una visión global y detallada del sistema de interacción que se lleva a cabo en este patosistema y particularmente al interior celular del hospedero [17].

Biología de sistemas: una aproximación para la compresión de sistemas complejos

Durante la segunda mitad del siglo XX, la biología estuvo fuertemente influenciada por enfoques reduccionistas que buscaban generar información sobre los componentes individuales de los sistemas biológicos, con la cual se buscaba inferir sobre las propiedades de un organismo como un todo [33]–[35]. Este enfoque ha generado importantes logros y

una enorme cantidad de datos, sin embargo, no permite la comprensión de sistemas complejos [36], [37], como por ejemplo, la relación genotipo – fenotipo de rasgos fenotípicos que son el resultado de múltiples interacciones de productos génicos[32], [38].

Actualmente, la compresión de procesos biológicos más complejos, como por ejemplo los que se llevan a cabo durante el conjunto de reacciones bioquímicas que denominamos metabolismo [39], no se centra en el estudio de componentes individuales (ej. enzimas y metabolitos), sino en la naturaleza de los vínculos que los conectan [33]. Esta orientación, de naturaleza más holista, es conocida como Biología de Sistemas o biología sistémica, y busca predecir el comportamiento y entender las propiedades que emergen de los sistemas biológicos ante estímulos y cambios en el ambiente, entendiéndolos como un conjunto de componentes que interactúan de manera compleja a nivel de red. Estos sistemas son susceptibles de ser modelados mediante métodos computacionales con alta capacidad predictiva, diferenciándose así de la aproximación reduccionista que se enfoca en el estudio de los componentes individuales de manera aislada.

Reconstrucciones metabólicas a escala genómica

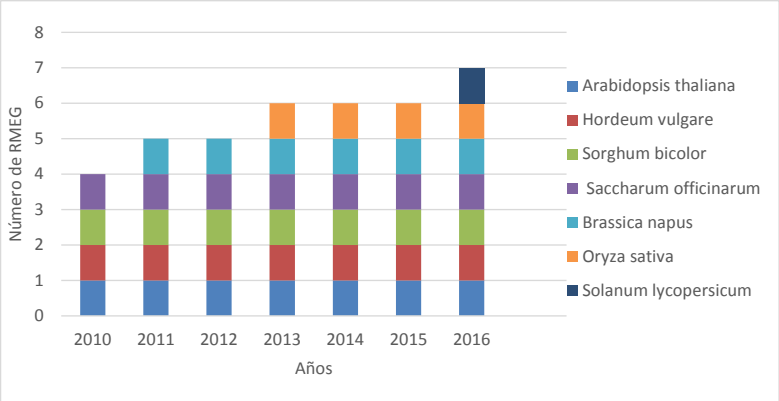
El surgimiento y avance de las tecnologías secuenciación de siguiente generación, ha permitido anotar con un alto nivel de precisión las funciones del genoma de los organismos [40]. Estas anotaciones son fundamentales para la reconstrucción de redes metabólicas a escala genómica (RMEG), las cuales son una representación de las reacciones bioquímicas presentes en un organismo, que pueden ser analizadas, interpretadas y predichas, mediante simulación computacional [33]. Un método computacional ampliamente utilizado para generar predicciones fenotípicas de la actividad metabólica de un organismo se encuentra en el análisis de balance de flujo (FBA, por sus siglas en inglés). EL FBA implica la optimización lineal del conjunto de reacciones de una red metabólica en estado estacionario, con el fin de predecir una distribución óptima de flujos metabólicos, que maximiza objetivos metabólicos inferidos, bajo un conjunto de restricciones que limitan la solución de sistema [41]. El establecimiento de los objetivos metabólicos supone que la célula desempeña óptimamente una función metabólica [42]. Ejemplos de funciones objetivo incluyen, la

maximización de biomasa, la producción de ATP y la producción de un metabolito de interés [43].Una vez que se fija la función objetivo, el sistema en estado estacionario puede resolverse para obtener una distribución de flujos óptima que permiten interpretar las capacidades metabólicas de la red [42]. Para que la optimización obtenga una distribución de flujos fisiológicamente viables, el espacio de solución del sistema en el FBA se debe restringir con restricciones termodinámicas y restricciones dadas por funciones metabólicas en condiciones particulares. Estas restricciones son impuestas a los límites de los flujos (superior e inferior) de las reacciones en la red metabólica [44], [45].

La modelación de RMEG a través de FBA tiene importantes alcances sobre los sistemas biológicos complejos, permite por ejemplo: (1) inferir la función de redes bioquímicas; (2) plantear hipótesis dirigidas que pueden ser resueltas experimentalmente para la incorporación de nuevo conocimiento biológico, y (3) descubrir propiedades emergentes de las redes, ya que se enfoca en toda la red metabólica, más que, en genes o rutas individuales [46].

Para plantas, hasta el 2009 solo se reportaban las RMEG de *Arabidopsis thaliana* [47], [48], y de la semilla de cebada (*Hordeum vulgare)* [49]. Esta escasez de reconstrucciones se justificaba en el bajo número de genomas completos disponibles; actualmente se han secuenciado y publicado más de 100 genomas de plantas en la base de datos Genbank del National Center for Biotechnology Information – NCBI [50], [51]. Ante este panorama, las reconstrucciones metabólicas en plantas están creciendo con modelos disponibles para *Zea mays*[52]–[54], *Sorghum bicolor* [52], *Saccharum officinarum*[52]; *Brassica napus* [55], [56], *Oryza sativa* [57] y *Solanum lycopersicum* [58] (Figura 0-1).

Figure 0-1: Reportes de redes metabólicas a escala genómica en plantas



A partir de estas RMEGs, múltiples análisis y predicciones metabólicas se han se llevado a cabo. Particularmente, para recrear fenotipos metabólicos específicos de las plantas, algunos estudios han integrado datos ómicos como restricciones de flujo de reacción en las RMEGs [59]. Töpfer *et al.*, (2013) incorporaron, mediante el método E-Flux [60], datos de expresión de genes [61] a la RMEG de *Arabidopsis thaliana* [62] para intentar comprender funciones metabólicas de aclimatación a diferentes temperaturas e intensidades de luz [63]. Por su parte, Simons *et al*., (2014) con el propósito de profundizar en el metabolismo del nitrógeno en la planta del maíz, incorporaron a la RMEG de *Zea mays* [53] datos transcriptómicos y proteómicos de experimentos de plantas silvestres y mutantes, sometidas a condiciones óptimas y bajas de nitrógeno. Para ello, apagaron las reacciones asociadas a transcritos y proteínas que presentaron bajos niveles de expresión en los experimentos [54].

Generación del modelo metabólico de la interacción entre *S. tuberosum* y *P. infestans*

Con la disponibilidad del genoma de *S. tuberosum* y datos de expresión génica durante la interacción compatible, este trabajo generó un modelo computacional que representa el metabolismo celular de la hoja de una planta enferma, con el fin de comprender los mecanismos metabólicos que subyacen la posible disminución de la capacidad de fijación de CO_2, mencionada previamente[19]. Este trabajo pone a disposición de la comunidad científica las bases fundamentales para la modelación de la red metabólica a escala genómica de *S.tuberosum*, sobre las cuales se podrán entender otros mecanismos metabólicos de la

planta durante la interacción compatible o bajo otras condiciones ambientales, plantear novedosas hipótesis dirigidas, orientar investigaciones para realizar ingeniería metabólica en la planta e identificar propiedades emergentes del metabolismo de la planta, que no podrían observarse con el estudio de moléculas y procesos individuales.

Adicionalmente, se diseñaron e implementaron el paquete "g2f" para la etapa de refinamiento de la red metabólica y el paquete "exp2flux" para la etapa de generación del modelo metabólico de la interacción planta-patógeno. Ambos paquetes están desarrollados en el lenguaje de programación R y están disponibles en *The Comprehensive R Archive Network* (CRAN).

En este sentido, este trabajo presenta tres capítulos en los que se exponen los principales conceptos, métodos, desarrollos y aplicaciones necesarios para la generación del modelo metabólico a escala genómica de la interacción compatible entre *S. tuberosum y P. infestans*. En el primer capítulo titulado "A genome-Scale metabolic model of potato suggests a mechanism of photosynthetic suppression during the compatible interaction with *Phytophthora infestans"* se presenta el desarrollo de la primer red metabólica a escala genómica de *S. tuberosum* y el primer modelo metabólico a escala genómica de la interacción compatible entre la planta y *P. infestans.* El modelo fue evaluado principalmente para observar las variaciones de los flujos metabólicos asociados a fotosíntesis, lo que reflejó disminución de la capacidad fotosintética durante la infección. En el segundo capítulo titulado "Finding and filling gaps in metabolic networks through the 'g2f' R package", se presenta el paquete "g2f", el cual puede realizar gap find y gap fill en reconstrucciones metabólicas a escala genómica. Este paquete fue usado para el refinamiento semiautomático y automático de la red metabólica de la papa. En el último capítulo titulado "Constraining the metabolic reconstruction: Incorporating expression data as FBA limits through the 'exp2flux' R package", se presenta el paquete "exp2flux", el cual permite incorporar datos de expresión a modelos metabólicos a escala genómica. Este paquete fue usado para integrar datos de expresión de la interacción compatible entre *S. tuberosum y P. infestans* a la red metabólica de la planta. .

1. A Genome-Scale Metabolic Model of Potato suggests a mechanism of photosynthetic suppression during the compatible interaction with *Phytophthora infestans*

Abstract:

Phytophthora infestans is a plant pathogen that causes an economically important plant disease known as late blight in potato (*Solanum tuberosum*) plants and several other hosts. In spite of the importance of the disease, the molecular mechanisms underlying the compatibility between the pathogen and its hosts are still unknown. To elucidate the metabolic response of this disease, particularly, photosynthesis inhibition shown in infected plants, we reconstructed PstM1, a genome-scale metabolic network of the *S. tuberosum* leaf. This metabolic network is used to simulate the effect of potato late blight in the leaf metabolism. PstM1 accounts for 2765 genes, 1113 metabolic functions, 1771 gene-protein-reaction associations and 1938 metabolites involved in 2072 reactions. The model optimization for biomass synthesis maximization in three infection times suggests a suppression of the photosynthetic capacity related to the decrease of metabolic flux in light reactions and carbon fixation reactions. In addition, we were also able to identify a variation pattern in the flux of carboxylation to oxygenation reactions catalyzed by RuBisCO, likely associated to a defense response in the compatible interaction between *P. infestans* and *S. tuberosum.*

Keywords: Late blight, *Phytophthora infestans*, *Solanum tuberosum*, systems biology, metabolic model, flux balance analysis.

1.1 Introduction

Plants use various strategies to resist pathogen attack and avoid the development of diseases [64]. In those cases, plants are considered as *resistant* to that specific disease, and pathogens are referred to as avirulent on that plant. The whole plant-pathogen interaction is thus known as incompatible. On the other hand, sometimes plant pathogens develop strategies to evade plant defense responses, become virulent and establish the disease. In those cases, the plant-pathogen interaction is known as *compatible* and the infected plant is considered as *non*-resistant or susceptible [8], [9].

For some plant pathogens these evasion mechanisms are, at least, partially known [64], however for some others these mechanisms are still obscure. The latter is the case of *Phytophthora infestans*, one of the most destructive pathogens of *Solanum tuberosum*. This hemi-biotrophic plant pathogen is the causing agent of the disease known as Late Blight of Potato, the same plant disease responsible for the Irish famine in the mid-nineteenth century [1]. *S. tuberosum* is the fourth most important food crop (after corn, rice, and wheat) for human consumption [11], and it is one of the most produced crops worldwide. Historically, potato late blight is the main factor affecting potato crop production [13], [14]. After its first pandemic in the middle of the XIX century, *P. infestans* has been the most destructive pathogen in plantations of this crop, leading to annual losses that could be enough to feed several hundred million people [15]. The economic value of this loss and the cost of crop protection is estimated at 6.7 billion dollars a year [16].

Even though computational modeling of potato late blight has been studied since the early 1950s, it was not until recent years that important molecular information to feed these models was accessible to the research community. Typically, this information has been gathered by heterogeneous approaches, which have covered several research fields, such as structural and comparative genomics, protein-protein interactions and differential gene expression [17]. In particular, in the field of gene expression, several investigations have been developed with microarrays based on cDNA clones [19], transcriptome analysis through DeepSAGE [29] and RNA-Seq analysis [18].

These approaches have generated large amounts of data and have allowed us to partially understand some of the main mechanisms involved in potato late blight. Nevertheless, due to the heterogeneity of this information, our view of the disease is still fragmented. Thus, in order to better understand the molecular mechanisms underlying the compatible interaction inside the host cell, it is highly convenient and necessary to gather and integrate this information into a single computational model of the disease. This model should capture the time-dependent nature of this biological system, integrate various ranges of spatial and temporal scales, and allow us to explore the possible molecular mechanisms underlying the compatible interaction.

In recent years, the scientific community has undergone multiple efforts to integrate functional genomic characterization and biochemical knowledge into models known as genomic-scale metabolic reconstructions (GEMR), which are based on a detailed knowledge of the system's individual components (functional annotation), in order to perform the reconstruction through a Bottom-up approach, with the aim of understanding the properties that arise from the interactions between the metabolic system's components, and how these behave through time and in response to environmental stimuli [46], [65]–[67].

During the past years, this approach has been widely implemented in plants, with the development of GEMRs for several species, including *Arabidopsis thaliana* [47], [48], [62], [68], *Hordeum vulgare* [49], *Zea mays* [52]–[54], *Sorghum bicolor* [52], *Saccharum officinarum* [52]; *Brassica napus* [55], [56], *Oryza sativa* [57], and *Solanum lycopersicum* [58]. In general, the development of these models can be synthesized into four fundamental steps: 1) automatic reconstruction based on a genome annotation and biochemical databases; 2) manual refinement of the reconstruction through a literature review and biochemical and metabolic databases. In this step, each gene and reaction are verified, so that they are correctly located and connected; 3) mathematical and computational formalization of the biochemical information, where the system's specific conditions and limits are defined, this model is validated through multiple iterations and is used to prospectively simulate the system's phenotypic behavior; and 4) verification, evaluation, and validation of the

reconstruction, in other words, the model's capacity to synthesize, for example, biomass precursors, is evaluated. Generally, this evaluation leads to the identification of incorrect metabolic functions in the reconstruction (network gaps), which are once again evaluated by steps 2 and 3. Therefore, the whole process is iterative and the model is generally susceptible to being continually refined [69].

Once the mathematical model is generated in step 3, the phenotypic predictions of the organism's metabolic activity can be established through modelling methods based on restrictions, which include an *in silico* simulation of the model under inferred metabolic objectives and a set of restrictions that represent genetic or environmental conditions [70], [71]. A widely used method is flux balance analysis (FBA), which performs the linear optimization of a metabolic network in steady-state, in order to predict an optimal set of flux values that are coherent with the maximization or minimization of the chosen objective (which will depend on the purpose of the study) and under restrictions imposed on the reaction fluxes [41].

Given that genomic information used to generate the metabolic reconstructions does not consider the real expression of each gene or the subcellular localization of gene products, flux restrictions based on different "omics" data (such as transcriptomics, proteomics, and metabolomics) must be integrated into the GEMRs, in order to recreate specific metabolic phenotypes [54], [59], [63]. This approach has a powerful reach for gaining knowledge of the molecular and biochemical mechanisms of plants under specific environmental or genetic conditions. In addition, this approach allow us to contextualize high-throughput data as well as to guide hypothesis-driven discovery or to identify novel network properties [46].

In this study, we report the first genome-scale reconstruction of *S. tuberosum* and generate a genome-scale metabolic model of the compatible interaction between *S. tuberosum* and *P. infestans*, through the incorporation of expression data of the pathosystem into the model (see Materials and Methods for details). Previous work on the compatible interaction has already identified a decrease in photosynthetic activity of infected plants, however, the

underpinnings behind it remain unknown [19]. Hence, this model was aimed to follow up on those findings and try to understand the molecular basis of this typical photosynthesis turn off during plant disease. Furthermore, the model will be useful for understanding other molecular mechanisms involved during the pathosystem and for proposing novel directed hypotheses, guiding research in order to conduct metabolic engineering in the plant, and identifying emerging properties of the compatible interaction, which could otherwise not be observed through the study of individual molecules and processes.

1.2 Results and discussion

The reconstruction of the genome-scale metabolic model follows six major steps: (1) automatic reconstruction of draft network via homology searches for the identification of metabolic activities and biochemical reactions; (2) manual and semiautomatic refinement of the reconstruction (3) establishment of gene-protein-reaction (GPR) association; (4) generation of a genome-scale metabolic model in steady state; (5) incorporation of gene expression data of the compatible interaction plant-pathogen into the metabolic model; (6) flux balance analysis using objective function. In addition, model predictions are contrasted against experimental observations.

1.2.1 Metabolic reconstruction

The draft network included 1288 reactions and 1482 metabolites, 217 of which were dead-ends that decreased the metabolic network's connectivity. A manual search of these metabolites allowed us to identify 448 biochemical reactions, with biological and/or genomic evidence for potato, which refined 40 metabolic pathways totally or partially, including glycolysis/gluconeogenesis, pyruvate metabolism, glyoxylate and dicarboxylate metabolism (photorespiration module), oxidative phosphorylation, carbon fixation in photosynthetic organisms, amongst others (Supplementary Table 1-2). Additionally, based on previous reports, we manually reviewed complete or partial pathways that induced a defense response in the plant, such as those related to accumulation of salicylic acid (ubiquinone and other terpenoid-quinone biosynthesis) and jasmonic acid (alpha-Linolenic acid metabolism) [25], as well as other pathways related to PAMP (pathogen associated

molecular patterns) signaling cascades (Supplementary Table 1-2). By implementing automatic-specific gap filling and semiautomatic-specific gap find and gap fill, we were able to include 508 reactions in the reconstruction. The manual refinement and semiautomatic processes of the reconstruction are summarized in table 1.

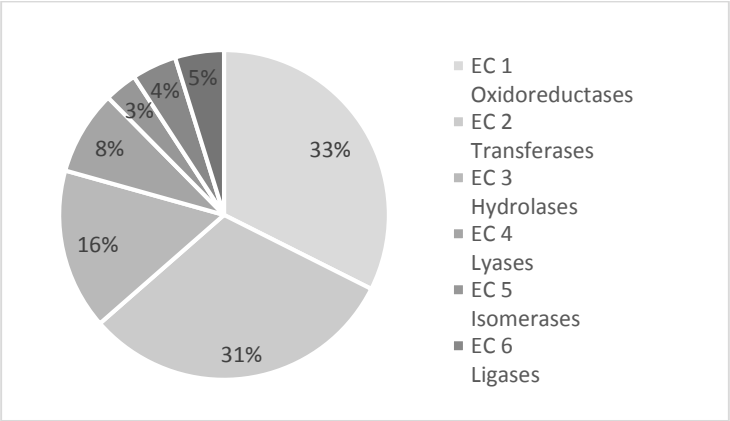
Table1-1. Summary of the manual and semi-automatic refinement connectivity reconstruction

Refinement	New reactions	New metabolites	New pathways	Complete and partially complete pathways
Manual	448	215	6	42
Automatic and Semiautomatic	508	324	x	x
Total	956	539	6	42

1.2.2 General model properties

Hereby, we present a metabolic model of *S. tuberosum* in SBML format, hereafter denoted PstM1 (Supplementary Data File S1). This model accounts for 2765 genes, 1113 metabolic functions, 1773 GPR associations and 1938 metabolites involved in 87 central and peripheral metabolic pathways and 2072 reactions, of which 1254 could carry a non-zero flux given different objective functions and the specified biomass components. The required enzymatic activities (according to Enzyme Commission - EC) to catalyze the reactions in the model are shown in (Figure 1-1). Among the 2072 reactions, 2059 represent biochemical conversions and 13 represent exchange reactions with the environment, to describe the uptake/secretion of inorganic compounds (NO_3, NO_2^-, Nitrile, CO_2, O_2, SO_4^{2-}, H_2S, PO_4^{3-}, H_2O, SO_4^{2-}, SeO_4^{2-}, NH_3, Fe_2^+) and light (photon). FBA solutions showed that the model was able to simulate leaf biomass synthesis, which was represented by the conversions of biomass precursors: protein (amino acids), sugars, nucleotides, cell wall (cellulose, lignin precursors) and fatty acids (hexadecanoic acid).

Figure 1-1: Required enzymatic activities to catalyze the reactions in the model.



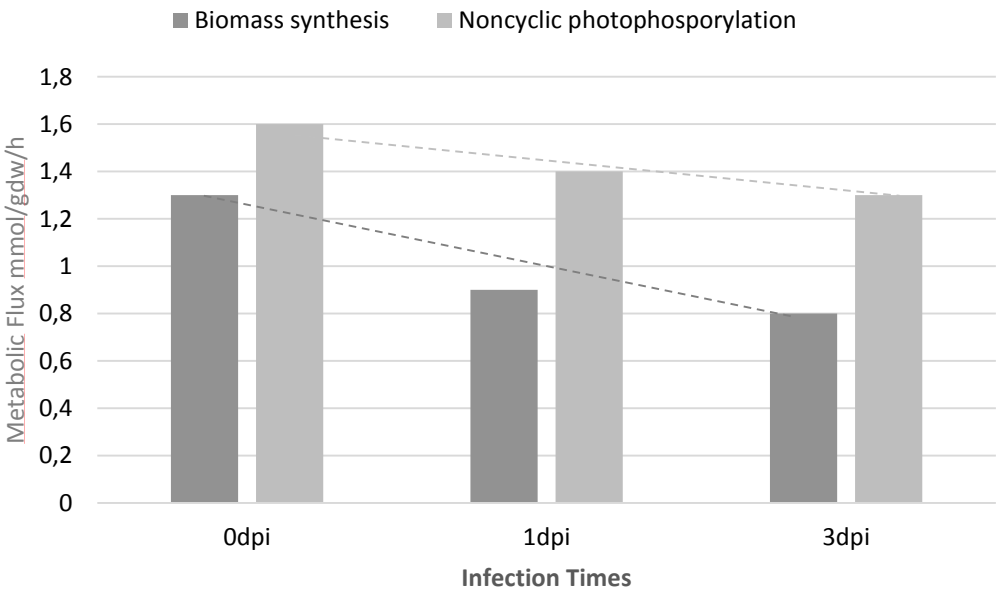
1.2.3 Metabolic phenotype analysis of the compatible interaction between *S. tuberosum* and *P. infestans*

In this study, we evaluated the plant metabolic phenotype during the first several hours and days of compatible potato late blight interaction. Biomass synthesis was selected as the objective function, because generally the mode of penetration of *P. infestans* in potato foliage is direct, through the epidermal cells [72]–[74], where the main method of entry is into the leaves [75].

To determine the metabolic profile for each infection time, we performed a FBA with flux boundaries established from the gene expression data (see Methods). Our study evidenced a decrease in the objective function (biomass synthesis of leaf) during the compatible interaction: 28% one day post inoculation (dpi) and 33% at 3dpi, from the initial time 0dpi (Figure 1-2). These model predictions can be understood if we take into account that for the first interaction days, the pathogen grows well intercellularly and then intracellularly [19], [72] mainly in the leaves [2]. We hypothesize that this pathogenic invasion causes loss of metabolic reactions capacity of the primary metabolism to synthesize macromolecules in the leaf and therefore in whole plant. The plants absorb water and minerals from the soil through their root system. These organic nutrients are translocated upward through the xylem until the leaves. All organic nutrients of plants are produced in the leaf cells, following photosynthesis, and are translocated downward and distributed to all the living plant cells.

When a pathogen destroys part of the photosynthetic tissue of plants and causes reduced photosynthetic output, this often results in smaller growth of these plants and smaller yields [76].

Figure 1-2: Metabolic fluxes of the biomass synthesis and noncyclic photophosphorylation reactions during compatible interaction between *S. tuberosum* and *P. infestans*. The flux reaction of biomass synthesis and noncyclic photophosphorylation decrease at 1 dpi and 3dpi.



1.2.4 Photosynthesis and carbon fixation metabolism |

Photophosphorylation, the synthesis of ATP in photosynthesis, occurs in chloroplasts [77]. Noncyclic photophosphorylation requires photosystem I and photosystem II, and involves water oxidation, oxygen evolution and reduction of an acceptor. In contrast, cyclic photophosphorylation is driven only by photosystem I, works with light of wavelengths over 680 nm, and does not require water oxidation and oxygen evolution [78]. *S. tuberosum* PstM1 includes the definition of the following balanced photosynthetic light reactions taken from literature sources [57], [79], [80]:

7 Photon[c] + 3 ADP[c] + 3 Orthophosphate[c] => 3 ATP[c]

(1-1)

28 Photon[c] + 9 ADP[c] + 7 NADP+[c] + 9 Orthophosphate[c] + 9 H2O[c] => 7 Oxygen[c] + 7 H+[c] + 7 NADPH[c] + 9 ATP[c]

(1-2)

Our results showed that the metabolic flux for noncyclic photophosphorylation lightly diminished through infection times (Figure 1-2) and the flux for the cyclic photophosphorylation reaction was consistently zero. The cyclic photophosphorylation reaction supplies only ATP, and the participation of this reaction in CO_2 assimilation is needed because the ATP formed in noncyclic reaction alone is insufficient for CO_2 assimilation to the level of carbohydrate [81]. The results suggested that during the interaction there was a reduction in the metabolic capacity of noncyclic photophosphorylation, and cyclic phosphorylation cannot supply the ATP deficit. These light reactions in a healthy plant convert light energy into chemical energy that is used for many cellular reactions that contribute to biomass synthesis [82], [83]. The low capacity in light reactions observed here could affect the flux of precursor reactions of biomass and, consequently, affect the capacity to synthesize biomass in the model. However, the flux of noncyclic photophosphorylation decreased less than biomass synthesis, which could be explained by the energy from photophosphorylation that is consumed by maintenance reactions. Some of these are associated with growth, for example, to maintain the electrochemical gradients across the plasma membrane, whereas others are independent of the specific growth rate of the cells [83]

The ATP formed at the expense of light energy in both cyclic and noncyclic photophosphorylation is used for converting $C0_2$ into carbohydrates [81]. The decrease in photosynthetic ability during compatible interaction between *P. infestans* and *S. tuberosum* has been reported in other studies. For example, the decline in the efficiency of photosystem II [84] and downregulation of genes encoding proteins involved in photosynthesis [19]. In the optimization of the metabolic network in steady state, we are able to predict the reaction flux, which is the overall rate of metabolite conversion [85]. [85]. We evaluated the fluxes distribution of the fixation carbon pathway to evaluate the photosynthetic capacity of the

plant during the infection times. The corresponding metabolic flux distribution patterns are shown in (Figure 1-3). A table listing the simulated flux values for each infection time is given in Supplementary Table 1-3.

In general, the metabolic fluxes in the carbon fixation cycle (also known as the Calvin cycle) during potato late blight is characterized by the decrease and loss of reaction fluxes necessary to convert the inorganic carbon in organic carbon (Figure 1-3). At 0 dpi, the metabolic fluxes of the reactions in the fixation carbon cycle varied between 5.79 and 51.42 mmol/gdw/h, suggesting interconversion of metabolites in all reactions in this pathway. At 1 dpi and 3 dpi, the variation of the metabolic fluxes in response to the compatible interaction was from 0 to 28 mmol/gdw/h and from 0 to 25.7 mmol/gdw/h, respectively. At 3 dpi, six reactions completely lost their flux capacity, showing a higher decrease of the global metabolic capacity in this time compared with the two previous times of infection.

The Calvin cycle involves three main phases: 1) carboxylation of ribulose-1,5-bisphosphate (RuBP) to form 3-phosphoglycerate (PGA), mediated by Ribulose-1,5-bisphosphate carboxylase/oxigenase enzyme (RuBisCO); 2) reduction of PGA to the level of a sugar (CH_2O) by formation of glyceraldehyde 3-phosphate (GAP) using NADPH an ATP produced in the light reactions; and 3) the regeneration of RuBP [86]. Hereafter, these stages are described for the three times of infection and the subsequent changes in the metabolic behavior are specified.

The highest flux of carboxylation of RuBP was obtained in the model optimization of 0 dpi (Figure 1-3a). At this time, we observed that the RuBisCO enzyme had the best carboxylase activity. In contrast, at 1 dpi, this enzymatic activity decreased, but increased the oxygenase activity of the RuBisCO, reducing the energy efficiency of photosynthesis in the plant [87], [88].

The subsequent metabolism of glycolate produced by the oxygenation of the RuBP is known as photorespiration, and is associated with high light intensity, uptake of O_2 and increased

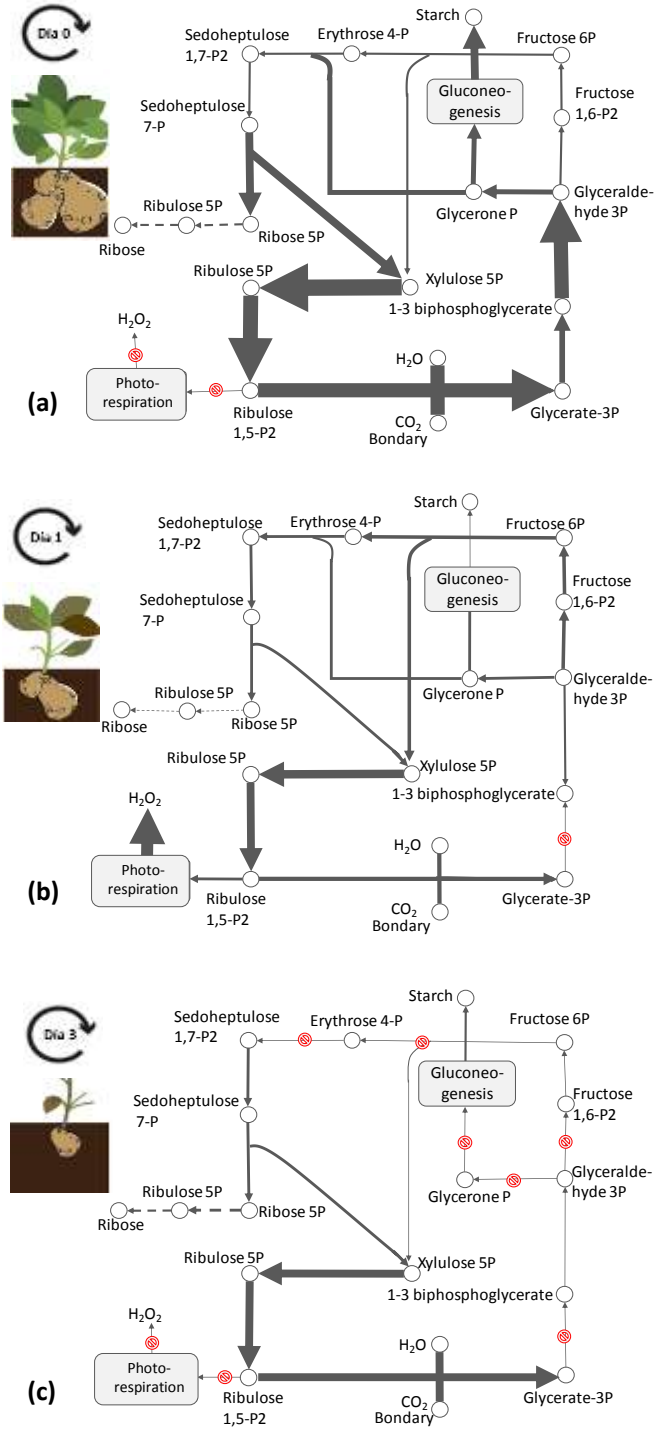
H_2O_2 production [89], [90]. In the solution of this metabolic model at 1 dpi the uptake of photons is not enhanced, but H_2O_2 production is increased, which is one of the major and the most stable reactive oxygen species (ROS) that regulates basic acclamatory, developmental and defense processes in plants [91], including programmed cell death (PCD) [92].

One of the most rapid defense responses against pathogen attack is the oxidative burst, which consists of the production of ROS, primarily superoxide and H_2O_2, at the site of attempted invasion [93]. The quick generation of H_2O_2 in potato tuber tissue following inoculation with *P. infestans* was previously reported [94]. Our results suggested that the increase in photorespiration compared to the decrease in carboxylation of RuBP can be associated to the plant's need to trigger a quick defense mechanism, given that photorespiration would appear to be the fastest process for generating H_2O_2 [95].

At 3 dpi, the oxygenation of the RuBP flux reaction disappeared, allowing an increase in the carboxylase activity of the RuBisCO. However, the flux of the carboxylation reaction was reduced by 55% compared to 0 dpi. Despite of the defense response through the oxidative burst observed in the previous time, the metabolic flux to H_2O_2 production here was zero. During incompatible plant- pathogen interaction, an initial and very rapid accumulation of H_2O_2 is followed by a second and prolonged burst of H_2O_2 production. In addition, the activity and levels of the ROS detoxifying enzymes are suppressed by salicylic acid (SA) [96]. The suppression of ROS detoxifying mechanisms is crucial for the induction of PCD [97], [98], which potentially limits the spread of disease [99]. In our model optimization, only one peak of H_2O_2 production occurs (Figure 1-3b). Interestingly, this behavior has been reported during compatible plant-pathogen interaction [100]. In addition, the metabolic flux that represents the synthesis of SA in PstM1 is consistently zero at the time points of infection evaluated. This can indicate that ROS-scavenging mechanisms are not downregulated in the metabolic model. Thus, just one peak of

Figure 1-3: Metabolic flux distribution patterns of the fixation carbon pathway and some reactions associated with photorespiration and carbohydrate synthesis. (a) Metabolic flux distribution patterns

for 0dpi; (b) metabolic flux distribution patterns for 1 dpi; and (c) metabolic flux distribution patterns for 3 dpi.



oxidative burst observed at 1 dpi is not sufficient to trigger an effective defense response in the plant; first, because a prolonged production of H_2O_2 is required, and second, because

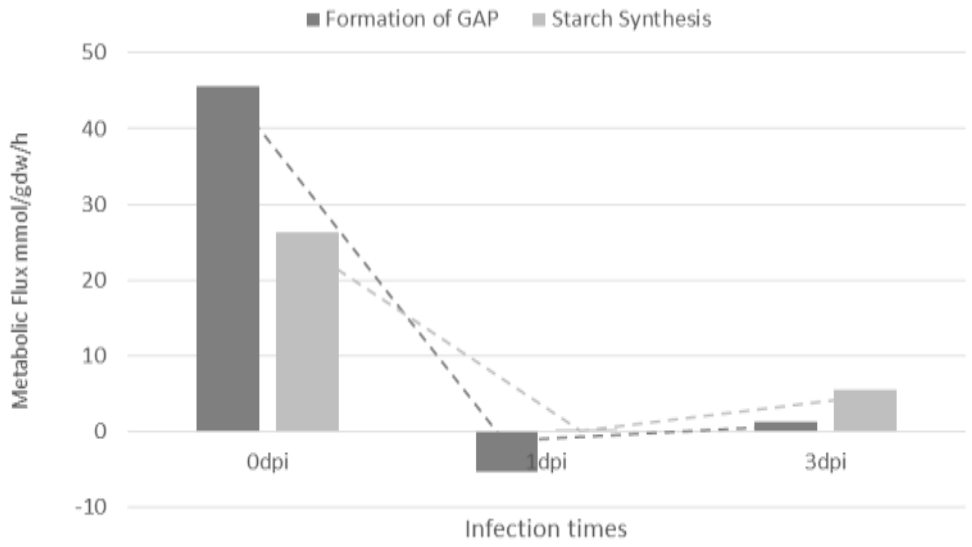
ROS production without suppression of ROS detoxification does not result in the induction of PCD [97], [98].

In the reaction of GAP formation, the fluxes fall deeply from 0 dpi to 1 dpi (Figure 1-3), where the anabolic capacity of this reaction is null. At 3 dpi, the metabolic flux increased incipiently, likely due to the recovery of carboxylation of the RuBisCO. The GAP molecules can be used for the synthesis of sucrose or starch, and alternatively can be used to regenerate RuBP [86], [101]. We compared the fluxes of the starch synthesis reaction with the fluxes of PGA transformation in GAP through all time points of infection, based on the idea that starch production is a good indicator of a healthy plant, which will only store extra reservoirs of starch in tissues if all their energy requirements are already fulfilled [102]. In addition, starch has been identified as an important integrator in plant growth regulation to cope with continual changes in carbon availability when the rate of photosynthesis is modified by environmental constraints [102], [103]. The comparison between these two reactions showed the same change trend in metabolic fluxes (Figure 1-4). At 1 dpi, the metabolic flux to synthesize GAP and starch is the lowest, and at 3 dpi, the flux of this reaction increased incipiently. Our results indicated that the loss of carbon fixation capacity of the RuBisCO and the subsequent decrease in GAP formation from PGA, can be related to the decreased efficiency of starch biosynthesis.

Previous studies have shown that starch biosynthesis regulation is linked with the expression of ADP-Glc pyrophosphorylase (AGPase) in leaves [103], [104]; likewise, AGPase activity is generally activated by PGA [105]. Therefore, AGPase activation or inactivation by PGA production allows starch synthesis to be adjusted in response to changes in photosynthesis [106]. We found that the metabolic flux of the AGPase activity decreased slightly through infection times (view Supplementary Table 1-3). This can be associated with the general decline in the efficiency of starch synthesis in response to the decrease of biosynthesis of GAP from PGA. Although the AGPase activity and starch biosynthesis both tend to decrease, they show particular tendencies that are not comparable between them, especially from 1 dpi to 3 dpi (view Supplementary Table 1-3). Hence, additional regulatory

mechanisms are required to achieve changes in the rate of starch biosynthesis, as previously reported [103], [107]–[110].

Figure 1-4: Metabolic fluxes of GAP formation and starch synthesis. The fluxes of these reactions showed the same trend through infection time.



In the Calvin cycle, the RuBP is both consumed and synthesized [111]. The synthesis of RuBP is known as regeneration, and involves a series of reactions from GAP, that are energetically favorable and do not consume ATP or NADPH, except in the phosphorylation reaction to transform ribulose 5-phosphate into RuBP [112]. The fluxes distribution of these reactions are affected during compatible interaction (Figure 1-3) mainly at 3 dpi, where a null flux is obtained for three reactions. The net flux of transformation of ribulose 5-phosphate in RuBP reaction decreases approximately by 55% at 1 dpi and 3 dpi compared to 0 dpi. This can be a consequence of both the decrease in the metabolic flux of its precursors reactions in the cycle as well as in the decrease of the metabolic capacity of ATP production for the light reactions. In addition, we observed that at 0 dpi and 3 dpi the metabolic fluxes of the regeneration and carboxylation of RuBP are directly proportional. At 1 dpi, the metabolic flux of the regeneration was directly proportional to the sum of the flux for carboxylation and oxygenation. Overall, these results can indicate that in our model

the RuBP is synthetized in the same rate that it is consumed. This observation agrees with previous studies that demonstrated that, in a photosynthetic steady state model, the rate at which RuBisCO consumes RuBP equals the rate at which RuBP is regenerated [113], [114].

1.3 Conclusions

In this work, we simultaneously introduced the first metabolic network of *S. tuberosum* and the first genome-scale metabolic model of the compatible interaction of a plant with *P. infestans*. The metabolic flux of the light reactions and carbon fixation cycle, including photorespiration and starch synthesis, suggest a suppression of the photosynthetic capacity as consequence of the compatible interaction between *P. infestans* and *S. tuberosum.* The results shown here are *in silico* metabolic predictions, which closely match previous studies of plant physiology. We recommend the compartmentalization of the model for a better evaluation of flux between cellular organelles.

1.4 Materials and Methods

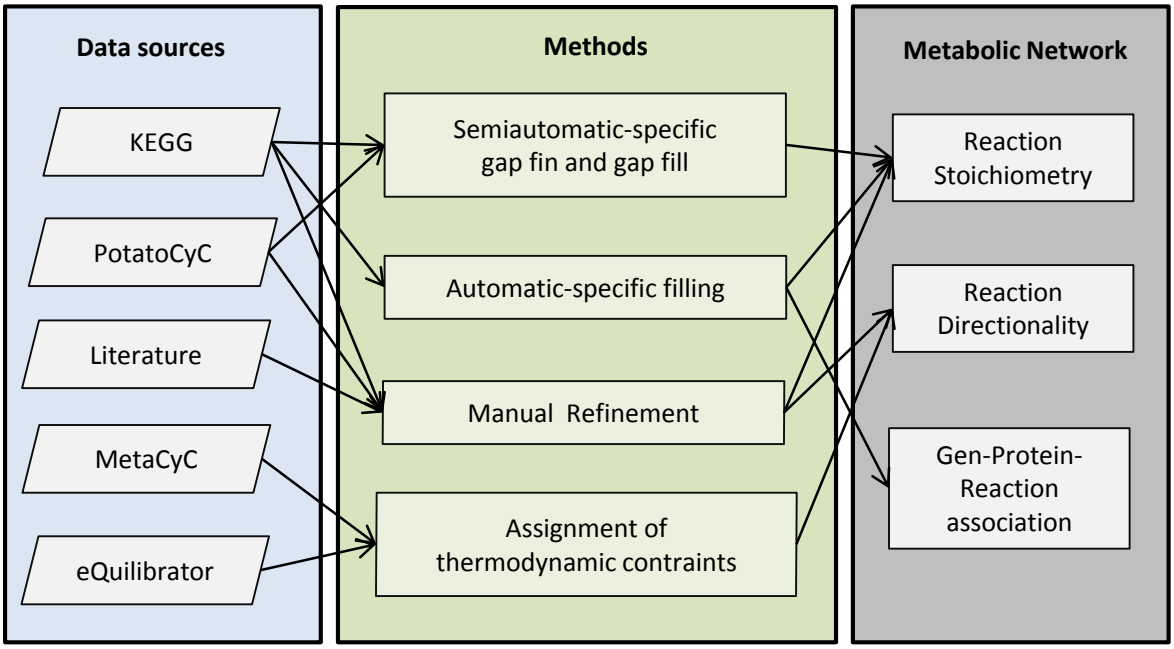
1.4.1 Metabolic reconstruction.

An initial automatic draft reconstruction was first created by means of Subliminal [115] and RAVEN [116] software toolboxes starting from the PGSC potato genome sequence [30]. Both automatic reconstructions were then conciliated and merged through in-house scripting. This merged reconstruction was taken as a starting point for manual reconstruction and then subjected to a manual refinement based on literature and available biological data.

1.4.2 Metabolic reconstruction refinement

The reconstruction refinement stage was divided into six phases: (1) manual gap refinement of the metabolic network and manual refinement of the reversibility and directionality of the reactions; (2) automatic-specific filling; (3) semiautomatic-specific gap find and gap fill; (4) establishment of the directionality and reversibility of the reactions through Gibbs free energy of reaction value ($\Delta rG'^{\circ}$) [117]; (5) inclusion of exchange reactions; and finally (6) gene-protein-reaction associations (Figure 1-5).

Figure 1-5: General scheme of the refinement of potato metabolic network. The lines indicate the trajectory from data sources to the refined metabolic network.



During manual refinement, the network's connectivity was verified in a pathway by pathway basis, as well as in a metabolite by metabolite basis [69]. For this, a search for dead-end metabolites in the reconstruction was performed by means of the R package minval [118]; these metabolites corresponded to those that are consumed in the set of biochemical reactions, but are not synthesized and viceversa. Based on the identified dead-end metabolites, missing reactions in the reconstruction were manually tracked within the KEGG Pathway Maps database [119], [120], which associates organism-specific genomic information to metabolic pathways maps. For reactions not reported for *S. tuberosum* in the KEGG Pathways database, we verified that their catalyzing enzymatic activity was reported in the Plant Metabolic Network (PMN) database specific for potato (PotatoCyc) (http://pmn.plantcyc.org/POTATO/organism-summary). In addition, based on biochemical literature, carbohydrate and energetic metabolic pathways were further refined. During the entire manual review process of the network's connectivity, reactions without any type of biological and/or genomic evidence for this plant were removed, and the reversibility and directionality of 120 reactions were corrected.

In order to verify that the metabolic pathways reported in this reconstruction were present in the metabolism of *S. tuberosum*, we implemented the KEGGREST R package, which allows to obtain all of the metabolic pathways reported for this organism in the KEGG pathways database [121]. By comparing the metabolic pathways obtained from KEGG with those in the reconstruction, we found 12 non-reported pathways, which were then removed. Since some of the reactions belonging to these pathways were catalyzed by enzymatic activities reported in PotatoCyc, these were included in the reconstruction without being associated to any specific metabolic pathway.

During the process of automatic-specific filling, a database construction of all the biochemical reactions reported for *S. tuberosum* in KEGG was done using the g2f R package [122]. Later, only missing reactions were added to the reconstruction, using the sot_g2f in-house script (available in GitHub: https://github.com/kellybotero/PotatoRecon). Since the network still showed gaps, a semiautomatic-specific gap finding and gap filling process was performed using the g2f package. For this, once again we identified all of the dead-end metabolites in the network and automatically tracked them to a reference database that contains all of the biochemical reactions stored in KEGG. The gap fill search method was restricted to retrieve only reactions that showed metabolites in the reconstruction. Finally, through manual validation, we included only reactions with enzymatic activities reported for *S. tuberosum* in PotatoCyc.

The reversibility and directionality of the reactions that were not manually corrected were determined through $\Delta_r G'^{\circ}$ values obtained from EQuilibrator [123] and MetaCyc [124] databases. These databases establish the value for $\Delta_r G'^{\circ}$ by calculating the Gibbs free energy of formation of a compound ($\Delta_f G'^{\circ}$), through the group contribution method for thermodynamic analysis [117]. The parameters used to calculate $\Delta_f G'^{\circ}$ in Equilibrator are pH 7.0 and ionic strength 0.1, and in MetaCyc these are pH 7.3 and ionic strength 0.25. Since the parameters are located in a pH range of 7.0 – 7.3 and ionic strength of 0.1 – 0.25, the reversibility R script (available in GitHub: https://github.com/kellybotero/PotatoRecon) was implemented, which allowed us to compared the $\Delta rG'^{\circ}$ values from both databases.

When both reversibilities were below -1 (integrated open framework for thermodynamics of reactions that combines accuracy and coverage), the GEMR reversibilities were established as forward irreversible. Due to lack of biological evidence and discrepancies between the $\Delta rG'°$ values of the databases, the remaining reactions were determined as reversible.

During the fifth phase of the reconstruction refinement process, we included 13 exchange reactions which transport metabolites that cannot by synthesized by the cell, within the limits of the system to the cytoplasm, and are precursors to other metabolites [48]. Finally, once the reconstruction was refined in terms of its reactions, metabolic pathways and reversibility, GPR associations were integrated using the g2f package. This package constructs GPRs based on KEGG ORTHOLOGY of the reactions present in the reconstruction.

1.4.3 Transformation of the reconstructed network into a genome-scale metabolic model

From the genome-scale metabolic reconstruction of potato, a SBML version was generated using the R package minval. Consecutively, the SBML file was imported to the R package Sybil [125], where the whole set of reactions present in metabolic reconstruction was transformed into a steady state metabolic model $S_{ij.}v_j=0$. Where S_{ij} is the entries of the stoichiometric matrix (Supplementary Data File S2). Rows in S represent metabolites and columns represent reactions, and v_j is the metabolic flux vector for each reaction. Substrates in the reaction have negative coefficients, while products have positive ones. This matrix also takes into account transport reactions across the cell membrane, which are represented as reactions interconverting intracellular and extracellular compounds. In the metabolic model, the reaction fluxes are subject constrains ($- v_{jmin} \leq v \leq v_{jmax}$) [49], [126].

Objective function was defined by the biomass synthesis reaction in the leaf $OF_{bioamass}$ (Supplementary Data File S2), previously reported in the genome-scale metabolic model of *Arabidopsis thaliana* [68]. This objective function was mathematically written as a

combination of metabolic coefficients of the biomass composition estimated for slow-growing species [127], [128].

1.4.4 Incorporation of gene expression data into the genome-scale metabolic model

Gene expression data of the compatible interaction were obtained from a study of Gyetvai et al., 2012. We used the normalized libraries of the untransformed cultivar Désireé at three infection time points with *P. infestans* (zero (0 dpi), one (1 dpi) and three (3 dpi) days post inoculation), with three biological replicates. The tag annotation was performed by BLASTN [129]. The source for the tag annotation was the available RNA sequence of "Potato 3.0" (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF_000226075.1_SolTub_3.0/). The unique genes and maximum value counts per gene were performed with the script summarization.r (available at GitHub: https://github.com/kellybotero/PotatoRecon). As the last step for building the gene expression database, all refseq gene identifiers were transformed to Entrez identifiers , by means of the R package UniProt.ws [130].

From the previously built gene expression database, we generated an expression set for each infection time point by means of the R package Biobase [131]. The gene expression values were incorporated directly into fluxes constraints of the reactions in the model using the R package ex2flux [132], which implements a method to integrate gene expression values into each GPR associated to the reactions of the model.

1.4.5 Metabolic flux model optimization

FBA represents a constraint-based modeling approach that allows the prediction of metabolic steady-state fluxes, by applying mass balance constraints into a stoichiometric model (Varma and Palsson, 1994). The reactions in the model can be represented by a linear system of equations, then, problems such as maximizing specific chemical production or growth can be solved by linear programming [46], [133]. With the purpose of obtaining a computational distribution of metabolic fluxes, FBA was employed assuming maximization of the objective function, by means of the R package Sybil (Equation. (1-3)).

$$\text{maximization } OF_{bioamass}$$

subject to

$$\sum_{i=1,2\ldots m;\ j=1,2\ldots n} S_{ij}.v_j = 0$$

$$-v_{jmin} \leq v \leq v_{jmax} \qquad (1\text{-}3)$$

To constrain the space of all the possible steady-state flux distributions in the optimization we impose thermodynamic constraints to reaction reversibility as well as upper and lower bounds constraints on reactions fluxes from known expression values for the particular enzyme that catalyzes the reaction.

1.5 Supplementary

Supplementary Table 1-2: Refined pathways in the metabolic network reconstruction. We show all the pathways included in the metabolic network of *S. tuberosum*. Pathway association was assigned based on the categorization of the KEGG Pathways database. The colors correspond to the level of curation of each pathway

Complete	Partially complete	Intervened	Not intervened

Alpha-Linolenic acid metabolism	Arginine and proline metabolism	Ascorbate and aldarate metabolism	Carbon fixation in photosynthetic organisms	Citrate cycle (TCA cycle)	Flavonoid biosynthesis	Galactose metabolism	Inositol phosphate metabolism
Nitrogen metabolism	Oxidative phosphorylation	Photosynthesis (ligth reactions)	Porphyrin and chlorophyll metabolism	Pyruvate metabolism	Steroid biosynthesis	Sulfur metabolism	Alanine, aspartate and glutamate metabolism
Benzoxazinoid biosynthesis	Brassinosteroid biosynthesis	Carotenoid biosynthesis	Cyanoamino acid metabolism	Cysteine and methionine metabolism	Diterpenoid biosynthesis	Fructose and mannose metabolism	Glutathione metabolism
Glycerolipid metabolism	Glycerophospholipid metabolism	Glycolysis / Gluconeogenesis	Glyoxylate and dicarboxylate metabolism	Histidine metabolism	Lysine biosynthesis	Lysine degradation	Other types of O-glycan biosynthesis
Pantothenate and CoA biosynthesis	Propanoate metabolism	Selenocompo-und metabolism	Sphingolipid metabolism	Starch and sucrose metabolism	Terpenoid backbone biosynthesis	Ubiquinone and other terpenoid-quinone biosynthesis	Zeatin biosynthesis
Thiamine metabolism	Amino sugar and nucleotide sugar metabolism	Biosynthesis of unsaturated fatty acids	Butanoate metabolism	Caffeine metabolism	Carbon metabolism	Fatty acid elongation	Fatty acid metabolism
Glycine, serine and threonine metabolism	Linoleic acid metabolism	Monoterpenoid biosynthesis	Pentose phosphate pathway	Phenylalanine metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis	Phenylpropanoid biosynthesis	Purine metabolism
Pyrimidine metabolism	Tryptophan metabolism	Valine, leucine and isoleucine biosynthesis	Aminoacyl-tRNA biosynthesis	Anthocyanin biosynthesis	Arachidonic acid metabolism	Arginine biosynthesis	beta-Alanine metabolism
Biotin metabolism	Ether lipid metabolism	Fatty acid biosynthesis	Flavone and flavonol biosynthesis	Folate biosynthesis	Glucosinolate biosynthesis	Glycosaminoglycan degradation	Glycosphingolipid biosynthesis - globo series
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	Lipoic acid metabolism	Limonene and pinene degradation	N-Glycan biosynthesis	Nicotinate and nicotinamide metabolism	One carbon pool by folate	Pentose and glucuronate interconversions	Riboflavin metabolism
Stilbenoid, diarylheptanoid and gingerol biosynthesis	Synthesis and degradation of ketone bodies	Taurine and hypotaurine metabolism	Tropane, piperidine and pyridine alkaloid biosynthesis	Tyrosine metabolism	Valine, leucine and isoleucine degradation	Vitamin B6 metabolism	

Supplementary Table 1-3: Reaction fluxes of the metabolic pathways of interest in *S. tuberosum* to evaluate photosynthetic capacity through compatible interaction with *P. infestans.*

Pathways	Reaction (ID KEGG)	Metabolic flux (0 dpi)	Metabolic flux (2 dpi)	Metabolic flux (3 dpi)
Biomass Synthesis	RBS01	1,313856	0,9435836	0,8790459
Cyclic Photophosphorylation	RK0004	0	0	0
Non Cyclic Photophosphorylation	RK0005	1,630357	1,408214	1,35
Carbon Fixation	R00024	46,38961	14,34539	25,7072
	R01512	19,59667	0	0
	R01061 *inv	-45,65	5,326667	-1,176667
	R01015	17,39	8,923333	0
	R01068 *inv	-5,79	-13,63667	0
	R00762	5,79667	13,63667	0
	R01067	5,79667	13,63667	0
	R01829	5,79667	12,67667	0
	R01845	5,79667	12,67667	13,94333
	R01641	25,52	12,91333	10,93469
	R01056	-5,036667	-2,386667	-2,251979
	R01523	46,3891	25,61669	25,7072
	R01529 *inv	-51,426	-28,00336	-25,65781
	R01051	-5,036667	-3.62	-5,69
Photorespiration	R03140	0	11,2713	0
	R01334	0	11,2713	0
	R00475	0	39,43	0
	R00009	0	36,37207	3,992193
	R00372	0	0	65,06771
	R00945	-40,9	2,556667	-2,646667
	R01221	28,57	10,43925	14,86955
	R00588	693,29	1543,01	854,7267
	R01388	-4,417	-3,146667	-19,57333
	R01514	0	0	0
AGPase activity	R00948	0.3481719	0.2500497	0.2329472
Starch Synthesis	R02110	26,29711	0,25	5,42
Salicylic Acid Synthesis	RK0003	0	0	0

* inv = The direction of the metabolic flux in the metabolic pathway is inverted

2. Finding and filling gaps in metabolic networks through the ‘g2f’ R package

- Original title: g2f: An R package for finding and filling gaps in metabolic networks.

Abstract

During the construction of a genome-scale metabolic network reconstruction, several dead-end metabolites which cannot be imported/produced, or that are not used as substrates or are released by not any of the reactions are incorporated into the network. The presence of dead-end metabolites can block the net flux of the objective function when it is evaluated through Flux Balance Analysis (FBA), and when it is not blocked, then the bias in the biological conclusions increases. The refinement to restore the connectivity of the network can be carried out manually or using computational algorithms. The ‘g2f’ package was designed as a tool to find the dead-end metabolites, and fill it from the stoichiometric reactions of a reference, filtering candidate reactions using a weighting function. Also, the option to download all the set of gene-associated stoichiometric reactions for a specific organism from the KEGG database is available.

2.1 Introduction

Genome-scale metabolic network reconstructions (GMNR) specify the chemical reactions catalyzed by hundreds of enzymes (registered in enzyme commission – EC) and cover the molecular function for a substantial fraction of a genome [134]. The main goal of these network reconstructions is to relate the genome of a given organism with its physiology, incorporating every metabolic transformation that this organism can perform [59], [116].

The GMNR are converted into computational models for metabolism simulation and gaining insight into the complex interactions that give rise to the metabolic capabilities [135], [136]. The predictive accuracy of a model depends on the comprehensiveness and biochemical fidelity of the reconstruction [137].

The GMNR construction process can be synthesized into two fundamental stages: (1) draft network reconstruction, here the reactions associated to the enzymes that participate in the metabolism of a particular organism are downloaded from specialized genome, biochemical and metabolic databases; and (2) manual or computational refinement of the network. Similar steps are performed during the construction of a tissue-specific metabolic reconstruction, defined as subsets of reactions included in a genome-scale metabolic reconstruction that are highly associated to the metabolism of a specific tissue. They are constructed from measured gene expression or proteomic data and enable to characterize or predict the metabolic behavior for each tissue under any physiological condition. Given that only the reactions associated to an enzyme or gene can be mapped from the measured data, spontaneous reactions and non-facilitated diffusion are missing in first stages of a tissue-specific reconstruction.

The refinement stage of the reconstruction is a process to restore the connectivity of the network, where network gaps in the draft reconstruction are identified, and candidate reactions to fill the gap are found in literature and databases [69], [138]. Since network reconstructions typically involve thousands of metabolic reactions, their refinement can be a very complex task [116]. Network gaps can be associated to dead-end metabolites which cannot be imported/produced by any of the reactions in the network; or to metabolites that are not used as substrates or released by any of the reactions in the network. When the metabolic network is transformed into a metabolic steady-state model in order to optimize the distribution of metabolic flux under an objective function, the presence of this type of metabolites can be problematic, since the flux cannot pass through them due to the incomplete connectivity with the rest of the network [138]. In a high quality model, all reactions should be able to carry flux if all relevant exchange reactions are available [116]. The lack of flux in dead-end metabolites is propagated downstream/upstream, depending if

the metabolites are not produced or not consumed, giving rise to additional metabolites that cannot carry any flux [138]. This can block the net flux of the objective function and when it is not blocked, then bias in the biological conclusions increases.

The manual refinement is an iterative process to assemble a higher confidence compendium of an organism-specific metabolism in a draft metabolic network reconstruction [67], [139], [140]. This type of refinement requires time and a labor intensive of use of available literature, databases and experimental data [67], [141]. For example, if the genome annotation of the target organism is present in KEGG [119], reactions associated to the genes can be identified in KEGG maps and dead-end metabolites under the organism-specific metabolic environment can be tracked [69]. Any given GMNR accounts for hundreds or thousands of biochemical reactions, thus, this task is very complex and can lead to both the introduction of new errors as well as the possibility of overlooking some others.

Network gaps refinement can also be performed using algorithms to identify blocked reactions through an optimization of the metabolic model, and to fill metabolic functions gaps, by means of searching for reactions associated to the missing functions in available databases [116], [137]. This approach, however, only fills gaps for a single biological objective function, while other implemented approaches based on the identification and filling of dead-end metabolites, included in GAMS [8] and COBRA [14] packages, allow to fill the networks gaps for all of the metabolism, generating a list of reactions that enable to restore the network connectivity and eliminate blocked reactions that no carry flux in a steady state model [11]. Some drawbacks, however, are that implemented COBRA [71] gap fill algorithms operate as tools under the commercial MATLAB® environment, and GAMS [138] only allow to perform gap fill using MetaCyc databases as reference [124].
With the aim of offering an open source tool that facilitates the refinement and depuration of draft network reconstructions and metabolic models, we introduce the ‘g2f’ R package. It includes four functions to identify and fill gaps, as well as, to calculate the addition cost of a reaction, depurate metabolic networks of blocked reactions (not activated under any scenario) and a function to associate GPR to reactions included in a metabolic reconstruction using the KEGG database as a reference.

2.2 Installation and functions

g2f includes four functions and is available for download and installation from CRAN, the Comprehensive R Archive Network. To install and load it, just type:

```
> install.packages("g2f")
> library(g2f)
```

The g2f package requires R version 2.10 or higher.

2.2.1 Downloading a reference from the KEGG database

The KEGG database is a collection of databases widely used as references in genomics, metagenomics, metabolomics and other omics studies, as well as for modeling and simulation in systems biology [16]. To date, the database includes genomes, biological pathways and its associated stoichiometric reactions for 346 eukaryotes, 3947 bacteria, and 238 archaea. The getReference function downloads all the KO-associated stoichiometric reactions, and their correspondent E.C. numbers from the KEGG database for a customized organism (using the KEGG organism ID). Based on the KOs associated to a reaction, their respective GPR is constructed as follows: All genes associated to a given KO are linked by an AND operator, after that, when a reaction has more than one associated KO, previously linked genes are now joined by an OR operator. As an example, to download all (1392 reactions) stoichiometric reactions associated to *Escherichia coli* just type:

```
> E.coli <- getReference(organism = "eco")
```

2.2.2 Calculating the addition cost

Metabolite-based reaction mapping, followed by the addition of new reactions, is a very basic solution for a gap fill process, which can increase the number of dead-end metabolites. As a way to reduce the addition of new dead-end metabolites, the additionCost function calculates a cost (in terms of new metabolites) to be added, based on metabolites that constitute the new reaction and those that constitute the stoichiometric reactions present in the metabolic reconstruction, following the equation 2-1.

$$additionCost = \frac{n(metabolites(\text{newReaction}) \notin metabolites(\text{reactionList}))}{n(metabolites(\text{newReaction}))} \quad (2\text{-}1)$$

As an example, we select a sample of reactions from the downloaded reference for *E. coli* and calculate the addition cost for the remaining reactions (6 rst values are showed).

```
>       reactionList <- sample(E.coli$reaction,10)
>       head(
+       additionCost(reaction = E.coli$reaction,

+       reference = reactionList)

+ )

[1] 0.4000000 0.4000000 0.4000000 0.4000000 0.3333333 0.3333333
```

2.2.3 Performing a gap find and fill

To identify network gaps in a metabolic network and fill them from a reference, the gap Fill function internally performs several steps: (1) Dead-end metabolites are identified from the stoichiometric matrix using functions included in the minval package, (2) the candidate reactions to be added are identified by metabolite mapping, (3) the addition cost of each candidate reaction is calculated, (4) the candidate reactions with an addition cost lower or equal to the user-defined limit are added to the reaction list, and, finally, (5) the process returns to step 1 until no more original-gaps can be filled under the user-defined limit. The function returns a set of candidate stoichiometric reactions to fill the original gaps included in the metabolic network.

As an example, we show how to fill the dead-end metabolites included in the previously selected sample using all downloaded stoichiometric reactions from the KEGG database for *E. coli* as the reference

```
> gapFill(reactionList = reactionList,

+               reference = E.coli$reaction,
+               limit = 1/4
+  )


23 Orphan reactants found
```

```
12 Orphan reactants found

11 Orphan reactants found

11 Orphan products found

 [1]  "D-Mannonate + NAD+ <=> D-Fructuronate + NADH + H+"
 [2]  "ATP + Thymidine <=> ADP + dTMP"
 [3]  "dTTP + H2O <=> dTMP + Diphosphate"
 [4]  "(R,R)-Tartaric acid + NAD+ <=> 2-Hydroxy-3-oxosuccinate + NADH + H+"
 [5]  "Hypoxanthine + NAD+ + H2O <=> Xanthine + NADH + H+"
 [6]  "Ammonia + 3 NAD+ + 2 H2O <=> Nitrite + 3 NADH + 3 H+"
 [7]  "L-Glutamine + H2O <=> L-Glutamate + Ammonia"
 [8]  "ATP + Deamino-NAD+ + Ammonia <=> AMP + Diphosphate + NAD+"
 [9]  "Ammonia + NAD+ + H2O <=> Hydroxylamine + NADH + H+"
[10]  "ATP + Glutathione + Spermidine <=> ADP + Orthophosphate + Glutathionylspermidine"
[11]  "2 Glutathione + NAD+ <=> Glutathione disulfide + NADH + H+"
[12]  "Glutathione + H2O <=> Cys-Gly + L-Glutamate"
[13]  "L-Proline + NAD+ <=> (S)-1-Pyrroline-5-carboxylate + NADH + H+"
[14]  "L-Glutamate 5-semialdehyde + NAD+ + H2O <=> L-Glutamate + NADH + H+"
[15]  "ATP + Pantothenate <=> ADP + D-4'-Phosphopantothenate"
```

2.2.4 Identifying blocked reactions

To identify the blocked reactions included in a metabolic model, the blockedReactions function sets each one of the reactions included in the model (one at a time) as the objective function and optimizes the model through Flux Balance Analysis. Reactions that do not participate in any possible solution during all evaluations are returned as blocked reactions.

As an example, we identify the blocked reactions in the *E. coli* core metabolic model included in the 'sybil' package.

```
> data("Ec_core")
> blockedReactions(Ec_core)


  |======================================================================| 100%
[1] "EX_fru(e)" "EX_fum(e)" "EX_mal_L(e)" "FUMt2_2" "MALt2_2"
```

2.3 Summary

We introduced the g2f package to find dead-end metabolites included in a metabolic reconstruction, and fill them from stoichiometric reactions of a reference, filtering candidate reactions by using a weighting function and a user-defined limit. We show step by step the

functionality of each procedure included in the package using a reference downloaded from the KEGG database for *E. coli*, and the core metabolic model included in the ‘sybil’ package.

.

3. Constraining the metabolic reconstruction: Incorporating expression data as FBA limits through the 'exp2flux' R package

Original title: exp2 ux: Convert Gene EXPression Data to FBA FLUXes

Abstract

Computational simulations of metabolism can help predict the metabolic phenotype of an organism in response to different stimuli, through constraint-based modeling approaches. To recreate specific metabolic phenotypes and enhance the model predictive accuracy, several methods for integrating transcriptomics data into constraint-based models have been proposed. The majority of available implemented methods are based on the discretization of data to incorporate constraints into the metabolic models via Boolean logic representation, which reduces the accuracy of physiological representations. The methods implemented for gene-expression data integration as continuous values are very few. The exp2flux package was designed as a tool to incorporate, in a continuous way, gene-expression data as FBA flux limit in a metabolic reconstruction. Also, a function to calculate the differences between fluxes in different metabolic scenarios was included.

3.1 Introduction

Metabolism is a cellular system suited for developing studies at the systemic level of the genotype-phenotype relationship and genetic interactions [134]. Genome sequencing projects have contributed to our understanding of the metabolic capabilities in cellular systems since functional annotation of gene products (enzymes) allow the reconstruction of

genome-scale metabolic networks (GMNs) that summarize these metabolic capabilities consistently and compactly in a stoichiometric matrix [142], [143]. GMNs can be converted into computational models to predict the metabolic phenotype of an organism in response to different stimuli, through constraint-based modeling approaches [70], [71]. A widely used approach to perform *in silico* metabolic simulations is flux balance analysis (FBA), a linear optimization method which uses imposed mass balance and constraints (that represent genetic or environmental conditions) to define the space of feasible steady-state flux distributions of the network and then identify optimal network states that maximize a defined objective function [41], [134].

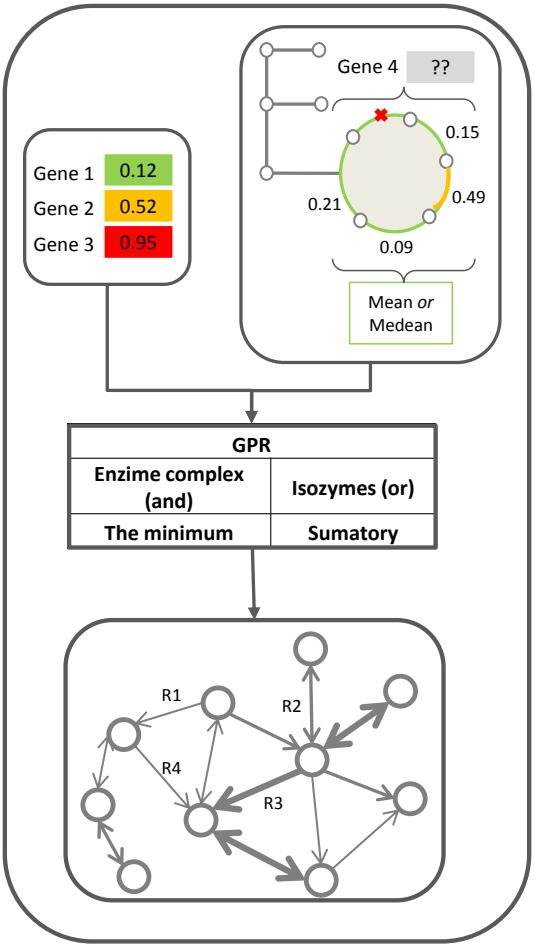
Given that steady-state simulations assume a constant enzymatic rate and does not consider the real expression of each gene or the subcellular localization of gene products, flux constraints based on different omics data (such as transcriptomics, proteomics, and metabolomics) must be integrated into GMNs, in order to recreate specific metabolic phenotypes and enhance model predictive accuracy [63], [144]. Integrative methods have a powerful potential to describe molecular and biochemical mechanisms of organisms' metabolism under specific environmental or genetic conditions, as well as to allow contextualizing high-throughput data [46].

Several algorithms and methods have been developed to integrate experimental data into GMNs [145]. Given the increase of gene expression data, these algorithms have focused on incorporating transcriptomics data into constraint-based models [146], [147] and, in this manner, constrain flux distribution (solution space) in GMNs [71], [148]. Each of the algorithms is based on the assumption that mRNA transcript levels are a strong indicator of the level of protein activity [149]. The algorithms differ mainly in the way to integrate expression data, some of the implemented algorithms incorporate data in a discrete or continuous way, using absolute values for a single condition, or relative expression levels between different conditions [145]. Most methods are based on the discretization of data to incorporate constraints into metabolic models via Boolean logic representation (activation/inactivation flux) [149]. Even when continuous integration could be more accurate for the physiological representation of the continuity of the reactions activity

gradient [63], the methods of continuous integration are very few [145].

With the aim of offering an open source tool that facilitates the integration of gene-expression data as a continuous constraint (FBA limits) in metabolic models, we introduce the "exp2flux" R package. The exp2flux package incorporates a previously described but not implemented continuous gene-expression data integration method [150], [151]. The implemented method is based on the association of omics data with genes included in the Gene-Protein-Reaction (GPR) related to each reaction in a genome-scale metabolic model [152] and considers different possible biological scenarios that occur during the catalysis of biochemical reactions [150][19] (Figure 3-1). Also, a function to calculate the difference in reaction fluxes between simulated scenarios was included.

Figure 3-1: Workflow to integrate omics data into reaction flux of a genome-scale metabolic model.



3.2 Installation and functions

exp2flux includes two functions and is available for download and installation from CRAN, the Comprehensive R Archive Network. To install and load it, just type:

```
> install.packages("exp2flux")
> library(exp2flux)
```

The exp2flux package requires R version 2.10 or higher.

3.2.1 Inputs

Functions included in the **exp2flux** package takes two kind of objects as input, metabolic models as an object of class **modelorg** for the 'sybil' R package, and gene-expression data as an object of class **ExpressionSet**, a container for high-throughput assays and experimental metadata described in the 'Biobase' Bioconductor package.

3.2.2 Converting gene expression data to FBA limits

Gene-protein-reaction (GPR) associations indicate which gene has which function to a genome-scale metabolic network and are represented as Boolean relationships between genes. This function calculates and assigns the flux boundaries for each reaction based on their associated GPR. The value is obtained as follows: (1) When two genes are associated with an AND operator according to the GPR rule, a minimum function is applied to their associated expression values. In the AND case, downregulated genes alter the reaction, acting as the limiting factor in enzyme formation, since both are required for the enzymatic complex formation. In turn, (2) when the genes are associated with an OR rule, each one of them can code an entire enzyme to act as a reaction catalyst. In this case, a sum function is applied for their associated expression values. For missing gene expression values, the function does a data imputation and assigns one of: `min', `1q',`mean', `median', `3q', or `max' expression values calculated from the genes associated to the same metabolic pathway. Metabolic pathway assignation for each reaction is performed through an organism specific search in the KEGG database. In the case of not being able to assign a pathway to a gene, the value to be assigned is calculated from all gene expression values. The fluxes

boundaries of exchange reactions are not modified. To show the potential use of data integration through the exp2flux function, in this example, we simulate values to represent gene expression data and integrate it into the *Escherichia coli* core metabolic model included in the 'sybil' R package.

Integration of gene-expression data begins by loading the 'exp2flux', 'sybil' and 'biobase' required packages.

```
>  library(exp2flux)
>  library(sybil)
>  library(Biobase)
```

After that, the *E. coli* core metabolic model can be loaded from the 'sybil' package. The model includes 95 biochemical reactions associated with 137 genes in 69 GPR rules.

```
> data("Ec_core")
```

Five different measures to represent the gene-expression data for each gene included in the metabolic model was simulated, and the generated matrix was converted to an **ExpressionSet**.

```
> geneExpression <- ExpressionSet(assayData = matrix(
+       data = runif(n = 5*length(Ec_core@allGenes), min = 0, max = 1000),
+       nrow = length(Ec_core@allGenes),
+ dimnames = list(c(Ec_core@allGenes)) + ))

> geneExpression

ExpressionSet (storageMode: lockedEnvironment) assayData:
137 features, 5 samples
    element names: exprs
protocolData: none
phenoData: none
featureData: none
experimentData: use 'experimentData(object)'
Annotation:
```

Incorporation of gene-expression data into the metabolic model through the exp2flux function requires two objects, a metabolic model and an **ExpressionSet** as arguments. The exp2flux function returns a constrained metabolic model.

```
> mEc_core <- exp2flux(
+       model = Ec_core,
+       expression = geneExpression
+       )
> mEc_core

model name:                    Ecoli_core_model
number of compartments  2
                               C_c
                               C_e
number of reactions:           95
number of metabolites:  72
number of unique genes: 137
objective function:            +1 Biomass_Ecoli_core_w_GAM
```

To visualize the metabolic changes induced by the incorporation of gene-expression data as FBA limits of the reactions included in the metabolic model, a Flux Variability Analysis was performed using the original *E. coli* core metabolic model and the constrained one.

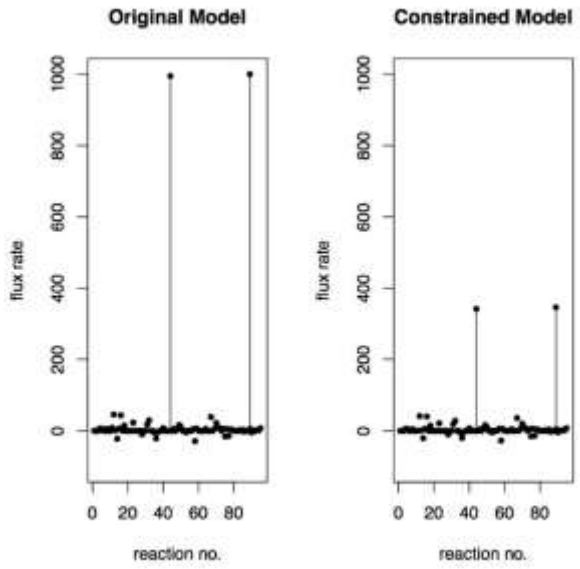
```
>   par(mfcol=c(1,2))
>   plot(fluxVar(Ec_core),main="Original Model",ylim=c(-100,1000))

  |                    :                    |                    :                    |   100 %
                       |=================================================|     :-)

> plot(fluxVar(mEc_core),main="Modified Model",ylim=c(-100,1000))

  |                    :                    |                    :                    |   100 %
                       |=================================================|     :-)
```

Figure 3-2: Flux differences between an unconstrained and a constrained model. Constraints were calculated through the exp2flux R package using simulated gene expression data.



3.2.3 Identifying flux changes between scenarios

The measurement of flux change for each reaction between metabolic scenarios is a task generally carried out manually and oriented directly towards the research objective. However, at a system level analysis, this process can become laborious. The **fluxDifferences** function calculates the fold change for each common reaction between metabolic scenarios. *Fold change* is a measure that describes how much a values changes going from an initial to a final value. The algorithm implemented in the **fluxDifferences** functions is described in equation 3-1; the function takes as an argument two valid models for the 'sybil' R package and a customizable threshold value to filter functions to be reported.

$$foldChange : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

$$(rFluxModel1, rFluxModel2) \mapsto \begin{cases} rFluxModel2, & rFluxModel1 = 0; \\ \dfrac{(rFluxModel2 - rFluxModel1)}{|rFluxModel1|}, & \text{Other cases} \end{cases} \tag{3-1}$$

As an example, we report the fold change of all reactions with an absolute change greater or

equal to 2-fold between the unconstrained and constrained metabolic scenarios previously simulated.

```
> fluxDifferences(
+       model1 = Ec_core,
+       model2 = mEc_core,
+       foldReport = 2
+       )
```

	fluxModel1	fluxModel2	foldChange
ADK1	4.547474e-13	1.000444e-11	21
D_LACt2	-6.821210e-13	1.364242e-12	3
LDH_D	-6.821210e-13	1.364242e-12	3

3.3 Summary

We introduced the exp2flux package, an implementation of a previously described method to integrate gene-expression data in a continuous way into genome-scale metabolic network reconstructions. We show, as an example, how the integration of a set of simulated data modifies the behavior of the *E.coli* core metabolic model using Flux Balance Analysis simulations. Also, an example of the measurement of flux change between metabolic scenarios was shown.

Bibliografía

[1] D. Ribeiro, O. Ribeiro, and D. Erwin, *Phytophthora diseases worldwide*. 1996.

[2] J. B. Ristaino and M. L. Gumpertz, "New frontiers in the study of dispersal and spatial analysis of epidemics caused by species in the genus phytophthora," *Phytopathol*, vol. 38, pp. 541–576, 2000.

[3] H. S. Judelson and F. A. Blanco, "The spores of Phytophthora: weapons of the plant destroyer," *Nat. Rev. Microbiol.*, vol. 3, no. 1, pp. 47–58, Jan. 2005.

[4] A. R. Hardham, "The cell biology behind Phytophthora pathogenicity," *Australas. Plant Pathol.*, vol. 30, no. 2, pp. 91–98, 2001.

[5] A.-M. Catanzariti, P. N. Dodds, and J. G. Ellis, "Avirulence proteins from haustoria-forming pathogens," *FEMS Microbiol. Lett.*, vol. 269, no. 2, pp. 181–188, 2007.

[6] S. Kamoun, "Molecular genetics of pathogenic oomycetes.," *Eukaryot. Cell*, vol. 2, no. 2, pp. 191–9, Apr. 2003.

[7] J. Ellis, A.-M. Catanzariti, and P. Dodds, "The problem of how fungal and oomycete avirulence proteins enter plant cells.," *Trends Plant Sci.*, vol. 11, no. 2, pp. 61–3, 2006.

[8] S. Kamoun, "A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes," *Annu. Rev. Phytopathol.*, vol. 44, no. 1, pp. 41–60, 2006.

[9] B. M. Tyler, "Molecular basis of recognition between phytophthora pathogens and their hosts.," *Annu. Rev. Phytopathol.*, vol. 40, pp. 137–67, Jan. 2002.

[10] J. T. Tippett, A. A. Holland, G. C. Marks, and T. P. O'Brien, "Penetration of Phytophthora cinnamomi into disease tolerant and susceptible eucalypts," *Arch. Microbiol.*, vol. 108, no. 3, pp. 231–242, 1976.

[11] M. a. Bolinder, T. Kätterer, C. Poeplau, G. Börjesson, and L. E. Parent, "Net primary productivity and below-ground crop residue inputs for root crops: Potato (*Solanum tuberosum* L.) and sugar beet (*Beta vulgaris* L.)," *Can. J. Soil Sci.*, vol. 95, no. 2, pp. 87–93, 2015.

[12] B. Flis, E. Zimnoch-Guzowska, and D. Mankowski, "Correlations among Yield, Taste, Tuber Characteristics and Mineral Contents of Potato Cultivars Grown at Different

Growing Conditions," *J. Agric. Sci.*, vol. 4, no. 7, p. p197, Jun. 2012.

[13] J. Li, L. Zhu, G. Lu, X.-B. Zhan, C.-C. Lin, and Z.-Y. Zheng, "Curdlan β-1,3-glucooligosaccharides induce the defense responses against Phytophthora infestans infection of potato (Solanum tuberosum L. cv. McCain G1) leaf cells.," *PLoS One*, vol. 9, no. 5, p. e97197, Jan. 2014.

[14] J. Solano, I. Acuña, F. Esnault, and P. Brabant, "Resistance to Phytophthora infestans in Solanum tuberosum landraces in Southern Chile," *Trop. Plant Pathol.*, vol. 39, no. 4, pp. 307–315, Aug. 2014.

[15] K. Yoshida, V. J. Schuenemann, L. M. Cano, M. Pais, B. Mishra, R. Sharma, C. Lanz, F. N. Martin, S. Kamoun, J. Krause, M. Thines, D. Weigel, and H. a Burbano, "The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine," *Elife*, vol. 2, pp. 1–25, 2013.

[16] B. J. Haas, S. Kamoun, M. C. Zody, R. H. Y. Jiang, R. E. Handsaker, L. M. Cano, M. Grabherr, C. D. Kodira, S. Raffaele, T. Torto-Alalibo, T. O. Bozkurt, A. M. V. Ah-Fong, L. Alvarado, V. L. Anderson, M. R. Armstrong, A. Avrova, L. Baxter, J. Beynon, P. C. Boevink, S. R. Bollmann, J. I. B. Bos, V. Bulone, G. Cai, C. Cakir, J. C. Carrington, M. Chawner, L. Conti, S. Costanzo, R. Ewan, N. Fahlgren, M. a. Fischbach, J. Fugelstad, E. M. Gilroy, S. Gnerre, P. J. Green, L. J. Grenville-Briggs, J. Griffith, N. J. Grünwald, K. Horn, N. R. Horner, C.-H. Hu, E. Huitema, D.-H. Jeong, A. M. E. Jones, J. D. G. Jones, R. W. Jones, E. K. Karlsson, S. G. Kunjeti, K. Lamour, Z. Liu, L. Ma, D. MacLean, M. C. Chibucos, H. McDonald, J. McWalters, H. J. G. Meijer, W. Morgan, P. F. Morris, C. a. Munro, K. O'Neill, M. Ospina-Giraldo, A. Pinzón, L. Pritchard, B. Ramsahoye, Q. Ren, S. Restrepo, S. Roy, A. Sadanandom, A. Savidor, S. Schornack, D. C. Schwartz, U. D. Schumann, B. Schwessinger, L. Seyer, T. Sharpe, C. Silvar, J. Song, D. J. Studholme, S. Sykes, M. Thines, P. J. I. van de Vondervoort, V. Phuntumart, S. Wawra, R. Weide, J. Win, C. Young, S. Zhou, W. Fry, B. C. Meyers, P. van West, J. Ristaino, F. Govers, P. R. J. Birch, S. C. Whisson, H. S. Judelson, and C. Nusbaum, "Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans," *Nature*, vol. 461, no. 7262, pp. 393–398, 2009.

[17] A. Pinzón, E. Barreto, A. Bernal, L. Achenie, A. F. González Barrios, R. Isea, and S. Restrepo, "Computational models in plant-pathogen interactions: the case of Phytophthora infestans," *Theor. Biol. Med. Model.*, vol. 6, no. 1, p. 24, 2009.

[18] L. Gao, Z. Tu, B. P. Millett, and J. M. Bradeen, "Insights into organ-specific pathogen

defense responses in plants: RNA-seq analysis of potato tuber-Phytophthora infestans interactions," *BMC Genomics*, vol. 14, no. 1, p. 340, 2013.

[19] S. Restrepo, K. L. Myers, O. del Pozo, G. B. Martin, a L. Hart, C. R. Buell, W. E. Fry, and C. D. Smart, "Gene profiling of a compatible interaction between Phytophthora infestans and Solanum tuberosum suggests a role for carbonic anhydrase.," *Mol. Plant. Microbe. Interact.*, vol. 18, no. 9, pp. 913–922, 2005.

[20] K. Beyer, A. Binder, T. Boller, and M. Collinge, "Identification of potato genes induced during colonization by Phytophthora infestans," *Mol. Plant Pathol.*, vol. 2, no. 3, pp. 125–134, May 2001.

[21] B. Wang, J. Liu, Z. Tian, B. Song, and C. Xie, "Monitoring the expression patterns of potato genes associated with quantitative resistance to late blight during Phytophthora infestans infection using cDNA microarrays," *Plant Sci.*, vol. 169, no. 6, pp. 1155–1167, 2005.

[22] H. Lindqvist-Kreuze, D. Carbajulca, G. Gonzalez-Escobedo, W. Pérez, and M. Bonierbale, "Comparison of transcript profiles in late blight-challenged Solanum cajamarquense and B3C1 potato clones.," *Mol. Plant Pathol.*, vol. 11, no. 4, pp. 513–30, Jul. 2010.

[23] R. W. Jones, "Multiple Copies of Genes Encoding XEGIPs are Harbored in an 85-kB Region of the Potato Genome," *Plant Mol. Biol. Report.*, vol. 30, no. 4, pp. 1040–1046, 2012.

[24] V. A. Halim, L. Eschen-Lippold, S. Altmann, M. Birschwilks, D. Scheel, and S. Rosahl, "Salicylic acid is important for basal defense of Solanum tuberosum against Phytophthora infestans.," *Mol. Plant. Microbe. Interact.*, vol. 20, no. 11, pp. 1346–52, Nov. 2007.

[25] V. a. Halim, S. Altmann, D. Ellinger, L. Eschen-Lippold, O. Miersch, D. Scheel, and S. Rosahl, "PAMP-induced defense responses in potato require both salicylic acid and jasmonic acid," *Plant J.*, vol. 57, no. 2, pp. 230–242, 2009.

[26] V. G. Vleeshouwers, W. van Dooijeweert, F. Govers, S. Kamoun, and L. T. Colon, "The hypersensitive response is associated with host and nonhost resistance to Phytophthora infestans.," *Planta*, vol. 210, no. 6, pp. 853–64, May 2000.

[27] P. R. J. Birch, A. O. Avrova, J. M. Duncan, G. D. Lyon, and R. L. Toth, "Isolation of Potato Genes That Are Induced During an Early Stage of the Hypersensitive Response to Phytophthora infestans," Feb. 2007.

[28] M. S. Muktar, J. Lübeck, J. Strahwald, and C. Gebhardt, "Selection and validation of potato candidate genes for maturity corrected resistance to Phytophthora infestans based on differential expression combined with SNP association and linkage mapping," *Front.*

Genet., vol. 6, no. September, pp. 1–19, 2015.

[29] G. Gyetvai, M. Sønderkær, U. Göbel, R. Basekow, A. Ballvora, M. Imhoff, B. Kersten, K.-L. Nielsen, and C. Gebhardt, "The Transcriptome of Compatible and Incompatible Interactions of Potato (Solanum tuberosum) with Phytophthora infestans Revealed by DeepSAGE Analysis," *PLoS One*, vol. 7, no. 2, p. e31526, 2012.

[30] Potato Genome Sequencing Consortium, "Genome sequence and analysis of the tuber crop potato," *Nature*, vol. 475, no. 7355, pp. 189–195, 2011.

[31] J. T. Greenberg and N. Yao, "The role and regulation of programmed cell death in plant-pathogen interactions," *Cell. Microbiol.*, vol. 6, no. 3, pp. 201–211, 2004.

[32] C. M. Rojas, M. Senthil-Kumar, V. Tzin, and K. S. Mysore, "Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense," *Front. Plant Sci.*, vol. 5, no. February, pp. 1–12, 2014.

[33] B. Ø. Palsson, *Systems Biology: properties of reconstruted networks*. California: Cambrigde University Press, 2006.

[34] P. Wellstead, E. Bullinger, D. Kalamatianos, O. Mason, and M. Verwoerd, "The rôle of control and system theory in systems biology," *Annu. Rev. Control*, vol. 32, pp. 33–47, 2008.

[35] R. Singh, "Darwin to DNA, molecules to morphology: the end of classical population genetics and the road ahead," *NRC Res. Press*, vol. 46, pp. 938–942, 2003.

[36] P. Waliszewski, M. Molski, and J. Konarski, "On the Holistic Approach in Cellular and Cancer Biology: Nonlinearity, Complexity, and Quasi-Determinism of the Dynamic Cellular Network," *J. Surg. Oncol.*, vol. 68, pp. 70–78, 1998.

[37] H. V Westerhoff and B. O. Palsson, "The evolution of molecular biology into systems biology.," *Nat. Biotechnol.*, vol. 22, no. 10, pp. 1249–52, Oct. 2004.

[38] B. Ø. Palsson, "Metabolic Systems Biology," *FEBS Lett*, vol. 583, no. 24, pp. 3900–3904, 2009.

[39] J. Schellenberger, J. O. Park, T. M. Conrad, and B. Ø. Palsson, "BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions.," *BMC Bioinformatics*, vol. 11, p. 213, 2010.

[40] C. Feuillet, J. E. Leach, J. Rogers, P. S. Schnable, and K. Eversole, "Crop genome sequencing: lessons and rationales.," *Trends Plant Sci.*, vol. 16, no. 2, pp. 77–88, Feb. 2011.

[41] A. Varma and B. O. Palsson, "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use," *Bio/Technology*, vol. 12, pp. 994–998, 1994.

[42] K. Raman and N. Chandra, “Flux balance analysis of biological systems: Applications and challenges,” *Briefings in Bioinformatics*, vol. 10, no. 4. pp. 435–449, 2009.

[43] N. E. Lewis, H. Nagarajan, and B. O. Palsson, “Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.,” *Nat. Rev. Microbiol.*, vol. 10, no. 4, pp. 291–305, Feb. 2012.

[44] N. D. Price, J. A. Papin, C. H. Schilling, and B. O. Palsson, “Genome-scale microbial in silico models: the constraints-based approach.,” *Trends Biotechnol.*, vol. 21, no. 4, pp. 162–9, Apr. 2003.

[45] P. A. Jensen and J. A. Papin, “Functional integration of a metabolic network model and expression data without arbitrary thresholding,” *Bioinformatics*, vol. 27, no. 4, pp. 541–547, Feb. 2011.

[46] M. a Oberhardt, B. Ø. Palsson, and J. a Papin, “Applications of genome-scale metabolic reconstructions.,” *Mol. Syst. Biol.*, vol. 5, no. 320, p. 320, 2009.

[47] K. Radrich, Y. Tsuruoka, P. Dobson, a. Gevorgyan, and N. Swainston, “Reconstruction ofAan In Silico Metabolic Model of Arabidopsis thaliana Through Database Integration,” *Nat. Proc.*, pp. 1–27, 2009.

[48] M. G. Poolman, L. Miguet, L. J. Sweetlove, and D. a Fell, “A genome-scale metabolic model of Arabidopsis and some of its properties.,” *Plant Physiol.*, vol. 151, no. 3, pp. 1570–1581, 2009.

[49] E. Grafahrend-Belau, F. Schreiber, D. Koschützki, and B. H. Junker, “Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism.,” *Plant Physiol.*, vol. 149, no. 1, pp. 585–598, 2009.

[50] NCBI Resource Coordinators, “Índice de /genomes/genbank/plant/,” 2016. [Online]. Available: ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/plant/. [Accessed: 05-Jan-2016].

[51] NCBI Resource Coordinators, “Database resources of the National Center for Biotechnology Information.,” *Nucleic Acids Res.*, vol. 45, pp. 6–17, Nov. 2015.

[52] C. G. D. O. Dal’Molin, L.-E. Quek, R. W. Palfreyman, S. M. Brumbley, and L. K. Nielsen, “C4GEM, a genome-scale metabolic model to study C4 plant metabolism.,” *Plant Physiol.*, vol. 154, no. 4, pp. 1871–1885, 2010.

[53] R. Saha, P. F. Suthers, and C. D. Maranas, “Zea mays irs1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism,” *PLoS One*, vol. 6, no. 7, 2011.

[54] M. Simons, R. Saha, N. Amiour, a. Kumar, L. Guillard, G. Clement, M. Miquel, Z. Li, G. Mouille, P. J. Lea, B. Hirel, and C. D. Maranas, “Assessing the Metabolic Impact of Nitrogen Availability Using a Compartmentalized Maize Leaf Genome-Scale Model,”

Plant Physiol., vol. 166, no. 3, pp. 1659–1674, 2014.

[55] E. Pilalis, A. Chatziioannou, B. Thomasset, and F. Kolisis, “An in silico compartmentalized metabolic model of Brassica napus enables the systemic study of regulatory aspects of plant central metabolism,” *Biotechnol. Bioeng.*, vol. 108, no. 7, pp. 1673–1682, 2011.

[56] J. Hay and J. Schwender, “Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (Brassica napus L.) embryos,” *Plant J.*, vol. 67, no. 3, pp. 526–541, 2011.

[57] M. G. Poolman, S. Kundu, R. Shaw, and D. a Fell, “Responses to light intensity in a genome-scale model of rice metabolism.,” *Plant Physiol.*, vol. 162, no. 2, pp. 1060–72, 2013.

[58] H. Yuan, C. Y. M. Cheung, M. G. Poolman, P. A. J. Hilbers, and N. A. W. van Riel, “A genome-scale metabolic network reconstruction of tomato (Solanum lycopersicum L.) and its application to photorespiratory metabolism.,” *Plant J.*, vol. 85, no. 2, pp. 289–304, Jan. 2016.

[59] N. Chen, I. J. Del Val, S. Kyriakopoulos, K. M. Polizzi, and C. Kontoravdi, “Metabolic network reconstruction: Advances in in silico interpretation of analytical information,” *Curr. Opin. Biotechnol.*, vol. 23, no. 1, pp. 77–82, 2012.

[60] C. Colijn, A. Brandes, J. Zucker, D. S. Lun, B. Weiner, M. R. Farhat, T. Y. Cheng, D. B. Moody, M. Murray, and J. E. Galagan, “Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production,” *PLoS Comput. Biol.*, vol. 5, no. 8, 2009.

[61] C. Caldana, T. Degenkolbe, A. Cuadros-Inostroza, S. Klie, R. Sulpice, A. Leisse, D. Steinhauser, A. R. Fernie, L. Willmitzer, and M. a. Hannah, “High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions,” *Plant J.*, vol. 67, no. 5, pp. 869–884, 2011.

[62] S. Mintz-Oron, S. Meir, S. Malitsky, E. Ruppin, a. Aharoni, and T. Shlomi, “Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 1, pp. 339–344, 2012.

[63] N. Töpfer, C. Caldana, S. Grimbs, L. Willmitzer, A. R. Fernie, and Z. Nikoloski, “Integration of genome-scale modeling and transcript profiling reveals metabolic pathways underlying light and temperature acclimation in Arabidopsis.,” *Plant Cell*, vol. 25, no. 4, pp. 1197–211, 2013.

[64] J. D. G. Jones and J. L. Dangl, “The plant immune system,” *Nature*, vol. 444, no. 7117, pp. 323–329, Nov. 2006.

[65] S. a Becker and B. Ø. Palsson, “Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation.,” *BMC Microbiol.*, vol. 5, p. 8, 2005.

[66] M. MacLeod and N. J. Nersessian, “Modeling systems-level dynamics: Understanding without mechanistic explanation in integrative systems biology,” *Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci.*, vol. 49, pp. 1–11, 2015.

[67] B. D. Heavner and N. D. Price, “Transparency in metabolic network reconstruction enables scalable biological discovery,” *Curr. Opin. Biotechnol.*, vol. 34, pp. 105–109, 2015.

[68] C. G. de Oliveira Dal’Molin, L.-E. Quek, R. W. Palfreyman, S. M. Brumbley, and L. K. Nielsen, “AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis,” *Plant Physiol.*, vol. 152, no. 2, pp. 579–589, 2010.

[69] I. Thiele and B. Ø. Palsson, “A protocol for generating a high-quality genome-scale metabolic reconstruction.,” *Nat. Protoc.*, vol. 5, no. 1, pp. 93–121, 2010.

[70] S. M. D. Seaver, C. S. Henry, and A. D. Hanson, “Frontiers in metabolic reconstruction and modeling of plant genomes,” *J. Exp. Bot.*, vol. 63, no. 6, pp. 2247–2258, 2012.

[71] J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Ø. Palsson, “Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.,” *Nat. Protoc.*, vol. 6, no. 9, pp. 1290–1307, 2011.

[72] M. D. Coffey and U. E. Wilson, “An ultrastructural study of the late-blight fungus Phytophthora infestans and its interaction with the foliage of two potato cultivars possessing different levels of general (field) resistance,” *Can. J. Bot*, vol. 61, pp. 2669–2685, 1983.

[73] R. Pristou and M. E. Gallegly, “Leaf penetration by Phytophthora infestans,” *Phytopathology*, vol. 44, pp. 81–86, 1954.

[74] C. Shimony and J. Friend, “Ultrastructure of the interaction between Phytophthora infestans and leaves of two cultivars of potato (Solanum tuberosum L.) orion and majestic,” *New Phytol.*, vol. 74, no. 1, pp. 59–65, Jan. 1975.

[75] H. R. Hohl and P. Stössel, “Host–parasite interfaces in a resistant and a susceptible cultivar of *Solanum tuberosum* inoculated with *Phytophthora infestans* : tuber tissue,” *Can. J. Bot.*, vol. 54, no. 9, pp. 900–912, May 1976.

[76] G. N. Agrios, “Effects of pathogens on plant physiological functions,” in *Plant Pathology*,

5th ed., California: Elsevier, 2004, pp. 105–123.

[77] D. I. Arnon, M. B. Allen, and F. R. Whatley, “Photosynthesis by Isolated Chloroplasts,” *Nature*, vol. 174, no. 4426, pp. 394–396, Aug. 1954.

[78] F. R. Whatley, M. B. Allen, A. V Trebst, and D. I. Arnon, “Photosynthesis by solated chloroplasts IX. Photosynthetic phosphorylation and CO(2) assimilation in different species.,” *Plant Physiol.*, vol. 35, no. 2, pp. 188–93, Mar. 1960.

[79] M. Hervás, J. A. Navarro, and M. A. De La Rosa, “Electron Transfer between Membrane Complexes and Soluble Proteins in Photosynthesis,” *Acc. Chem. Res.*, vol. 36, no. 10, pp. 798–805, 2003.

[80] J. F. Allen, “Photosynthesis of ATP-electrons, proton pumps, rotors, and poise,” *Cell*, vol. 110, no. 3, pp. 273–276, 2002.

[81] D. I. Arnon, “The chloroplast as a functional unit in photosynthesis,” in *Handbuch der Pflanzenphysiologie*, W. Ruhland, E. Ashby, J. Bonner, M. Geiger-Huber, W. O. James, A. Lang, D. Müller, and M. G. Stålfelt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1960, pp. 773–829.

[82] J. . Berg, J. . Tymoczko, and L. Stryer, “The Light Reactions of Photosynthesis,” in *Biochemistry. 5th edition*, New York: W.H. Freeman, 2002, p. 1050.

[83] A. A. Stephanopoulos, Gregory N. Aristidou and J. Nielsen, *Metabolic engineering. Principles and methodologies*. San Diego, CA: Academic Press, 1998.

[84] C. Koch, G. Noga, and G. Strittmatter, “Photosynthetic electron transport is differentially affected during early stages of cultivar/race-specific interactions between potato andPhytophthora infestans,” *Planta*, vol. 193, no. 4, pp. 551–557, May 1994.

[85] J. Yen, I. Tanniche, A. K. Fisher, G. E. Gillaspy, D. R. Bevan, and R. S. Senger, “Designing metabolic engineering strategies with genome-scale metabolic flux modeling,” *Adv. Genomics Genet.*, vol. 5, pp. 93–105, 2015.

[86] G. Karp, *Cell and molecular biology. Concepts and experiments*, 6th Editio. 2009.

[87] A. Wingler, P. J. Lea, W. P. Quick, and R. C. Leegood, “Photorespiration: metabolic pathways and their role in stress protection.,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 355, no. 1402, pp. 1517–29, Oct. 2000.

[88] N. A. Eckardt, “Photorespiration Revisited,” *PLANT CELL ONLINE*, vol. 17, no. 8, pp. 2139–2141, Aug. 2005.

[89] G. Noctor, A.-C. M. Arisi, L. Jouanin, and C. H. Foyer, “Photorespiratory glycine enhances glutathione accumulation in both the chloroplastic and cytosolic compartments,” *J. Exp.*

Bot., vol. 50, no. 336, pp. 1157–1167, Jul. 1999.

[90] C. Peterhansel, I. Horst, M. Niessen, C. Blume, R. Kebeish, S. Kürkcüoglu, and F. Kreuzaler, "Photorespiration.," *Arab. B.*, vol. 8, p. e0130, 2010.

[91] I. Slesak, M. Libik, B. Karpinska, S. Karpinski, and Z. Miszalski, "The role of hydrogen peroxide in regulation of plant metabolism and cellular signalling in response to environmental stresses," *Acta Biochim Pol*, vol. 54, no. 1, pp. 39–50, 2007.

[92] J. L. Dangl and J. D. G. Jones, "Plant pathogens and integrated defence responses to infection," *Nature*, vol. 411, no. 6839, pp. 826–833, Jun. 2001.

[93] I. Apostol, P. F. Heinstein, and P. S. Low, "Rapid Stimulation of an Oxidative Burst during Elicitation of Cultured Plant Cells : Role in Defense and Signal Transduction.," *Plant Physiol.*, vol. 90, no. 1, pp. 109–16, May 1989.

[94] N. Doke, "NADPH-dependent O2− generation in membrane fractions isolated from wounded potato tubers inoculated with Phytophthora infestans," *Physiol. Plant Pathol.*, vol. 27, no. 3, pp. 311–322, 1985.

[95] C. H. Foyer and G. Noctor, "Oxygen Processing in Photosynthesis: Regulation and Signalling," *New Phytol.*, vol. 146, pp. 359–388, 2000.

[96] D. F. Klessig, J. Durner, R. Noad, D. A. Navarre, D. Wendehenne, D. Kumar, J. M. Zhou, J. Shah, S. Zhang, P. Kachroo, Y. Trifa, D. Pontier, E. Lam, and H. Silva, "Nitric oxide and salicylic acid signaling in plant defense.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 16, pp. 8849–55, Aug. 2000.

[97] R. Mittler, E. H. Herr, B. L. Orvar, W. van Camp, H. Willekens, D. Inzé, and B. E. Ellis, "Transgenic tobacco plants with reduced capability to detoxify reactive oxygen intermediates are hyperresponsive to pathogen infection.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 24, pp. 14165–70, Nov. 1999.

[98] M. Delledonne, J. Zeier, A. Marocco, and C. Lamb, "Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 23, pp. 13454–9, Nov. 2001.

[99] K. Apel and H. Hirt, "Reactive oxygen species: Metabolism, Oxidative Stress, and Signal Transduction," *Annu. Rev. Plant Biol.*, vol. 55, no. 1, pp. 373–399, 2004.

[100] C. J. Baker and E. W. Orlandi, "Active Oxygen in Plant Pathogenesis," *Annu. Rev. Phytopathol.*, vol. 33, no. 1, pp. 299–321, Sep. 1995.

[101] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular Cell Biology*, 4th editio. New York: W. H. Freeman, 2000.

[102] R. Sulpice, E.-T. Pyl, H. Ishihara, S. Trenkamp, M. Steinfath, H. Witucka-Wall, Y. Gibon,

B. Usadel, F. Poree, M. C. Piques, M. Von Korff, M. C. Steinhauser, J. J. B. Keurentjes, M. Guenther, M. Hoehne, J. Selbig, A. R. Fernie, T. Altmann, and M. Stitt, "Starch as a major integrator in the regulation of plant growth.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 25, pp. 10348–53, Jun. 2009.

[103] P. Geigenberger, "Regulation of starch biosynthesis in response to a fluctuating environment.," *Plant Physiol.*, vol. 155, no. 4, pp. 1566–1577, 2011.

[104] Y. Obana, D. Omoto, C. Kato, K. Matsumoto, Y. Nagai, I. H. Kavakli, S. Hamada, G. E. Edwards, T. W. Okita, H. Matsui, and H. Ito, "Enhanced turnover of transitory starch by expression of up-regulated ADP-glucose pyrophosphorylases in Arabidopsis thaliana," *Plant Sci.*, vol. 170, no. 1, pp. 1–11, 2006.

[105] J. H. M. Hendriks, A. Kolbe, Y. Gibon, M. Stitt, and P. Geigenberger, "ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species.," *Plant Physiol.*, vol. 133, no. 2, pp. 838–49, Oct. 2003.

[106] M. Stitt, S. Huber, and P. Kerr, "Control of photosynthetic sucrose synthesis. In MD Hatch, NK Boardman, eds, The Biochemistry of Plants," in *The Biochemistry of Plants, The Biochemistry of Plants, Vol 10*, M. Hatch and N. Boardman, Eds. New York: Academic Press, 1987, pp. 327–409.

[107] I. J. Tetlow, K. G. Beisel, S. Cameron, A. Makhmoudova, F. Liu, N. S. Bresolin, R. Wait, M. K. Morell, and M. J. Emes, "Analysis of Protein Complexes in Wheat Amyloplasts Reveals Functional Interactions among Starch Biosynthetic Enzymes," *PLANT Physiol.*, vol. 146, no. 4, pp. 1878–1891, Apr. 2008.

[108] P. Geigenberger and M. Stitt, "Diurnal changes in sucrose, nucleotides, starch synthesis and AGPS transcript in growing potato tubers that are suppressed by decreased expression of sucrose phosphate synthase.," *Plant J.*, vol. 23, no. 6, pp. 795–806, Sep. 2000.

[109] Y. Gibon, O. E. Blaesing, J. Hannemann, P. Carillo, M. Höhne, J. H. M. Hendriks, N. Palacios, J. Cross, J. Selbig, and M. Stitt, "A Robot-Based Platform to Measure Multiple Enzyme Activities in Arabidopsis Using a Set of Cycling Assays: Comparison of Changes of Enzyme Activities and Transcript Levels during Diurnal Cycles and in Prolonged Darkness," *PLANT CELL ONLINE*, vol. 16, no. 12, pp. 3304–3325, Dec. 2004.

[110] S. M. Smith, D. C. Fulton, T. Chia, D. Thorneycroft, A. Chapple, H. Dunstan, C. Hylton, S. C. Zeeman, and A. M. Smith, "Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional

Regulation of Starch Metabolism in Arabidopsis Leaves," *PLANT Physiol.*, vol. 136, no. 1, pp. 2687–2699, Sep. 2004.

[111] D. W. Lawlor, W. Tezara, V. J. Mitchell, and S. D. Driscoll, "Water stress inhibits plant photosynthesis by decreasing coupling factor and ATP," *Nature*, vol. 401, no. 6756, pp. 914–917, Oct. 1999.

[112] A. J. Lack and David E. Evans, *Plant Biology*, 2th editio. New York: Taylor & Francis group, 2005.

[113] G. D. Farquhar and S. von Caemmerer, "Modelling of Photosynthetic Response to Environmental Conditions," in *Physiological Plant Ecology II*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 549–587.

[114] R. F. Sage, "A Model Describing the Regulation of Ribulose-1,5-Bisphosphate Carboxylase, Electron Transport, and Triose Phosphate Use in Response to Light Intensity and CO(2) in C(3) Plants.," *Plant Physiol.*, vol. 94, no. 4, pp. 1728–34, 1990.

[115] N. Swainston, K. Smallbone, P. Mendes, D. Kell, and N. Paton, "The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks.," *J. Integr. Bioinform.*, vol. 8, no. 2, p. 186, Jan. 2011.

[116] R. Agren, L. Liu, S. Shoaie, W. Vongsangnak, I. Nookaew, and J. Nielsen, "The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum," *PLoS Comput. Biol.*, vol. 9, no. 3, 2013.

[117] M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks," *Biophys. J.*, vol. 95, no. 3, pp. 1487–1499, 2008.

[118] D. Osorio, J. Gonzalez, and A. Pinzon-Velasco, "minval: MINimal VALidation for Stoichiometric Reactions," 2016.

[119] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, "From genomics to chemical genomics: new developments in KEGG.," *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. D354-7, Jan. 2006.

[120] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D109–D114, 2012.

[121] D. Tenenbaum, "KEGGREST: Client-side REST access to KEGG. .," *R Packag. version*, vol. 1, no. 1, 2013.

[122] K. Botero, D. Osorio, J. Gonzalez, and A. Pinzon, "g2f: Find and Fill Gaps in Metabolic

Networks," 2016.

[123] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo, "eQuilibrator--the biochemical thermodynamics calculator.," *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D770-5, Jan. 2012.

[124] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, and P. D. Karp, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D471–D480, 2014.

[125] G. Gelius-Dietrich, "sybil - Efficient Constrained Based Modelling in R," *BMC Syst. Biol.*, vol. 7, no. 1, p. 42, 2013.

[126] C. G. de Oliveira Dal'Molin, L.-E. Quek, R. W. Palfreyman, S. M. Brumbley, and L. K. Nielsen, "AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis.," *Plant Physiol.*, vol. 152, no. 2, pp. 579–589, 2010.

[127] H. Poorter and M. Bergkotte, "Chemical composition of 24 wild species differing in relative growth rate," *Plant, Cell Environ.*, vol. 15, no. 2, pp. 221–229, Feb. 1992.

[128] G. J. Niemann, J. B. M. Pureveen, G. B. Eijkel, H. Poorter, and J. J. Boon, "Differential chemical allocation and plant adaptation: A Py-MS Study of 24 species differing in relative growth rate," *Plant Soil*, vol. 175, no. 2, pp. 275–289, Aug. 1995.

[129] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool.," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–10, Oct. 1990.

[130] M. Carlson, "R Interface to UniProt Web Services," 2014.

[131] R. Gentleman, V. Carey, M. Morgan, and S. Falcon, "Biobase: Base functions for Bioconductor," 2016.

[132] D. Osorio, K. Botero, J. Gonzalez, and A. Pinzon, "exp2flux: Convert Gene EXPression Data to FBA FLUXes," 2016.

[133] N. Tomar and R. K. De, "Comparing methods for metabolic network analysis and an application to metabolic engineering," *Gene*, vol. 521, no. 1. Elsevier B.V., pp. 1–14, 2013.

[134] B. Szappanos, K. Kovács, B. Szamecz, F. Honti, M. Costanzo, A. Baryshnikova, G. Gelius-Dietrich, M. J. Lercher, M. Jelasity, C. L. Myers, B. J. Andrews, C. Boone, S. G. Oliver, C. Pál, and B. Papp, "An integrated approach to characterize genetic interaction networks in yeast metabolism.," *Nat. Genet.*, vol. 43, no. 7, pp. 656–62, 2011.

[135] H. Alper, Y.-S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the

metabolic engineering of lycopene biosynthesis in Escherichia coli," *Metab. Eng.*, vol. 7, no. 3, pp. 155–164, May 2005.

[136] S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. O. Palsson, "In silico design and adaptive evolution of Escherichia coli for production of lactic acid.," *Biotechnol. Bioeng.*, vol. 91, no. 5, pp. 643–8, Sep. 2005.

[137] I. Thiele, N. Vlassis, and R. M. T. Fleming, "FASTGAPFILL : Efficient gap filling in metabolic networks.," *Bioinformatics*, pp. 15–17, 2014.

[138] V. Satish Kumar, M. S. Dasika, and C. D. Maranas, "Optimization based automated curation of metabolic reconstructions," *BMC Bioinformatics*, vol. 8, no. 1, p. 212, 2007.

[139] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D. P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White, and S. Yon Rhee, "Big data: The future of biocuration," *Nature*, vol. 455, no. 7209, pp. 47–50, Sep. 2008.

[140] A. Bateman, "Curators of the world unite: the International Society of Biocuration," *Bioinformatics*, vol. 26, no. 8, pp. 991–991, Apr. 2010.

[141] M. Lakshmanan, G. Koh, B. K. S. Chung, and D.-Y. Lee, "Software applications for flux balance analysis.," *Brief. Bioinform.*, vol. 15, no. 1, pp. 108–22, 2014.

[142] N. D. Price, J. L. Reed, and B. Ø. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints.," *Nat. Rev. Microbiol.*, vol. 2, no. November, pp. 886–897, 2004.

[143] L. Liu, R. Agren, S. Bordel, and J. Nielsen, "Use of genome-scale metabolic models for understanding microbial physiology," *FEBS Lett.*, vol. 584, no. 12, pp. 2556–2564, Jun. 2010.

[144] B. Palsson, "In silico biology through 'omics,'" *Nat. Biotechnol.*, vol. 20, no. 7, pp. 649–650, Jul. 2002.

[145] D. Machado and M. Herrgård, "Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism," *PLoS Comput. Biol.*, vol. 10, no. 4, 2014.

[146] M. W. Covert and B. Ø. Palsson, "Transcriptional Regulation in Constraints-based Metabolic Models of Escherichia coli," *J. Biol. Chem.*, vol. 277, no. 31, pp. 28058–28064, 2002.

[147] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson, "Integrating high-throughput and computational data elucidates bacterial networks," *Nature*, vol. 429, no. 6987, pp. 92–96, May 2004.

[148] J. L. Reed, H. Kim, S. Kim, H. Jeong, T. Kim, J. Kim, J. Park, T. Kim, S. Lee, C. Pal, B.

Papp, M. Lercher, P. Csermely, S. Oliver, V. Kumar, C. Maranas, J. Reed, T. Patel, K. Chen, A. Joyce, M. Applebee, V. S. Kumar, M. Dasika, C. Maranas, A. Bordbar, N. Lewis, J. Schellenberger, B. Palsson, N. Jamshidi, S. Stolyar, S. Van Dien, K. Hillesland, N. Pinel, T. Lie, A. Zomorrodi, C. Maranas, K. Zhuang, M. Izallalen, P. Mouser, H. Richter, C. Risso, I. Thiele, B. Palsson, C. Henry, M. DeJongh, A. Best, P. Frybarger, B. Linsay, M. Kanehisa, S. Goto, P. Karp, S. Paley, M. Krummenacker, M. Latendresse, J. Dale, N. Lewis, H. Nagarajan, B. Palsson, J. Orth, I. Thiele, B. Palsson, J. Schellenberger, R. Que, R. Fleming, I. Thiele, J. Orth, P. Jensen, J. Papin, R. Fleming, I. Thiele, H. Zur, E. Ruppin, T. Shlomi, A. Kummel, S. Panke, M. Heinemann, R. Fleming, I. Thiele, H. Nasheuer, C. Henry, L. Broadbelt, V. Hatzimanikatis, Q. Beg, A. Vazquez, J. Ernst, M. de Menezes, Z. Bar-Joseph, A. Vazquez, Q. Beg, M. Demenezes, J. Ernst, Z. Bar-Joseph, A. Vazquez, M. de Menezes, A. Barabasi, Z. Oltvai, T. Shlomi, T. Benyamini, E. Gottlieb, R. Sharan, E. Ruppin, K. Zhuang, G. Vemuri, R. Mahadevan, G. Pinchuk, E. Hill, O. Geydebrekht, J. De Ingeniis, X. Zhang, A. Raghunathan, J. Reed, S. Shin, B. Palsson, S. Daefler, N. Lewis, K. Hixson, T. Conrad, J. Lerman, P. Charusanti, C. Colijn, A. Brandes, J. Zucker, D. Lun, B. Weiner, S. Becker, B. Palsson, T. Shlomi, M. Cabili, M. Herrgard, B. Palsson, E. Ruppin, J. Moxley, M. Jewett, M. Antoniewicz, S. Villas-Boas, H. Alper, R. De Smet, K. Marchal, D. Rodionov, M. Herrgard, M. Covert, B. Palsson, M. Covert, C. Schilling, B. Palsson, M. Covert, E. Knight, J. Reed, M. Herrgard, B. Palsson, M. Covert, B. Palsson, T. Shlomi, Y. Eisenberg, R. Sharan, E. Ruppin, S. Chandrasekaran, N. Price, X. Feng, Y. Xu, Y. Chen, Y. Tang, T. Hanly, M. Henson, I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, N. Jamshidi, B. Palsson, R. Fleming, I. Thiele, G. Provan, H. Nasheuer, K. Smallbone, E. Simeonidis, N. Swainston, P. Mendes, K. Soh, L. Miskovic, V. Hatzimanikatis, K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, T. Shlomi, R. Taffs, J. Aston, K. Brileya, Z. Jay, and C. Klatt, “Shrinking the Metabolic Solution Space Using Experimental Datasets,” *PLoS Comput. Biol.*, vol. 8, no. 8, p. e1002662, Aug. 2012.

[149] A. S. Blazier and J. A. Papin, “Integration of expression data in genome-scale metabolic network reconstructions.,” *Front. Physiol.*, vol. 3, p. 299, 2012.

[150] C. Colijn, A. Brandes, J. Zucker, D. S. Lun, B. Weiner, M. R. Farhat, T.-Y. Cheng, D. B. Moody, M. Murray, and J. E. Galagan, “Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production,” *PLoS Comput. Biol.*, vol. 5, no. 8, p. e1000489, Aug. 2009.

[151] P. Carbonell, “Hepatotoxicity prediction by systems biology modeling of disturbed metabolic pathways using gene expression data,” *ALTEX*, pp. 1–23, 2016.

[152] I. Thiele and B. Ø. Palsson, “A protocol for generating a high-quality genome-scale metabolic reconstruction.,” *Nat. Protoc.*, vol. 5, no. 1, pp. 93–121, 2010.