UADY

POSGRADO INSTITUCIONAL EN CIENCIAS QUÍMICAS Y BIOQUÍMICAS

RECONSTRUCCIÓN Y VALIDACIÓN DE MODELOS METABÓLICOS DE MICROORGANISMOS PARTICIPANTES EN UN CONSORCIO NITRIFICANTE

TESIS

PRESENTADA POR

DIEGO AARÓN TEC CAMPOS

EN OPCIÓN AL GRADO DE

MAESTRO EN CIENCIAS QUÍMICAS Y BIOQUÍMICAS

MÉRIDA, YUCATÁN, MÉXICO

2020

UADY

POSGRADO INSTITUCIONAL EN CIENCIAS QUÍMICAS Y BIOQUÍMICAS

RECONSTRUCCIÓN Y VALIDACIÓN DE MODELOS METABÓLICOS DE MICROORGANISMOS PARTICIPANTES EN UN CONSORCIO MICROBIANO NITRIFICANTE

TESIS

PRESENTADA POR

DIEGO AARÓN TEC CAMPOS

EN OPCIÓN AL GRADO DE

MAESTRO EN CIENCIAS QUÍMICAS Y BIOQUÍMICAS

MÉRIDA, YUCATÁN, MÉXICO

2020

UADY
UNIVERSIDAD AUTÓNOMA DE YUCATÁN

COORDINACIÓN GENERAL
DEL SISTEMA DE POSGRADO
INVESTIGACIÓN Y VINCULACIÓN

MAESTRÍA EN CIENCIAS
QUÍMICAS Y BIOQUÍMICAS

Mérida, Yuc., 28 de agosto de 2020
Oficio Num.: PICQB/182/2020

Asunto: Autorización de digitalización

La tesis "Reconstrucción y validación de modelos metabólicos de microorganismos participantes en un consorcio microbiano nitrificante" presentada por el I.B. Diego Aaron Tec Campos, en cumplimiento parcial de los requisitos para optar por el grado de Maestro en Ciencias Químicas y Bioquímicas, ha sido aprobada en cuanto a su contenido científico y en cuanto a lo establecido en el Manual de Procedimientos del Posgrado Institucional en Ciencias Químicas y Bioquímicas, por lo que se le autoriza la digitalización de los ejemplares correspondientes.

Directores de la tesis

Dr. Alejandro Zepeda Pedreguera

Dra. Nancy Cristal Zuñiga Peña

Sinodales	Firmas
Dr. Geovanny Nic Can	
Dr. Juan E. Ruíz Espinoza	
Dra. Neyi Estrella Gómez	
Dr. José Utrilla Carreri	

ATENTAMENTE
"LUZ, CIENCIA Y VERDAD"

DRA. MAIRA RUBI SEGURA CAMPOS
COORDINADORA DEL POSGRADO INSTITUCIONAL
EN CIENCIAS QUÍMICAS Y BIOQUÍMICAS

POSGRADO INSTITUCIONAL
EN CIENCIAS QUIMICAS
Y BIOQUIMICAS

c.c. Archivo
MRSC

AGRADECIMIENTOS

Quiero agradecer principalmente a mi familia, a cada uno de ellos, mi madre Aida, mi padre Jesús, a mis hermanos Jesús y Patricia, por apoyarme desde el proceso de ingreso a la maestría en la Facultad de Ingeniería Química, hasta la recta final de toda esta etapa. Siempre han sido el principal pilar de mi vida en todos los aspectos, estando presentes en cada uno de los momentos más importantes de mi formación dentro de esta gran universidad.

Además, quiero agradecer enormemente a mis dos directores de tesis; a la Dra. Nancy Cristal Zúñiga Peña por motivarme a realizar este trabajo y guiarme en cada uno de los detalles para su elaboración; a través de sus enseñanzas durante la estancia de investigación que realicé en la Universidad de California San Diego obtuve muchos conocimientos que me ayudaron a desarrollar este trabajo, siempre admiraré su gran labor como supervisora, investigadora y sobre todo como excelente persona. De igual forma quiero agradecer al Dr. Alejandro Zepeda, por ser mi mentor prácticamente durante toda mi formación universitaria, y ahora, por ser director y guía durante toda mi maestría. He tenido el placer de realizar diferentes trabajos de investigación bajo su tutela, y muchos de mis actuales éxitos no se hubiesen alcanzado sin él, su apoyo incondicional me ha permitido llegar hasta la recta final del proceso.

A cada una de las personas que forman parte del personal de la Facultad de Ingeniería Química, por hacer sentir que esta institución es mi segunda casa, por facilitar cada uno de los procesos y mantener siempre la cordialidad y el espíritu de servicio, además de brindar una excelente educación de calidad.

A la Universidad de California San Diego, por facilitar todo el proceso de estancia de maestría. Específicamente al laboratorio del Dr. Karsten Zengler, quien confió en mí desde el primer día y me dieron capacitación y motivación para realizar un gran trabajo. A los miembros del laboratorio de Zengler, a Juan, Max, Lisa, Livia, Manish, Anurag, Oriane, Bo Liang, Megan, Amber, Joana, Katrina y en especial al Dr. Karsten, su laboratorio siempre tuvo las puertas abiertas para mí.

A mis amigas que me han apoyado, motivado y brindado su ayuda y amistad en todo momento, especialmente quiero mencionar a Regina, Ana, Brigitte, Niní, Salette, Cristina, así como a mi grupo de trabajo de investigación en FIQ.

Por último, quiero agradecer a mis revisores de tesis, al Dr. Juan Enrique Ruiz, a la Dra. Neyi Estrella y al Dr. Geovanny Nic por apoyarme y asesorarme dentro de este trabajo, tuve la fortuna de tenerlos como profesores y estoy enormemente agradecido de todos los conocimientos transmitidos. Gracias a ellos, la calidad de mi trabajo tuvo mejoras significativas.

Este trabajo fue realizado en la Facultad de Ingeniería Química de la Universidad Autónoma de Yucatán y el Departamento de Pediatría de la Universidad de California San Diego, bajo la dirección del Dr. Alejandro Zepeda Pedreguera y la Dra. Nancy Cristal Zúñiga Peña y formó parte del proyecto “Generación de modelos metabólicos basados en el genoma de un consorcio microbiano utilizado en el tratamiento de aguas residuales” financiado por CONACYT (proyecto #284140), con el apoyo de una beca concedida por CONACYT (beca 932962).

RESUMEN

El uso de consorcios microbianos en el tratamiento de aguas residuales se ha posicionado como una alternativa viable entre los diferentes procesos existentes en la actualidad. Sin embargo, esta opción aún presenta desafíos debido a la complejidad existente en las interacciones microbianas, obteniendo diferentes resultados a los esperados, y, por tanto, un sistema con baja confiabilidad. Actualmente existen herramientas que permiten diseñar modelos y realizar inferencias del metabolismo de los miembros de comunidades microbianas usando enfoques basados en biología de sistemas y biología cuantitativa.

En el presente trabajo se analizó el metagenoma de un consorcio nitrificante para el tratamiento de aguas residuales. Seguidamente, se diseñaron modelos metabólicos para los principales miembros del consorcio utilizando estrategias semiautomáticas. Los modelos fueron refinados para aumentar la calidad de las asociaciones genéticas y la robustez en las predicciones. Seis modelos fueron reconstruidos por homología usando como templados modelos de microorganismos previamente validados de la base de datos BiGG. Los templados fueron escogidos basados en comparaciones por ARN 16S, similitud proteómica y fisiológica. Los modelos fueron desarrollados combinando algoritmos de módulos COBRA, RAVEN y algoritmos diseñados en este trabajo. Los parámetros óptimos para la reconstrucción semiautomática fueron calculados en el presente trabajo. Los modelos resultantes fueron optimizados a nivel genético y bioquímico a partir de revisión manual y de algoritmos semiautomáticos. Los modelos de seis microorganismos (*Rhodopseudomonas palustris* BisA53, *Thauera* sp. MZ1T, *Acinetobacter oleivorans* DR1, *Aromatoleum aromaticum* EbN1, *Dechloromonas aromatica* RCB y *Nitrosomonas europaea* ATCC 19718) pertenecientes al consorcio permitieron comprender mejor los procesos asociados con la fijación de carbono-nitrógeno, nitrificación-desnitrificación, y procesos fundamentales en el tratamiento de aguas residuales. Por último, las herramientas semiautomáticas fueron aplicadas en dos proyectos diferentes: el modelo de *Azotobacter vinelandii* (iDT1278) y los modelos de las cepas de *Liberibacter*. El modelo iDT1278 fue validado con datos de crecimiento en más de 300 fuentes de carbono y nitrógeno, obteniendo una precisión mayor al 90% de predicción. Con respecto a los modelos de *Liberibacter*, se evaluó la precisión en las predicciones de crecimiento utilizando diferentes fuentes de carbono y las relaciones genéticas fueron validadas a partir de información transcriptómica.

ABSTRACT

The use of microbial consortia for wastewater treatments has been positioned as a viable alternative among the different processes that currently exist. However, this option still presents challenges due to the complexity of the microbial interactions, obtaining different results than expected, and, therefore, a system with low reliability. Currently, there are tools to design models and make inferences of the metabolism of members of microbial communities using approaches based on systems biology and quantitative biology.

In the present work, the metagenome of a nitrifying microbial consortium for wastewater treatment was analyzed. From this information, metabolic models were designed for the main members of the consortium using semi-automatic strategies. The resulting models were refined to increase the quality of genetic associations and robustness in phenotypic predictions. Six metabolic models were homologically reconstructed using previously validated micro-organism models from the BiGG database. Templates were chosen based on comparisons by 16S RNA, BLAST between proteomes and physiological similarities. The models were developed combining algorithms from COBRA, RAVEN toolboxes and algorithms designed in this work. The optimal parameters for the semi-automatic reconstruction were calculated from information of different databases and experimental data. The resulting models were optimized at genetic and biochemical levels based on manual curation and semi-automatic algorithms. The models of the six microorganisms (*Rhodopseudomonas palustris* BisA53, *Thauera* sp. MZ1T, *Acinetobacter oleivorans* DR1, *Aromatoleum aromaticum* EbN1, *Dechloromonas aromatica* RCB and *Nitrosomonas europaea* ATCC 19718) belonging to the nitrifying microbial consortium will help to comprehend the metabolic processes associated with the carbon and nitrogen fixation, nitrification, denitrification, etc., fundamental in the wastewater treatment. Finally, the semi-automatic tools and strategies for reconstruction and validation developed were applied in two different projects: *Azotobacter vinelandii* DJ model (iDT1278) and the multiple metabolic models for *Liberibacter* microorganisms. iDT1278 was validated with growth data on more than 200 carbon sources and 95 nitrogen sources, obtaining an accuracy greater than 90% of prediction. Regarding the *Liberibacter* strains, the accuracy in the predictions from the models were validated through different carbon sources and the genes associations were evaluated using transcriptomic data.

ÍNDICE

INTRODUCCIÓN

El uso de consorcios microbianos para el tratamiento de aguas residuales se ha posicionado como una alternativa competitiva entre los diferentes procesos existentes (Monica et al., 2011; Zhu et al., 2019). Los sistemas biológicos son amigables con el medio ambiente y permiten la remoción de una gran cantidad de diferentes compuestos presentes en las aguas, aprovechando el metabolismo y las interacciones de los microorganismos pertenecientes a la comunidad microbiana (Zhang et al., 2017; Zuñiga et al., 2017; Zuñiga et al., 2019). Sin embargo, esta opción aún presenta algunos retos debido a diferentes factores físicos (temperatura, pH, oxígeno disuelto, etc.) y factores biológicos (ejemplo; complejidad que existe en las interacciones dentro de las comunidades microbianas) (Nadiri et al., 2018; Sulaiman et al., 2014), que generan variabilidad entre los resultados obtenidos experimentalmente y predichos, por tanto, estos sistemas biológicos registran una confiabilidad limitada (Perez-Garcia et al., 2016).

Actualmente han surgido diferentes herramientas y tecnologías computacionales que permiten comprender y describir las capacidades y el potencial de los microorganismos individualmente y en comunidades microbianas (Jhu & Zhang, 2015). Una de las principales herramientas para la comprensión de estos sistemas biológicos es el modelamiento matemático del metabolismo desde una perspectiva en biología de sistemas (Thiele & Palsson, 2010), cuyo enfoque se basa en la predicción de los fenotipos (crecimiento celular, flujos metabólicos, rutas metabólicas activas, análisis genético, etc.) a partir de la información genómica y bioquímica (O'Brien et al., 2013). Dentro de las alternativas para diseñar modelos metabólicos, las estrategias de modelamiento semiautomático se han posicionado como una de las opciones más viables debido a que permiten diseñar modelos de alta precisión en tiempos relativamente cortos (meses) (Zuñiga et al., 2016; Tec-Campos et al., 2020).

En este trabajo se diseñaron algoritmos y estrategias semiautomáticas para construir modelos matemáticos del metabolismo que permitan analizar con alta precisión y sensibilidad los genotipos y fenotipos de los microorganismos pertenecientes a un consorcio microbiano nitrificante desde un enfoque en biología de sistemas. A partir de estos modelos se realizaron inferencias y predicciones certeras de las rutas metabólicas de los principales miembros del consorcio microbiano bajo diferentes condiciones experimentales específicas para cada microorganismo. Por último, las herramientas y estrategias semiautomáticas fueron evaluadas en el microorganismo diazótrofo *Azotobacter vinelandii* DJ (iDT1278) y en el conjunto de cepas de *Liberibacter*. Se evaluaron para iDT1278 incialmente 38 fuentes de carbono y nitrógeno molecular a partir de datos de la literatura. A partir de datos obtenidos en el presente trabajo, se evaluaron 190 fuentes de carbono y 95 fuentes de nitrógeno para medir el crecimiento en estas condiciones. Además, se evaluaron datos de fluxómica recopilados en la literatura para la fijación de nitrógeno y producción de biopolímeros utilizando glucosa como fuente de

carbono. Con respecto a las cepas de *Liberibacter*, se validaron diferentes fuentes de carbono como carbohidratos, ácidos orgánicos y aminoácidos para determinar la capacidad de crecimiento en diferentes condiciones. Las asociaciones genéticas de estos modelos fueron evaluadas a partir de información transcriptómica.

HIPÓTESIS

La biología de sistemas permite obtener modelos a escala del genoma con una alta precisión. El uso de estas herramientas y estrategias computacionales resulta en la construcción y diseño de modelos metabólicos desde un enfoque semiautomático. Los modelos metabólicos permitirán realizar predicciones fenotípicas de crecimiento y entender la distribución de flujos metabólicos en microorganismos participantes en un consorcio microbiano nitrificante (bacterias fijadoras de nitrógeno, amonio oxidantes, desnitrificantes, oxidantes de compuestos orgánicos, etc.) para el tratamiento de aguas residuales.

JUSTIFICACIÓN

Los sistemas biológicos permiten la remoción parcial o total de una gran variedad de compuestos presentes en las aguas residuales. Sin embargo, existen dificultades en el control y manipulación de las condiciones de cultivo debido a la complejidad de las interacciones microbianas, al realizar modificaciones químicas o físicas. Entre las alternativas para su análisis y comprensión se encuentra el modelado basado en restricciones. A pesar de que existen algunas plataformas automatizadas (KBase, Model SEED), estas acarrean grandes problemas durante el proceso de reconstrucción de los modelos metabólicos (Fritzemeier et al., 2017), por tanto, surge la necesidad de construir herramientas que permitan construir con rapidez y confiabilidad modelos de comunidades microbianas y de esta manera poder elucidar el comportamiento de los microorganismos bajo diferentes condiciones de aguas residuales. En este trabajo se pretende desarrollar algoritmos que permitan la construcción de modelos individuales y de comunidad, presentando una nueva alternativa que permita generar modelos en un tiempo menor a las estrategias manuales, evitando la adición de bucles energéticos (EGCs por sus siglas en inglés) en el proceso de construcción. Específicamente, se pretende analizar procesos relacionados con el tratamiento de aguas residuales sintéticas, principalmente bajo condiciones nitrificantes y la presencia de compuestos orgánicos identificados como contaminantes, cuyos experimentos se han desarrollado y estudiado previamente por nuestro equipo de trabajo De esta forma se abarca un enfoque poco estudiado en la modelación de comunidades microbianas aplicado al tratamiento de aguas residuales, ya que los trabajos de modelos de consorcios microbianos principalmente se enfocan en el estudio del microbioma (Pérez-García et al., 2016; Zuñiga et al., 2017), estableciendo las condiciones adecuadas para la proliferación y funcionamiento de los microorganismos que tienen mayor participación dentro de la comunidad, optimizando *in silico* los parámetros del proceso. El modelo de comunidad resultante permitirá validar la información experimental previamente obtenida por medio de un consorcio microbiano nitrificante, y de esta manera poder elucidar el comportamiento y la interacción de los microorganismos integrantes (bacterias fijadoras de nitrógeno, amonio oxidantes, desnitrificantes, oxidantes de compuestos orgánicos, etc.) del consorcio microbiano nitrificante utilizando la información fisiológica y molecular obtenida en trabajos previamente realizados por el grupo.

OBJETIVOS

Objetivo general:

- Reconstruir y validar modelos metabólicos a escala del genoma de miembros de una comunidad microbiana nitrificante presente en aguas residuales, representando el metabolismo de cada miembro utilizando métodos y principios de biología de sistemas

Objetivos específicos:

- Desarrollar modelos metabólicos individuales de los microorganismos identificados en el metagenoma del consorcio nitrificante (SRR1801934) utilizando como herramientas complementarias las cajas de herramientas COBRA y RAVEN

- Validar la exactitud y precisión de los modelos metabólicos individuales a partir de la información presente en la literatura

- Evaluar *in silico* las capacidades metabólicas de cada uno de los microorganismos modelables para intercambiar metabolitos
- Identificar la participación *in silico* de los microorganismos modelables bajo condiciones experimentales de tratamiento de aguas residuales sintéticas previamente obtenidas por el grupo del Dr. Zepeda

REFERENCIAS

Campos, D. T., Zuñiga, C., Passi, A., Del Toro, J., Tibocha-Bonilla, J. D., Zepeda, A., Betenbaugh, M. J., & Zengler, K. (2020). Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ. Metabolic Engineering Communications, e00132. https://doi.org/10.1016/j.mec.2020.e00132

Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B., & Lercher, M. J. (2017). Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLOS Computational Biology, 13(4), e1005494. https://doi.org/10.1371/journal.pcbi.1005494

Ju, F., & Zhang, T. (2015). Experimental Design and Bioinformatics Analysis for the Application of Metagenomics in Environmental Sciences and Biotechnology. Environmental Science & Technology, 49(21), 12628–12640. https://doi.org/10.1021/acs.est.5b03719

Monica, S., Karthik, L., Mythili, S., & Sathiavelu, A. (2011). Formulation of Effective Microbial Consortia and its Application for Sewage Treatment. Journal of Microbial & Biochemical Technology, 3(3). https://doi.org/10.4172/1948-5948.1000051

Nadiri, A. A., Shokri, S., Tsai, F. T.-C., & Asghari Moghaddam, A. (2018). Prediction of effluent quality parameters of a wastewater treatment plant using a supervised committee fuzzy logic model. Journal of Cleaner Production, 180, 539–549. https://doi.org/10.1016/j.jclepro.2018.01.139

O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Molecular Systems Biology, 9(1), 693. https://doi.org/10.1038/msb.2013.52

Perez-Garcia, O., Lear, G., & Singhal, N. (2016). Metabolic Network Modeling of Microbial Interactions in Natural and Engineered Environmental Systems. Frontiers in Microbiology, 7. https://doi.org/10.3389/fmicb.2016.00673

Sulaiman, S., Khamis, M., Nir, S., Scrano, L., Bufo, S. A., & Karaman, R. (2014). Diazepam stability in wastewater and removal by advanced membrane technology, activated carbon, and micelle–clay complex. Desalination and Water Treatment, 57(7), 3098–3106. https://doi.org/10.1080/19443994.2014.981225

Tan, J., Zuniga, C., & Zengler, K. (2015). Unraveling interactions in microbial communities – from co-cultures to microbiomes. Journal of Microbiology, 53(5), 295–305. https://doi.org/10.1007/s12275-015-5060-1

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols, 5(1), 93–121. https://doi.org/10.1038/nprot.2009.203

Zhang, B., Xu, X., & Zhu, L. (2017). Structure and function of the microbial consortia of activated sludge in typical municipal wastewater treatment plants in winter. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-17743-x

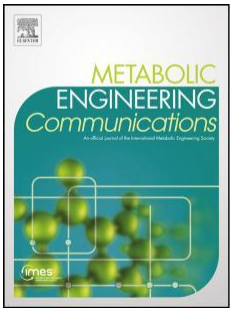
Zhu, S., Huo, S., & Feng, P. (2019). Developing Designer Microalgal Consortia: A Suitable Approach to Sustainable Wastewater Treatment. In Microalgae Biotechnology for Development of Biofuel and Wastewater Treatment (pp. 569–598). https://doi.org/10.1007/978-981-13-2264-8_22

Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D. C., McConnell, B. O., Long, C. P., Knoshaug, E. P., Guarnieri, M. T., Antoniewicz, M. R., Betenbaugh, M. J., & Zengler, K. (2016). Genome-Scale Metabolic Model for the Green Alga *Chlorella vulgaris* UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiology, 172(1), 589–602. https://doi.org/10.1104/pp.16.00593

Zuñiga, C., Li, C.-T., Yu, G., Al-Bassam, M. M., Li, T., Jiang, L., … Zengler, K. (2019). Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. Nature Microbiology, 4(12), 2184–2191. https://doi.org/10.1038/s41564-019-0567-6

Zuñiga, C., Zaramela, L., & Zengler, K. (2017). Elucidation of complexity and prediction of interactions in microbial communities. Microbial Biotechnology, 10(6), 1500–1522. https://doi.org/10.1111/1751-7915.12855

PRIMER ARTÍCULO DE INVESTIGACIÓN



Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ

Diego Tec Campos, Cristal Zuñiga, Anurag Passi, John Del Toro, Juan D. Tibocha-Bonilla, Alejandro Zepeda, Michael J. Betenbaugh, Karsten Zengler

DOI: https://doi.org/10.1016/j.mec.2020.e00132

Reference: MEC 132

To appear in: *Metabolic Engineering Communications*

Received Date: 29 January 2020

Revised Date: 9 May 2020

Accepted Date: 11 May 2020

Please cite this article as: Campos, D.T., Zuñiga, C., Passi, A., Del Toro, J., Tibocha-Bonilla, J.D., Zepeda, A., Betenbaugh, M.J., Zengler, K., Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ, *Metabolic Engineering Communications* (2020), doi: https://doi.org/10.1016/j.mec.2020.e00132.

CARTA DE ACEPTACIÓN DE PRIMER ARTÍCULO

Dear Dr Tec,

Thank you for publishing your article in Metabolic Engineering Communications. Professor. Zengler completed the Rights and Access Form for your article *Modeling of nitrogen metabolism and polymer production in the heterotrophic diazotroph Azotobacter vinelandii DJ* on June 05, 2020.

The Order Summary is attached to this email. A copy of the Order Summary is also sent to all co-authors for whom we have contact details. Your article is free for everyone to read online at https://doi.org/10.1016/j.mec.2020.e00132

If you have any questions, please do not hesitate to contact us. To help us assist you, please quote our article reference MEC_e00132 in all correspondence.

Now that your article has been accepted, you will want to maximize the impact of your work. Elsevier facilitates and encourages authors to share their article responsibly. To learn about the many ways in which you can share your article whilst respecting copyright, visit: www.elsevier.com/sharing-articles.

Kind regards,
Elsevier Researcher Support

Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ

Short title: Genome-scale metabolic modeling of *Azotobacter vinelandii* DJ

Diego Tec Campos[1,2], Cristal Zuñiga[1], Anurag Passi[1], John Del Toro[3], Juan D. Tibocha-Bonilla[4], Alejandro Zepeda[2], Michael J. Betenbaugh[3], Karsten Zengler[1,5,6]*

[1]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA
[2]Facultad de Ingeniería Química, Universidad Autónoma de Yucatán, Mérida, Yucatán, México
[3]Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[4]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla CA 92093-0412, USA.
[5]Department of Bioengineering, University of California, San Diego, La Jolla CA 92093-0412, USA.
[6]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0403, USA

*Corresponding author: Karsten Zengler
Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA. Phone: (858) 504-0823, E-mail: kzengler@ucsd.edu

Abstract

Nitrogen fixation is an important metabolic process carried out by microorganisms, which converts molecular nitrogen into inorganic nitrogenous compounds such as ammonia (NH_3). These nitrogenous compounds are crucial for biogeochemical cycles and for the synthesis of essential biomolecules, i.e. nucleic acids, amino acids and proteins. *Azotobacter vinelandii* is a bacterial non-photosynthetic model organism to study aerobic nitrogen fixation (diazotrophy) and hydrogen production. Moreover, the diazotroph can produce biopolymers like alginate and polyhydroxybutyrate (PHB) that have important industrial applications. However, many metabolic processes such as partitioning of carbon and nitrogen metabolism in *A. vinelandii* remain unknown to date.

Genome-scale metabolic models (M-models) represent reliable tools to unravel and optimize metabolic functions at genome-scale. M-models are mathematical representations that contain information about genes, reactions, metabolites and their associations. M-models can simulate optimal reaction fluxes under a wide variety of conditions using experimentally determined constraints. Here we report on the development of a M-model of the wild type bacteria *A. vinelandii* DJ (*i*DT1278) which consists of 2,003 metabolites, 2,469 reactions, and 1,278 genes. We validated the model using high-throughput phenotypic and physiological data, testing 180 carbon sources and 95 nitrogen sources. *i*DT1278 was able to achieve an accuracy of 89% and 91% for growth with carbon sources and nitrogen source, respectively. This comprehensive M-model will help to comprehend metabolic processes associated with nitrogen fixation, ammonium assimilation, and production of organic nitrogen in an environmentally important microorganism.

1. Introduction

Azotobacter vinelandii is a gram-negative soil bacterium capable of converting atmospheric nitrogen gas (N_2) into soluble ammonia as well as into other important nitrogenous compounds (Gyurján et al., 1995; Howard and Rees, 1996). *Azotobacter* and related *Azospirillium* are estimated to fix up to 10-30% of the total nitrogen in the rhizosphere (Cleveland et al., 1999). Nitrogen fixation can be carried out under ambient conditions by any of the three highly specialized metal-dependent nitrogenases, referred to as molybdenum nitrogenase, vanadium nitrogenase, and iron-only nitrogenase (Setubal and Almeida, 2015; Sippel et al., 2017). Nitrogenases produce high concentrations of fixed ammonium, which is excreted and serves as essential nutrient for other organisms (Ambrosio et al., 2017). However, the activity of these enzymes is highly sensitive to molecular oxygen and energetically costly. Diazotrophs, such as *A. vinelandii* have developed specific strategies to protect the nitrogenase complex in diazotrophic conditions (Setubal and Almeida, 2015). One of the most studied mechanisms for nitrogenase protection is alginate biosynthesis. Alginate is transported to the extracellular space where it works as a barrier that decreases oxygen diffusion into the cytoplasm and thus maintains high functionality of oxygen-sensitive nitrogenases in anoxic environments (García et al., 2018; Pacheco-Leyva et al., 2016).

Alginate is of great industrial value because of its use as biocompatible and biodegradable exopolysaccharide. This polymer is used as gel-film-stabilizing, -thickening, or -forming agent in the food and pharmaceutical industry (Remminghorst and Rehm, 2006). Besides alginate bioproduction, *A. vinelandii* produces another attractive commercial polymer, i.e. polyhydroxybutyrate (PHB) (Yoneyama et al., 2015). PHB is synthetized by this microorganism under high carbon/nitrogen ratios as a carbon and energy reserve in the form of cysts (Stevenson and Socolofsky, 1966). Both biopolymers can be produced in elevated concentrations, representing 30-70% of the dry biomass (Pacheco-Leyva et al., 2016; Yoneyama et al., 2015).

vinelandii has been shown to grow under a broad range of heterotrophic conditions and is able to metabolize, different sugars, alcohols, and organic acids as well as nitrogen-containing compounds (Nagai et al., 1972; Quiroz-Rocha et al., 2017; Sahoo et al., 2014; Shawky et al., 1987). Despite this metabolic versatility to use different carbon and nitrogen sources, several of the internal metabolic processes regarding carbon and nitrogen partitioning (division and distribution of an element into metabolic, structural or storage pools) in *A. vinelandii* remain unknown. Today there are five fully sequenced genomes available for *A. vinelandii* strains (e.g. *A. vinelandii* CA, DJ, CA6, DSM 279, and NBRC13581) (Noar et al., 2015; Setubal et al., 2009; Setubal and Almeida, 2015), enabling a comprehensive functional characterization of *Azotobacter* metabolism at genome-scale.

To comprehend the metabolic capabilities of *Azotobacter vinelandii* DJ we used a systems biology approach, which offers tools to predict the organism behavior based on mathematical representations of biological data. M-models can be reconstructed using semi-automated tools that generate a draft model. This draft model is further curated manually to increase its quality. To date, only two core M-models of *Azotobacter vinelandii* are available that contain a reduced number of metabolic reactions. These core reactions are in general related to nitrogen fixation or PHB and alginate production, disregarding most of the central metabolism of the microorganism (e.g. TCA cycle, lipid metabolism and some amino acids synthesis) (García et al., 2018; Inomura et al., 2018). Here we have developed a M-model for *Azotobacter vinelandii* DJ to contextualize metabolic processes associated with nitrogen fixation, ammonium assimilation, and production of organic nitrogen on genome-scale. Our model was successfully validated using high-throughput phenotypic data and physiological data.

2. Material and Methods

2.1 Draft model generation

The draft model of *A. vinelandii* DJ was generated using The COBRA (Heirendt et al., 2019) and The RAVEN (Agren et al., 2013) Toolboxes. The proteome sequence was obtained from PATRIC database (Genome ID: 322710.5) and was used as input sequence to reconstruct the draft model based on protein homology. We selected five reference models as templates after alignment of the complete genome sequences of *A. vinelandii* DJ with all bacteria with available models in the BiGG Database (King et al., 2016). Templates included *Escherichia coli* str. K-12 substr. MG1655, model *i*ML1515 (Monk et al., 2017), *Klebsiella pneumoniae* subsp. Pneumoniae MGH 78578, model *i*YL1228 (Liao et al., 2011), *Geobacter metallireducens* GS-15, model *i*AF987 (Feist et al., 2014), *Clostridium ljungdahlii* DSM 13528, model *i*HN637 (Nagarajan et al., 2013), and *Methanosarcina barkeri* str. Fusaro, model *i*AF692 (Feist et al., 2006). Template models contained reactions associated with nitrogen fixation, H_2 production, acetate consumption, amino acids catabolism and sugar degradation (Fig. 2). The generated draft model also contained genes (exogenous genes) from template models, which were later removed during the manual curation step.

2.2 Model Refinement

2.2.1 Manual Curation

We used PATRIC (Wattam et al., 2017) to identify essential genes for *A. vinelandii DJ* in the final model. We only extracted those genes that had a given enzyme commission (EC) number that could be used to obtain the GPR (gene-protein reaction) associations. The final list of reactions with EC number and gene association not previously present in the model were balanced and added to the model before analyzing GPR associations.

Model refinement included two major steps: manual curation/review of the GPR associations and gap-filling by adding new metabolic reactions in the model. In the first step of manual curation, we determined sequence similarity among *A. vinelandii DJ* proteins and the exogenous proteins in the GPRs to identify *A. vinelandii* (AVIN) genes closely related to the exogenous proteins. We identified proteins based on BLASTp criteria of >= 40% identity, e-value <=1e-4, and query coverage >=85%. A second step of manual curation was performed based on protein function, type of metabolic reaction, and GPR associations. Then all the GPR associations were manually curated to catalyze biological reactions that they were associated with. PATRIC essential genes previously identified were added in this step of manual curation. Remaining reactions with mixed AVIN and exogenous genes in the GPR association were manually curated in order to remove the genes that did not belong to *A. vinelandii*. Reactions with exclusively exogenous GPR associations were identified through previous manual curation steps.

Afterwards, Flux Balance Analysis (FBA) was performed to identify which of these reactions carry any flux under experimental conditions (Orth et al., 2010). From these evaluated reactions, those with no flux and exogenous GPR associations were removed from the model.

2.2.3 Gap-filling

Gap analysis was performed in order to identify the metabolites disconnected in the model. These metabolites were classified depending on the number of reactions present in the model or their capability to be consumed, produced, or both. Disconnected reactions were manually curated using information from different bioinformatic databases (e.g. KEGG, Biocyc). From these results, gap-filling was used to connect pathways through the data retrieved. A second step of gap-filling was accomplished to connect the metabolites from the medium conditions retrieved using literature information (Wong and Maier, 1985) through algorithms to identify the reactions involved in the carbon source assimilation. A total of 38 carbon sources were used under nitrogen fixation and ammonium assimilation conditions. Complementary, experimental data were generated using Biolog plates to test different carbon and nitrogen sources. This was employed to improve the quality of model predictions under a wide variety of conditions. A set of 190 carbon sources and 95 nitrogen sources were used to connect the networks properly. Subsequently, the GPR associations were verified for each reaction added during the gap-filling to maintain the quality of the model. Those reactions with no gene information and literature validation were conserved as orphan reactions.

2.2.3 Final Quality Control and Quality Analysis

Final quality check was performed by a person who did not perform the manual curation to assess the quality of the data. We performed *in-silico* GPR simulations to verify if the GPR associations are correctly assigned using the COBRA Toolbox algorithms. Next, we performed Mass Balance simulations on the model to check for unbalanced reactions added during the model refinement. Ultimately, the final model was tested looking for ATP, NADH, and NADPH free energy production, removing exchange reactions, and calculating their accumulation.

2.3 Constraints and Growth simulations

Experimental data from the literature were retrieved to calculate the initial medium constraints. For each growth condition, the carbon, nitrogen, and hydrogen fluxes were initially determined depending on every value obtained from the literature. The constraints related to mineral compounds and exchange reactions are summarized in Table S1. Initially, a set of six different conditions were used to measure the accuracy of the model. The carbon sources verified in this stage of validation were carbohydrates under nitrogen fixation or ammonium assimilation conditions. The simulation results were compared to this set of experimental values to identify the quality in the model predictions. Subsequently, 38 carbon sources under nitrogen fixation and H_2 consumption conditions

from the literature (Wong and Maier, 1985) were used to test and increase the quality of the model. The uptake rates were estimated from the experimental conditions and set for all the carbon sources; nitrogen and H_2 uptake rates were not fixed to a specific value according to the experimental conditions. Finally, model benchmarking was performed for 190 carbon sources (Biolog plates PM1 and PM2) and 95 nitrogen sources (Biolog plates PM3) to validate model predictions. Biolog microplates experiments to test carbon and nitrogen assimilation were performed in the present work, measuring the growth rate values in the plate reader for 96 h. For carbon sources evaluation, ammonium assimilation was not fixed to a specific value (non-diazotroph conditions). The experimental results from Biolog plates were matched with data retrieved from the literature to determine and evaluate model precision during the simulations. During the nitrogen condition simulations, pyruvate was used as the unique carbon source. Statistical parameters were calculated according to the comparison between the metabolic predictions and the experimental values. The model accuracy from the Biolog plates results was compared with the *in-silico* predictions of the *A. vinelandii* model from CarveMe to identify the quality of the model simulations of the present work. The alginate production capability of the model was tested using four different carbon sources (carbohydrates) from the literature (Revin et al., 2018).The carbon compound uptake rates were calculated according to the experimental values. The simulations were performed initially setting ammonium as unique nitrogen source and subsequently molecular nitrogen was established as the unique nitrogen source. Furthermore, the predicted values were compared to determine which conditions allow a higher alginate production rate. For polyhydroxybutyrate (PHB) production, the metabolic internal fluxes for the principal pathways related to the PHB synthesis were calculated (glycolysis, pentose phosphate pathway, Entner-Doudorrff pathway, and TCA cycle) and compared with fluxomic data determined by Wu et al. (2019). *In silico* predictions were performed through FBA, using The COBRA Toolbox and the Gurobi Optimizer v.8.0.1 solver (Gurobi Optimization) for MATLAB (MathWorks). Percent error between experimental values and *in silico* results were calculated to obtain model accuracy.

2.4 Carbon and nitrogen partitioning analysis

The growth results from the Biolog plates experiments were used to determine the carbon and nitrogen distribution across the metabolism. The internal fluxes for all the reactions of the model were calculated *in-silico* for all the carbon sources (PM1 and PM2) experiments. The reactions were grouped in general subsystems that represented the complete metabolism of *A. vinelandii* DJ. Subsequently, an average flux per subsystem was calculated using the flux values of all the reactions belonging to the subsystem. This procedure was performed to calculate the carbon and nitrogen distribution in each general subsystem under diazotrophic and non-diazotrophic conditions. Ultimately, the carbon and nitrogen distributions (the grouped average fluxes per subsystem) were compared through a linear correlation analysis in order to determine how the fluxes change through the experimental conditions for carbon sources set in the Biolog plates.

High linear correlation coefficients I indicate a similar distribution of a specific subsystem (above 0.9), while low coefficients suggest different carbon or nitrogen distributions when comparing diazotroph and non-diazotroph conditions.

3 Results

3.1 Metabolic network reconstruction of *A. vinelandii* DJ

We used a semiautomatic approach to reconstruct the M-model of *A. vinelandii* DJ (Fig. 1). This approach has been previously applied for the reconstruction of M-models (Zuñiga et al., 2016). First, a draft model of *A. vinelandii* DJ was reconstructed using the genome annotation from PATRIC (Genome ID: 322710.5). Five manually curated and validated M-Models were used as protein homology templates: *Escherichia coli* str. K-12 substr. MG1655 (Monk et al., 2017), *Klebsiella pneumoniae* subsp. Pneumoniae MGH 78578 (Liao et al., 2011), *Geobacter metallireducens* GS-15 (Feist et al., 2014), *Clostridium ljungdahlii* DSM 13528 (Nagarajan et al., 2013), and *Methanosarcina barkeri* str. Fusaro (Feist et al., 2006). The RAVEN and COBRA Toolboxes (Agren et al., 2013; Heirendt et al., 2019) were used to generate the draft reconstruction. Each reaction in the draft model was evaluated for energy (ATP, NADH and NADPH accumulation) and mass balances as part of the quality control tests to guarantee model functionality and accuracy. Reactions associated with template genes were conserved in the first draft model to ensure model connectivity as well as the model's capability to perform simulations. Nitrogen fixation and hydrogen consumption reactions were imported from the M-model templates. The resulting draft model contained 2,432 metabolic reactions and 1,918 metabolites divided into three different compartments (cytoplasm, periplasm, and extracellular space).

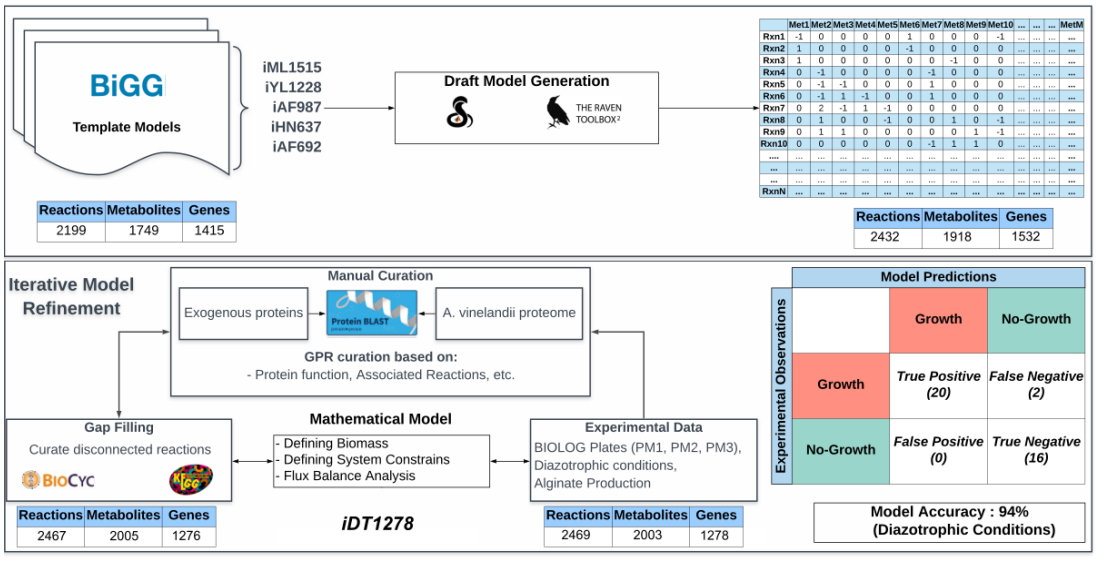


Fig. 1 Workflow used to reconstruct a metabolic model of A. vinelandii DJ. A draft model was created from five template models present in BiGG (Escherichia coli str. K-12 substr. MG1655, Klebsiella pneumoniae subsp. Pneumoniae MGH 78578, Geobacter metallireducens GS-15, Clostridium ljungdahlii DSM 13528 and Methanosarcina barkeri

str. Fusaro). The RAVEN toolbox for MATLAB was used to create the draft model from stoichiometric data. The initial draft model contained 2,432 reactions, 1,918 metabolites, and 1532 genes. The iterative process of model refinement included manual curation, gap-filling and curation using experimental data. The resultant final model contained A. vinelandii specific metabolic processes such as nitrogen fixation, and production of alginate and PHB. The final model, containing 2,469 reactions, 2003 metabolites, and 1278 genes, predicted with 94% accuracy.

3.1.1 Model Refinement

Model refinement was performed using two principal steps: manual curation and gap filling. Every gene-protein-reaction (GPR) association was verified using multiple databases (e.g. KEGG, Biocyc, BRENDA, and MetaNetX) and available information from the literature. Manual curation was based on protein sequence similarity. The genes annotated in the GPR associations were aligned to protein sequences of *A. vinelandii*. Sequences, which passed the BLASTp parameters (see Methods), were assigned functionality based on information in the bioinformatics databases. The assigned genes in the GPR rules were replaced with the *A. vinelandii* genes (AVIN).

The original draft model consisted of 1,532 genes (Fig. 2A) corresponding to 934 AVIN genes and 598 genes from the template. At the end of the second manual curation, the AVIN genes increased to 1,233 and the total number of template genes was 102. Intuitively, as the level of curation increases the number of genes from template models decreases. When the genes in the model were curated by functionality, the number of genes from template models was zero.

3.1.2 Gap filling

After the manual curation, the total number of reactions and metabolites in the model was 2,416 and 1,976, respectively. We used literature information and experimental data from the Biolog plates results (PM1 and PM2 for carbon sources and PM3 for nitrogen sources) to add or remove reactions in the model. Each reaction added to the model in this step was manually reviewed to maintain concordance in the GPR associations. Overall, a total of 51 reactions and 29 metabolites (mostly reactions related to carbohydrate and amino acids catabolism) were added to the model during the gap filling process. The reactions added to the model were mainly transport and interconversion reactions to connect the carbon or nitrogen sources with intracellular metabolites from the model.

3.1.3 Model properties

The Azotobacter vinelandii DJ metabolic model (*i*DT1278) consists of 2,003 metabolites, 2,469 reactions and 1,278 genes (around 26% of all annotated coding genes in the genome). Specific details about the reactions and metabolites from the model are summarized in Table S1. *i*DT1278 was validated using experimental data under nitrogen fixation (diazotroph) and ammonium assimilation (non-diazotroph) conditions. *i*DT1278 contains all the reactions and genes involved in nitrogen fixation, PHB, and alginate biosynthesis (Fig. 2B).

The properties of *i*DT1278 are shown in Fig. 2. Most of the reactions in the model belong to amino acid metabolism, lipid metabolism, and cofactor and vitamins metabolism (60% of total reactions of the model). Specific metabolic capabilities of *A. vinelandii* DJ such as nitrogen fixation (nitrogen metabolism), PHB and alginate production (glycan and secondary metabolites biosynthesis) represent around 3% of the metabolic reactions. Template models used during the reconstruction share 208 reactions with *i*DT1278. Most of the reactions (a total of 2139 reactions) were taken from the first template (*Escherichia coli* K12 substr. MG1555, *i*ML1515). Nitrogen fixation and H_2 consumption pathways were obtained from the templates *i*HN637(*Clostridium ljungdahlii* DSM 13528) and *i*AF987 (*Geobacter metallireducens* GS-15). *i*DT1278 shares 208 reactions among all the template models (Fig. 2C) which are related to core metabolic pathways (TCA cycle, oxidative phosphorylation, amino acids metabolism, etc.). Table 1 shows a comparison of the properties of the different metabolic models reconstructed for *A. vinelandii*. As a result, *i*DT1278 represents, to our knowledge, the most comprehensive M-model of the diazotroph *A. vinelandii* available to date. However, the *A. vinelandii* metabolic model from CarveMe (Machado et al., 2018) contains the closest number of reactions, metabolites and genes using the BiGG database information.

Table 1 Comparison of the principal model properties (reactions, metabolites and genes) available for *A. vinelandii*.

Model provenience	Reactions	Metabolites	Genes	Reference
Present work	2469	2003	1278	Present work
Model SEED	1570	1416	903	(Henry et al., 2010)
CarveMe	2422	1978	1395	(Machado et al., 2018)
García (Alginate and PHB production)	46	39	0	(García et al., 2018)
Inomura (Nitrogen fixation)	33	17	0	(Inomura et al., 2018)

3.1.4 Biomass Objective Function

The biomass objective function (BOF) contains the principal constituents and the abundance of each metabolite involved in biomass production. The proportion of each metabolite participating in the BOF composition is determined per gram of biomass. *i*DT1278 includes two biomass reactions: 1) An initial BOF was obtained from the first template (*Escherichia coli* K12 substr. MG1555, *i*ML1515) based on their physiological similarity (Gram-negative bacteria); the stoichiometric coefficients of the amino acids present in the BOF were calculated based on the theoretical amino acid abundance in the genome, using 55% of the biomass composition from amino acids. 2) A second BOF was determined from the first reaction to predict the alginate production since *A. vinelandii* DJ produces alginate only under specific metabolic conditions (Noar et al., 2015). The second BOF contains the same constituents present in the first BOF plus periplasmic alginate in order to simulate the complete metabolism and alginate production of *A. vinelandii*.

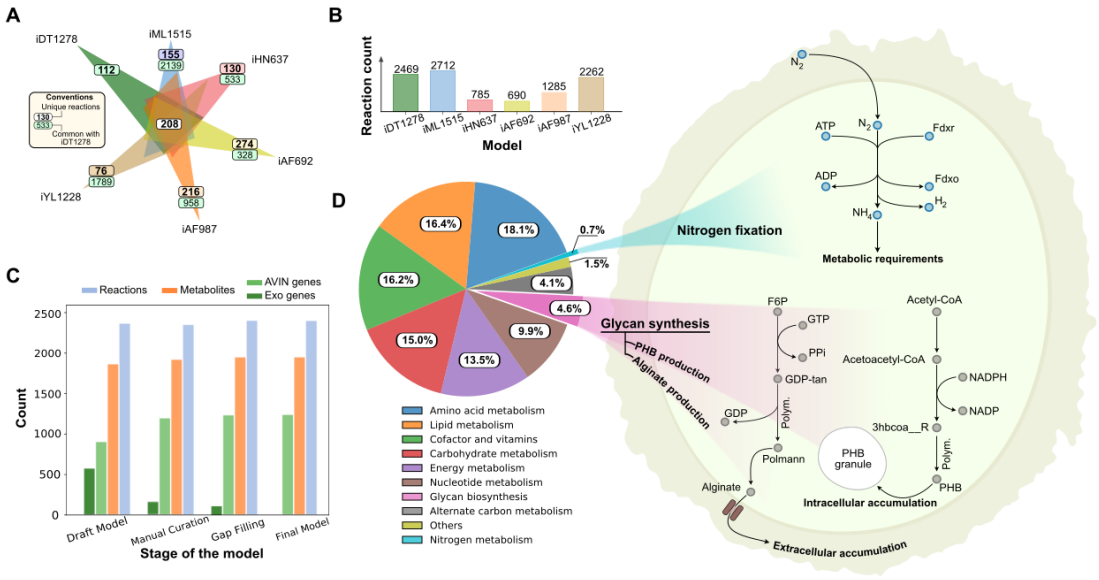


Fig. 2. Characteristics of iDT1278. A) Comparison among the template models (Escherichia coli str. K-12 substr. MG1655, Klebsiella pneumoniae subsp. Pneumoniae MGH 78578, Geobacter metallireducens GS-15, Clostridium ljungdahlii DSM 13528 and Methanosarcina barkeri str. Fusaro) and iDT1278 reactions. The six models share 208 core metabolic reactions. A. vinelandii DJ model contains 112 unique reactions related to aromatic compounds metabolism, alginate and PHB production, etc. B) Number of reactions in the template models and A. vinelandii DJ. C) Change in the number of reactions, metabolites and genes at the different stages of the reconstruction process/manual curation of A. vinelandii DJ. D) Reactions distribution through the subsystems in the genome-scale model; subsystems were grouped into 11 groups summarizing the complete metabolism of A. vinelandii DJ. Nitrogen fixation, PHB, and alginate production are highlighted in the condensed pathway diagrams. Alginate

accumulation occurs in the extracellular space meanwhile PHB storage happens in the cytoplasm compartment. Nitrogen fixation in A. vinelandii can be performed by different specialized nitrogenases.

3.2 iDT1278 predicts accurately phenotypic experimental data

3.2.1 Growth rates validation in carbon and nitrogen sources

The model was validated under a wide range of different growth conditions (diazotrophic and non-diazotrophic growth), using high-throughput phenotypic data as well as literature information. Initially, *i*DT1278 was tested under six different experimental conditions, specifically, carbohydrates under diazotrophic and non-diazotrophic conditions (Table 2). The M-model predicted precisely the growth rates for all the carbon sources using ammonium or molecular nitrogen as nitrogen sources. For the carbon sources in non-diazotrophic conditions (sucrose, mannitol and glucose), the predicted growth rates are consistent with experimental values obtained from the literature, resulting in an average accuracy close to 95%. For example, the predicted growth using mannitol (uptake rate of 0.83 mmol/gDWh) as sole carbon source was 0.0472 h^{-1}, agreeing with the experimental data (0.045±0.003 h^{-1}). Average precision under nitrogen fixation conditions decreased significantly to 83%. Table 2 shows the comparison between experimental data from literature and predicted values for *A. vinelandii* DJ. Initial results showed higher model accuracy (12% more) when predicting growth rates using ammonium as nitrogen source compared to N_2. Subsequently, flux balance analysis (FBA, Orth et al., 2010) was performed for a group of 38 carbon sources in diazotrophic and H_2-consuming conditions (Wong and Maier, 1985). Statistical results show for the subset of 38 carbon sources (Fig. 3C) an accuracy of 95%, with 20 true positive predictions (100% positive predicted) and 16 true negative predicted results (89% negative predicted). Matthews correlation coefficient (MCC) was calculated under the conditions previously mentioned, obtaining a value of 0.67 (Fig. 3C). The false negative predictions obtained during the validation of aconitate and lactose are related to the absence of literature information about the enzymes which metabolize the carbon sources into familiar metabolites for the microorganism (e.g. the model lacks enzymes to convert lactose into glucose and fructose).

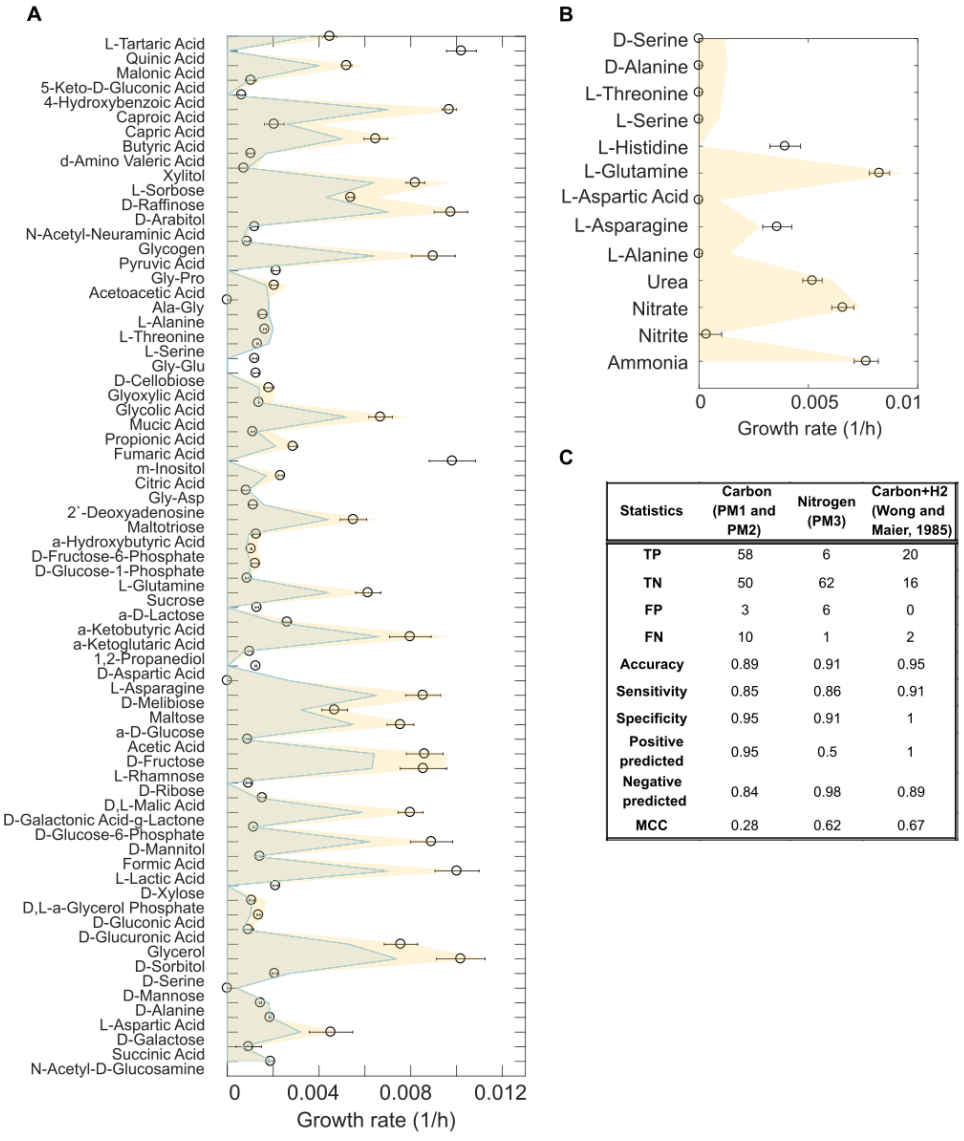
Table 2 Predicted and experimental growth rates reported for *A. vinelandii* DJ under different carbon and nitrogen sources.

Carbon source	Nitrogen source	Experimental value (h^{-1})	Predicted growth (h^{-1})	Reference
Glucose	Ammonium	0.0505	0.0486	(Clementi, 1997)
Mannitol	Ammonium	0.045±0.003	0.0472	(Revin et al., 2018)
Sucrose	Ammonium	0.076±0.004	0.07	(Díaz-Barrera et al., 2016)
Glucose	Nitrogen	0.06±0.0002	0.09	(Wong, 1988)
Fructose	Nitrogen	0.048±0.002	0.0517	(Wong, 1988)
Galactose	Nitrogen	0.074±0.007	0.065	(Wong, 1988)

Additional experimental validation was performed using Biolog plates for a set of carbon (PM1 and PM2) and nitrogen (PM3) sources to determine the growth rate values of *A. vinelandii* DJ. Out of 190 carbon sources from the Biolog plates, 123 compounds were identified in the model; the simulations were performed under two specific conditions: diazotrophic and non-diazotrophic simulation conditions. However, experimental results were obtained only under growth with ammonium as the unique nitrogen source. The same procedure used in PM1 and PM2 experiments was followed to estimate the growth rates with 75 different nitrogen sources. For this case, simulations were performed using pyruvate as the carbon source. Fig. 3 shows the complete analysis of the experimental and predicted data for all carbon (Fig. 3A) and nitrogen sources (Fig. 3B); statistical parameters (accuracy, sensitivity, specificity, positive predicted, negative predicted and Matthews correlation coefficient) were calculated for non-diazotrophic conditions. Subsequently, the same experimental results from the Biolog plates were used to calculate the statistical parameters of the CarveMe model to establish a comparison between iDT1278 and the CarveMe metabolic model accuracy. Details of CarveMe predictions and statistical parameters using Biolog plates data are presented in Table S3.

For carbon sources validation, an accuracy of 89% was achieved with 58 true positive predictions (95% positive predicted) and 50 true negative estimations (84% negative predicted). For this subset of compounds, accuracy decreased significantly in negative predictions (10 false negative estimations). These negative prediction disagreements involve carbohydrates and some amino acids as carbon sources. Some false negatives appear to be related to the lack of evidence about required transporters. Simulated growth rate values in ammonium assimilation conditions were significantly higher than in the nitrogen fixation conditions (>26%). Higher accuracy was observed in both positive and negative predictions in iDT1278 compared to CarveMe simulations (global accuracy of 61%, 72% positive predicted and 54% negative predicted). These statistical results support the clear differences between the quality predictions of iDT1278 and other metabolic models available for *A. vinelandii*. Significant changes in the growth rate values

were not observed when using amino acids as carbon sources since these organic molecules contain nitrogen, which provides an indirect supply of this element to the microbe. For the nitrogen source validation, an overall accuracy of 91% was accomplished, nonetheless, the number of non-growing conditions increased considerably in comparison with the carbon condition experiments (almost 91% of nitrogen experiments resulted in no growth), indicating a reduced capability to grow in a wide range of nitrogen sources.



Statistics	Carbon (PM1 and PM2)	Nitrogen (PM3)	Carbon+H2 (Wong and Maier, 1985)
TP	58	6	20
TN	50	62	16
FP	3	6	0
FN	10	1	2
Accuracy	0.89	0.91	0.95
Sensitivity	0.85	0.86	0.91
Specificity	0.95	0.91	1
Positive predicted	0.95	0.5	1
Negative predicted	0.84	0.98	0.89
MCC	0.28	0.62	0.67

Fig. 3. Model validation using high-throughput phenotypic data for different carbon and nitrogen sources. A) Predicted growth values of A. vinelandii DJ in nitrogen fixation (N_2 + Carbon source) or ammonium assimilation (NH_4 + Carbon source) conditions using 121 carbon sources (PM1 and PM2 from Biolog Plates). However, only 71 carbon sources

are shown in Fig. 3. B) Estimated growth for 75 nitrogen sources, nevertheless just 12 nitrogen compounds are presented in the graph (PM3 from Biolog Plates). Pyruvate was used to constrain the model as the unique carbon source during the nitrogen validation. C) statistical parameters and Matthews correlation coefficient of the predictions for the carbon (first column) and nitrogen sources (second column) using the Biolog Plates information; true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Statistical analysis of the estimations for 38 carbon sources under nitrogen fixation and H_2 consumption conditions (third column). The data used during the simulations were obtained from Wong and Maier, 1985.

3.2.3 Carbon and nitrogen partitioning analysis

The average metabolic fluxes of carbon and nitrogen elements were determined for all carbon sources with growth rate values greater than 0.001 from Biolog plates data (61 total). The complete dataset with the experimental data is summarized in Table S2. The metabolic fluxes for both elements were grouped in 43 specific subsystems to identify the activity of the main pathways for all the experimental conditions. The highest average carbon fluxes (see Methods) were obtained from energy metabolism (82 mmol/gDWh), oxidative phosphorylation (67.5 mmol/gDWh), biomass and maintenance functions (36 mmol/gDWh), and TCA cycle (25 mmol/gDWh) for both nitrogen (N_2 and NH_4) conditions (disregarding transport fluxes). Similar tendencies were observed for the internal nitrogen flux distributions across the subsystems for diazotrophic and non-diazotrophic growth. However, the average flux distributions of carbon and nitrogen decline under diazotrophic conditions. Regarding global metabolic fluxes, the global carbon flux drops 4.9% and the nitrogen global flux value decreases 5.5%. The pathways with higher variation between carbon and nitrogen fluxes in diazotrophic and non-diazotrophic conditions were riboflavin metabolism (45%), glycine, serine and threonine metabolism (22%), alanine, aspartate and glutamate metabolism (26%), and nucleotide synthesis (10%). These subsystems are well-known for containing metabolites with high nitrogen content. The declinein the carbon and nitrogen global flux values of these pathways can be related to the low available nitrogen under diazotroph conditions due to the high energy cost of nitrogen fixation.

The grouped average fluxes of carbon and nitrogen elements per subsystem calculated for diazotrophic and non-diazotrophic conditions were compared through a linear correlation analysis to determine how the subsystem flux values behave under both nitrogen (N_2 and NH_4) conditions. Fig. 4 presents the linear correlations values between diazotrophic and non-diazotrophic conditions of the active subsystems. For carbon and nitrogen partitioning analysis, the highest correlation coefficients (Pearson correlation >0.95, p-value $<1x10^{-30}$) were observed in all the amino acids pathways, biomass and maintenance functions, energy metabolism, carbohydrate metabolism, and some subsystems related to lipid metabolism (lipopolysaccharide biosynthesis and

glycerophospholipid metabolism), demonstrating similar metabolic flux distributions through these specific subsystems when comparing diazotrophic and non-diazotrophic conditions. However, weak correlation values were obtained in riboflavin metabolism (r=0.66), oxidative phosphorylation (r=0.32 in carbon fluxes and r=0.37 in nitrogen distributions) and transport to the inner membrane (r=0.12 and r=0.31, respectively), showing an average flux distribution decrease of 45% for these specific subsystems when comparing carbon and nitrogen activity in diazotrophic against non-diazotrophic conditions. Additionally, strong correlation values were observed in specific subsystem clusters for both nitrogen conditions.

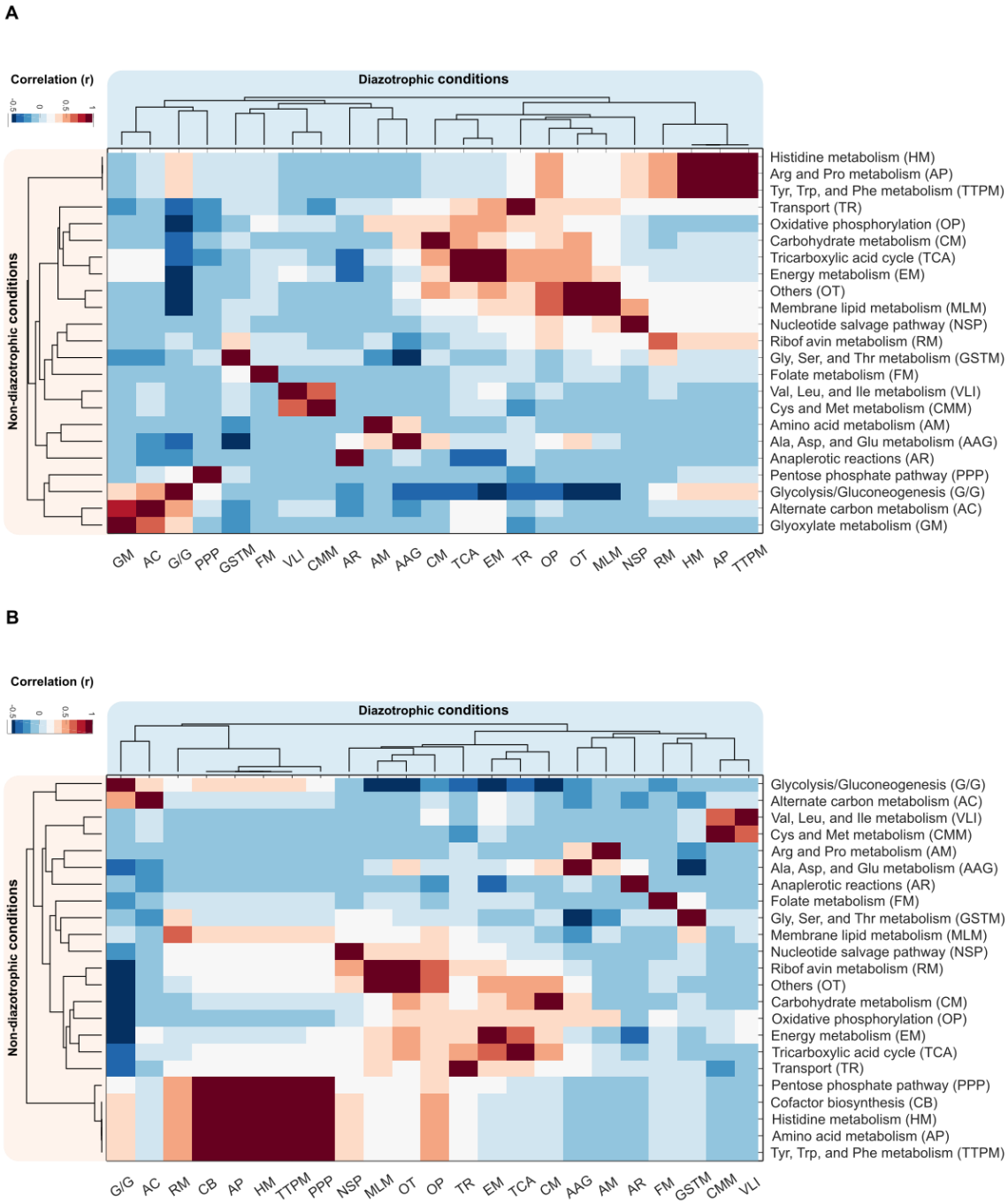


Fig. 4. Carbon and nitrogen partitioning distribution from Biolog plates experimental data. A) Average carbon flux correlation coefficients through all subsystems in diazotrophic and non-diazotrophic conditions; B) average nitrogen flux correlation coefficients through all subsystems in diazotrophic and non-diazotrophic conditions.

3.3 Alginate and PHB production estimated through *in-silico* experiments

A. vinelandii DJ contains specific mechanisms to produce and secrete alginate into the extracellular space. In *i*DT1278 we manually curated six specific reactions related to alginate biosynthesis including 16 different genes. One of the most important reactions involved in this pathway is the alginate epimerase (EC 5.1.3.37), which encompass a complex protein system to synthetize the alginate polymer (Pacheco-Leyva et al., 2016).

We evaluated the model accuracy to growth and alginate production using four carbon sources under diazotrophic and non-diazotrophic conditions. Experimental data retrieved from the literature (Revin et al., 2018) was only available for growth with ammonium. Simulations were confirmed to accurately predict (true positive predictions) alginate production rates with three carbon sources (glucose, mannitol, and sucrose). For example, the predicted growth using glucose (uptake rate of 0.33 mmol/gDWh) as the sole carbon source was 0.1478 h^{-1} and an alginate production rate of 0.25 mmol/gDWh, agreeing with the experimental data (growth rate of 0.152±0.012 h^{-1} and alginate production rate of 0.265±0.018 mmol/gDWh). Lactose was the only compound with a mismatch between reported values and *in-silico* outputs (false negative). The disagreement appears to be associated to the lack of information in the literature and bioinformatic databases about the essential enzymes in the metabolism of this disaccharide in *A. vinelandii* DJ.

We compared the growth rates determined under diazotrophic and non-diazotrophic conditions for alginate production. The simulation analysis showed a significant increase in the growth rates (close to 28%) when the *A. vinelandii* consumes ammonium as a nitrogen source instead of molecular nitrogen. Additionally, a second comparison of alginate production rates between ammonium and molecular nitrogen exhibited the same increasing tendency when ammonium is used as a unique nitrogen source instead of molecular nitrogen (around 27%).

While alginate is transported to the extracellular space, PHB is intracellularly stored (Yoneyama et al., 2015; Zúñiga et al., 2011). To simulate PHB accumulation in *A. vinelandii* we incorporated a sink reaction to the model *i*DT1278. Simulated flux distributions about PHB production were validated using fluxomic data retrieved from Wu et al. (2019). The metabolic fluxes of the reactions involved in the PHB synthesis and related pathways (glycolysis, pentose phosphate pathway, the Entner-Deundoroff pathway, and the TCA cycle) were calculated through FBA for diazotrophic and non-diazotrophic conditions. The simulation results were compared with the experimental measured fluxes (Wu et al., 2019) and the percent error was estimated by reaction (Fig. 5). A general agreement in the reaction fluxes was observed under both nitrogen (N_2 and NH_4) conditions. A total of 16 out of 19 reaction flux estimations presented a global accuracy above 90% for diazotrophic and non-diazotrophic conditions. Disagreements were detected in three specific reactions: the transketolase (TKT2), aconitase

(ACONTa/ACONTb) and isocitrate dehydrogenase (ICDHyr) in which the percent errors were above 20%. According to the predicted and experimental results, a higher PHB production is obtained under non-diazotroph conditions due to higher energy cost for nitrogen fixation. Additionally, most of the carbon coming from the glucose is metabolized through the Entner-Doudoroff (ED) pathway, generating less energy than glycolysis, but producing pyruvate (PHB precursor) in fewer steps. The available pyruvate generated though the ED pathway is mostly used in the TCA cycle and the PHB synthesis (Castillo et al., 2013; Wu et al., 2019), allowing the growth of the microorganism and the production of this biopolymer.

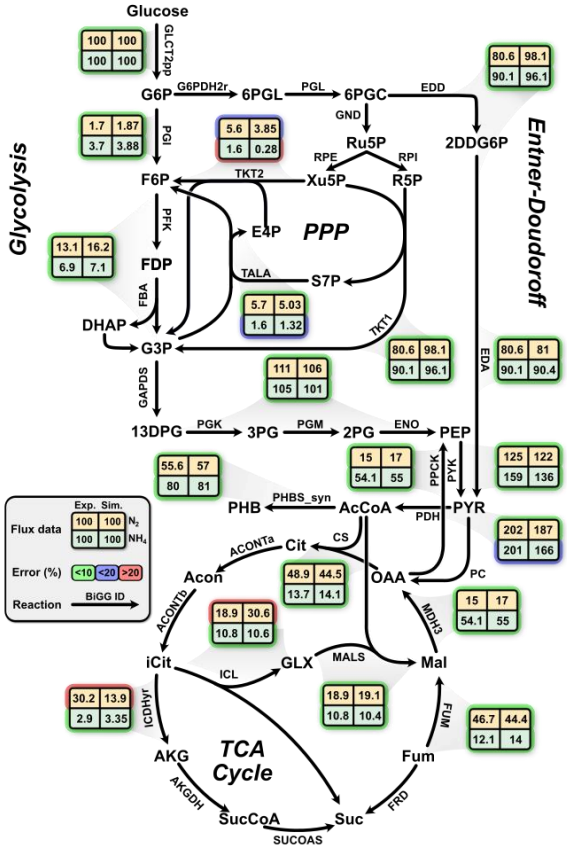


Fig. 5. Metabolic flux map distribution of A. vinelandii under diazotrophic and non-diazotrophic conditions. The map displays major metabolic pathways involved in the PHB production. Values of metabolic flux are normalized to a glucose uptake rate of 100. Predicted metabolic fluxes were compared against fluxomic data determined by Wu et al. in 2019. The reactions were labeled according to their percent error (green, blue and red) and nitrogen source (green and yellow). Abbreviations: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FDP, fructose-1,6-bisphosphate; G3P, 3-phosphoglycerate; DHAP, dihydroxyacetone phosphate; 13DPG, 3-phosphoglyceroil phosphate; 2PG, 2-

phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; Cit, citrate; Acon, aconitate; iCit, isocitrate; AKG, a-ketoglutarate; sucCoA, succinyl coenzyme A; Suc, succinate; Fum, fumarate; Mal, malate; OAA, oxaloacetate; GLX, glyoxylate; PHB, polyhydroxybutyrate; 6PGL, 6-phospho-glucono-1,5-lactone; 6PGC, 6-phospho-gluconate; 2DDG6P, 2-Dehydro-3-deoxygluconate 6-phosphate; Ru5P, ribulose-5-phosphate; R5P, ribose-5-phosphate; S7P, sedoheptulose-7-phosphate; X5P, xylulose-5-phosphate; E4P, erythrose-4-phosphate.

4 Discussion

4.1 Model Reconstruction

Here we have created the first comprehensive genome-scale metabolic model for *A. vinelandii* DJ (*i*DT1278) that focuses on nitrogen assimilation, nitrogen fixation, as well as on alginate and PHB production. The model consists of 1278 genes involved in 2469 reactions. Compared to the first microorganism template used in the present work (Escherichia coli str. K-12 substr. MG1655, *i*ML1515), the percentage of metabolic genes per genome decreases from 31% (*i*ML1515) to 26% (*i*DT1278). However, a higher percentage of metabolic genes (from 26% of *A. vinelandii* to 12% of the average genes in the photosynthetic models) was observed when *i*DT1278 was compared to three photosynthetic organisms in the BiGG database : *C. vulgaris* UTEX 395, *i*CZ843 (Zuñiga et al., 2016), *Synechocystis* sp. PCC 6803, *i*JN678 (Nogales et al., 2012) and *C. reinhardtii*, *i*CR1080 (Chang et al., 2011). The M-model is accurate to 89% for all the carbon sources and 91% for nitrogen sources. The model was validated using a wide variety of carbon (159 compounds) and nitrogen (75 metabolites) sources. iDT1278 shown a significant higher accuracy (27% upper) in the predictions compared to the CarveMe model simulations. Additionally, *i*DT1278 predicted accurately the growth ratio and production values of alginate and PHB production under diazotrophic and non-diazotrophic conditions. To our knowledge, this is the first M-model at genome-scale capable to simulate several carbon and nitrogen conditions (close to 250 conditions) with a high precision even when comparing internal metabolic fluxes.

4.2 Model Validation

4.2.1 Nitrogen Fixation

*i*DT1278 accurately predicts the growth of *A. vinelandii* using different carbon sources under diazotrophic and non-diazotrophic conditions. The model contains all required reactions and constraints to successfully simulate the BOF representing the growth of the organism. Model predictions have been confirmed by experimental validation using Biolog plates (PM1, PM2 and PM3). With this information we elucidated the preferred

mechanism used by the *A. vinelandii* DJ to fix nitrogen while growing with different carbon sources. The N_2 uptake depends on the ammonium concentration and metal cofactor concentrations required for the nitrogenases (García et al., 2018; Inomura et al., 2018). In our model, a difference in growth rate can be observed under diazotrophic and non-diazotrophic conditions. The growth rate under different carbon sources is higher during NH_4 assimilation than during N_2 fixation. However, the growth rates are quite similar for growth with amino acids under diazotrophic and non-diazotrophic conditions. This can be explained by the fact that amino acids release ammonium when metabolized which then becomes readily available to the organism. N_2 fixation on the other hand is an ATP-dependent process and the organism must employ more energy to convert the N_2 into ammonium.

Growth rate values decrease considerably when *A. vinelandii* DJ grows diazotrophically but nitrogen and carbon flux distributions per subsystem (e.g. amino acid, lipid, and carbohydrate metabolism) behave very similar in both nitrogen conditions. These flux distribution correlations suggest that the major discrepancies in the metabolism of *A. vinelandii* fixing nitrogen are related to the cofactor and vitamin pathways (riboflavin metabolism) and oxidative phosphorylation (energy generation). According to the model predictions (*i*DT1278), the vitamin, cofactor and oxidative phosphorylation pathways in *A. vinelandii* are involved in the generation of the precursors for the BOF; when ammonium is used as a nitrogen source, the carbon and nitrogen fluxes in these pathways are significantly higher (30%) than the fluxes estimated using molecular nitrogen as nitrogen source.

4.4.3 Alginate production

Alginate represents an important exopolysaccharide for *A. vinelandii* and is synthetized to reduce oxygen availability and thus increase nitrogenase activity for enhanced nitrogen fixation (Galindo et al., 2007; Nosrati et al., 2012). Additionally, this polymer has industrial relevance to multiple fields such as in pharmaceutical (Azevedo et al., 2014), biotechnological (Tomida et al., 2010), and food industry applications (Kuda et al., 1998). Therefore, elucidating the mechanism of alginate production could potentially provide insights for increasing production of this valuable biopolymer. *i*DT1278 accurately predicts three out of four carbon sources capable of producing alginate and the accompanying nitrogen sources to maximize biosynthesis of this polymer. The model also successfully predicts the decline in the growth and alginate production when molecular nitrogen is used as a nitrogen source. Alginate metabolism has been studied using genetic and regulation approaches to explain the synthesis of this valuable biopolymer, since most of the genes involved in this metabolic pathway are regulated by the presence of oxygen (Ertesvåg et al., 1995; Lloret et al., 1996; Núñez et al., 2000). In the present work we show that alginate production can also be explained based exclusively on metabolic requirements using mathematical and metabolic representations.

4.2.3 PHB production

PHB, like alginate, is synthesized by *A. vinelandii* to reduce oxygen availability and promote nitrogen fixation. *i*DT1278 contains all the genes and specific reactions involved in the production of PHB. The polymer is a high value product used in the production of biodegradable plastics and other environmental friendly polymers (Galindo et al., 2007). *i*DT1278 accurately predicts most of the metabolic flux values in the reactions (85% of the reactions) involved in PHB synthesis and pathways related to the generation of PHB precursors and energy metabolism (glycolysis, pentose phosphate pathway, the Entner-Doudoroff pathway, and the TCA cycle) for diazotrophic and non-diazotrophic conditions. The model is capable of accurately simulating all the variations presented in the flux distribution values for every reaction involved in the PHB synthesis under different conditions (ammonium or molecular nitrogen as unique nitrogen sources) using only metabolic data (stoichiometry and biochemistry data).

As with alginate production, PHB metabolism in *A. vinelandii* has been well studied using exclusively genetic and regulation perspectives since PHB accumulation mainly occurs during oxygen limitations (Castillo et al., 2013; Vijayendra et al., 2007); however, PHB metabolic distribution and biosynthesis can also be described using metabolic data according to the predictions performed by *i*DT1278. Our model could as a result be deployed to potentially optimize PHB production in *A. vinelandii.*

4.3 Network properties

Most of the False Negative (6) and False Positive (6) predictions from the model related to carbohydrates and amino acids substrates could not be confirmed to belong to a pathway or to an enzyme that converts these carbon sources to internal metabolites present in the model either from literature information or available genomic and metabolomic databases. Moreover, when using amino acids as growth substrates, the model predicts very similar growth rate values for diazotrophic and non-diazotrophic conditions since some amino acids can release ammonium further supporting growth. These discrepancies between experimental evidence and model predictions could be related to the limited information available concerning specific metabolic mechanisms for metabolizing carbon sources as carbohydrates, lipids and amino acids in *A. vinelandii* or processes related to genetic, transcriptional and regulation. Indeed, while nitrogen metabolism is tightly regulated in *A. vinelandii* (Hamilton et al., 2011; Toukdarian and Kennedy, 1986), transcriptional and translational regulation is not currently part of this M-model but could in part be recapitulated by including all macromolecular synthesis as part of a subsequent ME-models (Lerman et al., 2012; Liu et al., 2019, 2014; O'Brien et al., 2013). Furthermore, nitrogenase activity and activity of other nitrogen-associated pathways are also regulated at the enzyme level (Pacheco-Leyva et al., 2016), potentially also contributing further to the lack of agreement between model predictions and experimental data.

4.4 Future Aspects

Insight into the metabolic processes using the genome-scale model *i*DT1278 could benefit low-cost media optimization for *A. vinelandii* and to further improve production processes as PHB, alginate and other biopolymers with high industrial value. Although this environmentally important model organism has been isolated and studied for more than 100 years, its physiology and some metabolic pathways have yet to be fully understood. We believe the current model provides a valuable step along the path towards better characterization of this important microbe. Unraveling this knowledge for *A. vinelandii* could improve its use in industrial applications (Noar and Bruno-Bárcena, 2018), and the systems biology approaches presented here may provide a tool to help in achieving that goal.

Author contributions

CZ, AZ, and KZ conceived the study. DT and CZ reconstructed the model and performed all simulations. DT and AP performed manual curation. JDT performed growth experiments. CZ, DT, AP, JDT, and MJB discussed the results. DT, AP, CZ, and KZ wrote the manuscript with input from all co-authors.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 1332344 and CBET-1804187, and the U.S. Department of Energy (DOE), Office of Science, Office of Biological & Environmental Research under Awards DE-SC0012658 and DE-SC0019388. D.T. was in part supported by Mexican National Research Council, CONACYT, fellowship No. 932962.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We acknowledge Manish Kumar and Erick Armingol (UC San Diego) for providing input and assistance throughout all stages of this work.

Research Data for this article

iDT1278 is available in .mat and .xml format as well as the carbon and nitrogen partitioning analysis in the repository from Zengler Lab: https://github.com/cristalzucsd/AvinelandiiDJ

References

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., Nielsen, J., 2013. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. PLoS Comput. Biol. 9, e1002980. https://doi.org/10.1371/journal.pcbi.1002980

Ambrosio, R., Ortiz-Marquez, J.C.F., Curatti, L., 2017. Metabolic engineering of a diazotrophic bacterium improves ammonium release and biofertilization of plants and microalgae. Metab. Eng. 40, 59–68. https://doi.org/10.1016/j.ymben.2017.01.002

Azevedo, M.A., Bourbon, A.I., Vicente, A.A., Cerqueira, M.A., 2014. Alginate/chitosan nanoparticles for encapsulation and controlled release of vitamin B2. Int. J. Biol. Macromol. 71, 141–146. https://doi.org/10.1016/j.ijbiomac.2014.05.036

Castillo, T., Heinzle, E., Peifer, S., Schneider, K., Peña M, C.F., 2013. Oxygen supply strongly influences metabolic fluxes, the production of poly(3-hydroxybutyrate) and alginate, and the degree of acetylation of alginate in Azotobacter vinelandii. Process Biochem. 48, 995–1003. https://doi.org/10.1016/j.procbio.2013.04.014

Chang, R.L., Ghamsari, L., Manichaikul, A., Hom, E.F.Y., Balaji, S., Fu, W., Shen, Y., Hao, T., Palsson, B.Ø., Salehi-Ashtiani, K., Papin, J.A., 2011. Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. Mol. Syst. Biol. 7, 518. https://doi.org/10.1038/msb.2011.52

Clementi, F., 1997. Alginate Production by Azotobacter Vinelandii. Crit. Rev. Biotechnol. 17, 327–361. https://doi.org/10.3109/07388559709146618

Cleveland, C.C., Townsend, A.R., Schimel, D.S., Fisher, H., Howarth, R.W., Hedin, L.O., Perakis, S.S., Latty, E.F., Von Fischer, J.C., Elseroad, A., Wasson, M.F., 1999. Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. Global Biogeochem. Cycles 13, 623–645. https://doi.org/10.1029/1999GB900014

Díaz-Barrera, A., Andler, R., Martínez, I., Peña, C., 2016. Poly-3-hydroxybutyrate production by Azotobacter vinelandii strains in batch cultures at different oxygen transfer rates. J. Chem. Technol. Biotechnol. 91, 1063–1071. https://doi.org/10.1002/jctb.4684

Ertesvåg, H., Høidal, H.K., Hals, I.K., Rian, A., Doseth, B., Valla, S., 1995. A family of modular type mannuronan C-5-epimerase genes controls alginate structure in Azotobacter vinelandii. Mol. Microbiol. 16, 719–31. https://doi.org/10.1111/j.1365-2958.1995.tb02433.x

Feist, A.M., Nagarajan, H., Rotaru, A.-E., Tremblay, P.-L., Zhang, T., Nevin, K.P., Lovley, D.R., Zengler, K., 2014. Constraint-based modeling of carbon fixation and the energetics of electron transfer in Geobacter metallireducens. PLoS Comput. Biol. 10, e1003575. https://doi.org/10.1371/journal.pcbi.1003575

Feist, A.M., Scholten, J.C.M., Palsson, B.Ø., Brockman, F.J., Ideker, T., 2006. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina

barkeri. Mol. Syst. Biol. 2. https://doi.org/10.1038/msb4100046
Galindo, E., Peña, C., Núñez, C., Segura, D., Espín, G., 2007. Molecular and bioengineering strategies to improve alginate and polydydroxyalkanoate production by Azotobacter vinelandii. Microb. Cell Fact. 6, 1–16. https://doi.org/10.1186/1475-2859-6-7
García, A., Ferrer, P., Albiol, J., Castillo, T., Segura, D., Peña, C., 2018. Metabolic flux analysis and the NAD(P)H/NAD(P)+ ratios in chemostat cultures of Azotobacter vinelandii. Microb. Cell Fact. 17, 10. https://doi.org/10.1186/s12934-018-0860-8
Gyurján, I., Korányi, P., Preininger, É., Varga, S.S., Paless, G., 1995. Artificial Plant-Azotobacter Symbiosis for Atmospheric Nitrogen Fixation, in: Azospirillum VI and Related Microorganisms. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 401–413. https://doi.org/10.1007/978-3-642-79906-8_46
Hamilton, T.L., Ludwig, M., Dixon, R., Boyd, E.S., Dos Santos, P.C., Setubal, J.C., Bryant, D.A., Dean, D.R., Peters, J.W., 2011. Transcriptional Profiling of Nitrogen Fixation in Azotobacter vinelandii. J. Bacteriol. 193, 4477–4486. https://doi.org/10.1128/JB.05099-11
Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Žagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L. V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H.S., Bryant, W.A., Aragón Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B.Ø., Thiele, I., Fleming, R.M.T., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. 14, 639–702. https://doi.org/10.1038/s41596-018-0098-2
Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977–982. https://doi.org/10.1038/nbt.1672
Howard, J.B., Rees, D.C., 1996. Structural Basis of Biological Nitrogen Fixation. Chem. Rev. 96, 2965–2982. https://doi.org/10.1021/cr9500545
Inomura, K., Bragg, J., Riemann, L., Follows, M.J., 2018. A quantitative model of nitrogen fixation in the presence of ammonium. PLoS One 13, e0208282. https://doi.org/10.1371/journal.pone.0208282
King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 44, D515–D522. https://doi.org/10.1093/nar/gkv1049
Kuda, T., Goto, H., Yokoyama, M., Fujii, T., 1998. Effects of Dietary Concentration of Laminaran and Depolymerised Alginate on Rat Cecal Microflora and Plasma

Lipids. Fish. Sci. 64, 589–593. https://doi.org/10.2331/fishsci.64.589
Lerman, J.A., Hyduke, D.R., Latif, H., Portnoy, V.A., Lewis, N.E., Orth, J.D., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Zengler, K., Palsson, B.O., 2012. In silico method for modelling metabolism and gene product expression at genome scale. Nat. Commun. 3, 929. https://doi.org/10.1038/ncomms1928
Liao, Y.-C., Huang, T.-W., Chen, F.-C., Charusanti, P., Hong, J.S.J., Chang, H.-Y., Tsai, S.-F., Palsson, B.O., Hsiung, C.A., 2011. An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228. J. Bacteriol. 193, 1710–7. https://doi.org/10.1128/JB.01218-10
Liu, J.K., Lloyd, C., Al-Bassam, M.M., Ebrahim, A., Kim, J.-N., Olson, C., Aksenov, A., Dorrestein, P., Zengler, K., 2019. Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. PLoS Comput. Biol. 15, e1006848. https://doi.org/10.1371/journal.pcbi.1006848
Liu, J.K., O'Brien, E.J., Lerman, J.A., Zengler, K., Palsson, B.O., Feist, A.M., 2014. Reconstruction and modeling protein translocation and compartmentalization in Escherichia coli at the genome-scale. BMC Syst. Biol. 8, 110. https://doi.org/10.1186/s12918-014-0110-6
Lloret, L., Barreto, R., León, R., Moreno, S., MartÍnez-Salazar, J., Espín, G., Soberón-Chávez, G., 1996. Genetic analysis of the transcriptional arrangement of Azotobacter vinelandii alginate biosynthetic genes: identification of two independent promoters. Mol. Microbiol. 21, 449–457. https://doi.org/10.1111/j.1365-2958.1996.tb02554.x
Machado, D., Andrejev, S., Tramontano, M., Patil, K.R., 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553. https://doi.org/10.1093/nar/gky537
Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A.M., Palsson, B.O., 2017. iML1515, a knowledgebase that computes Escherichia coli traits. Nat. Biotechnol. 35, 904–908. https://doi.org/10.1038/nbt.3956
Nagai, S., Nishizawa, Y., Doin, P.A., Aiba, S., 1972. Energy difference in metabolism between glucose and acetate in chemostat culture of Azotobacter vinelandii. J. Gen. Appl. Microbiol. 18, 201–208. https://doi.org/10.2323/jgam.18.201
Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D.R., Ebrahim, A., Zengler, K., 2013. Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. Microb. Cell Fact. 12, 118. https://doi.org/10.1186/1475-2859-12-118
Noar, J., Loveless, T., Navarro-Herrero, J.L., Olson, J.W., Bruno-Bárcena, J.M., 2015. Aerobic Hydrogen Production via Nitrogenase in Azotobacter vinelandii CA6. Appl. Environ. Microbiol. 81, 4507–4516. https://doi.org/10.1128/AEM.00679-15
Noar, J.D., Bruno-Bárcena, J.M., 2018. Azotobacter vinelandii: the source of 100 years of discoveries and many more to come. Microbiology 164, 421–436.

https://doi.org/10.1099/mic.0.000643
Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., Thiele, I., 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc. Natl. Acad. Sci. U. S. A. 109, 2678–83. https://doi.org/10.1073/pnas.1117907109
Nosrati, R., Owlia, P., Saderi, H., Olamaee, M., Rasooli, I., Akhavian, T.A., 2012. Correlation between nitrogen fixation rate and alginate productivity of an indigenous Azotobacter vinelandii from Iran. Iran. J. Microbiol. 4, 153–9. https://doi.org/23066492
Núñez, C., León, R., Guzmán, J., Espín, G., Soberón-Chávez, G., 2000. Role of Azotobacter vinelandii mucA and mucC gene products in alginate production. J. Bacteriol. 182, 6550–6. https://doi.org/10.1128/jb.182.23.6550-6556.2000
O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, 693. https://doi.org/10.1038/msb.2013.52
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248. https://doi.org/10.1038/nbt.1614
Pacheco-Leyva, I., Guevara Pezoa, F., Díaz-Barrera, A., 2016. Alginate Biosynthesis in Azotobacter vinelandii : Overview of Molecular Mechanisms in Connection with the Oxygen Availability. Int. J. Polym. Sci. 2016, 1–12. https://doi.org/10.1155/2016/2062360
Quiroz-Rocha, E., Moreno, R., Hernández-Ortíz, A., Fragoso-Jiménez, J.C., Muriel-Millán, L.F., Guzmán, J., Espín, G., Rojo, F., Núñez, C., 2017. Glucose uptake in Azotobacter vinelandii occurs through a GluP transporter that is under the control of the CbrA/CbrB and Hfq-Crc systems. Sci. Rep. 7, 858. https://doi.org/10.1038/s41598-017-00980-5
Remminghorst, U., Rehm, B.H.A., 2006. Bacterial alginates: from biosynthesis to applications. Biotechnol. Lett. 28, 1701–1712. https://doi.org/10.1007/s10529-006-9156-x
Revin, V.V., Kostina, E.G., Revina, N.V., Shutova, V.V., 2018. Effect of Nutrient Sources on the Alginate Accumulation in the Culture Liquid of Azotobacter vinelandii D-05 and Obtaining Biocomposite Materials. Brazilian Arch. Biol. Technol. 61. https://doi.org/10.1590/1678-4324-2018160406
Sahoo, R.K., Ansari, M.W., Dangar, T.K., Mohanty, S., Tuteja, N., 2014. Phenotypic and molecular characterisation of efficient nitrogen-fixing Azotobacter strains from rice fields for crop improvement. Protoplasma 251, 511–523. https://doi.org/10.1007/s00709-013-0547-2
Setubal, J.C., Almeida, N.F., 2015. The Azotobacter vinelandii Genome: An Update, in: Biological Nitrogen Fixation. John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 225–234. https://doi.org/10.1002/9781119053095.ch22
Setubal, J.C., dos Santos, P., Goldman, B.S., Ertesvag, H., Espin, G., Rubio, L.M., Valla, S., Almeida, N.F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z.,

Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J.A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T.J., Latreille, P., Ligon, L.S., Lu, J., Maerk, M., Miller, N.M., Norton, S., O'Carroll, I.P., Paulsen, I., Raulfs, E.C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S.L., Studholme, D.J., Sun, J., Viana, C.J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D.R., Dixon, R., Wood, D., 2009. Genome Sequence of Azotobacter vinelandii, an Obligate Aerobe Specialized To Support Diverse Anaerobic Metabolic Processes. J. Bacteriol. 191, 4534–4545. https://doi.org/10.1128/JB.00504-09

Shawky, B.T., Ghali, Y., Ahmed, F.A., Kahil, T., 1987. Biochemical studies on the effect of various carbon sources on growth, nitrogen fixtion, and main cellular constituents of azotobacter vinelandii, strain IV grown under various cultivation conditions. Acta Biotechnol. 7, 361–369. https://doi.org/10.1002/abio.370070417

Sippel, D., Schlesier, J., Rohde, M., Trncik, C., Decamps, L., Djurdjevic, I., Spatzal, T., Andrade, S.L.A., Einsle, O., 2017. Production and isolation of vanadium nitrogenase from Azotobacter vinelandii by molybdenum depletion. JBIC J. Biol. Inorg. Chem. 22, 161–168. https://doi.org/10.1007/s00775-016-1423-2

Stevenson, L.H., Socolofsky, M.D., 1966. Cyst formation and poly-beta-hydroxybutyric acid accumulation in Azotobacter. J. Bacteriol. 91, 304–10.

Tomida, H., Yasufuku, T., Fujii, T., Kondo, Y., Kai, T., Anraku, M., 2010. Polysaccharides as potential antioxidative compounds for extended-release matrix tablets. Carbohydr. Res. 345, 82–86. https://doi.org/10.1016/j.carres.2009.10.015

Toukdarian, A., Kennedy, C., 1986. Regulation of nitrogen metabolism in Azotobacter vinelandii: isolation of ntr and glnA genes and construction of ntr mutants. EMBO J. 5, 399–407.

Vijayendra, S.V.N., Rastogi, N.K., Shamala, T.R., Anil Kumar, P.K., Kshama, L., Joshi, G.J., 2007. Optimization of polyhydroxybutyrate production by Bacillus sp. CFR 256 with corn steep liquor as a nitrogen source. Indian J. Microbiol. 47, 170–175. https://doi.org/10.1007/s12088-007-0033-7

Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., Gerdes, S., Henry, C.S., Kenyon, R.W., Machi, D., Mao, C., Nordberg, E.K., Olsen, G.J., Murphy-Olson, D.E., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H., Stevens, R.L., 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 45, D535–D542. https://doi.org/10.1093/nar/gkw1017

Wong, T.Y., 1988. Effects of Mannose on the Growth of N(2)-Fixing Azotobacter vinelandii. Appl. Environ. Microbiol. 54, 473–5.

Wong, T.Y., Maier, R.J., 1985. H2-dependent mixotrophic growth of N2-fixing Azotobacter vinelandii. J. Bacteriol. 163, 528–33.

Wu, C., Herold, R.A., Knoshaug, E.P., Wang, B., Xiong, W., Laurens, L.M.L., 2019. Fluxomic Analysis Reveals Central Carbon Metabolism Adaptation for Diazotroph Azotobacter vinelandii Ammonium Excretion. Sci. Rep. 9, 13209.

https://doi.org/10.1038/s41598-019-49717-6
Yoneyama, F., Yamamoto, M., Hashimoto, W., Murata, K., 2015. Production of polyhydroxybutyrate and alginate from glycerol by Azotobacter vinelandii under nitrogen-free conditions. Bioengineered 6, 209–217. https://doi.org/10.1080/21655979.2015.1040209
Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D.C., McConnell, B.O., Long, C.P., Knoshaug, E.P., Guarnieri, M.T., Antoniewicz, M.R., Betenbaugh, M.J., Zengler, K., 2016. Genome-Scale Metabolic Model for the Green Alga Chlorella vulgaris UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiol. 172, 589–602. https://doi.org/10.1104/pp.16.00593
Zúñiga, C., Morales, M., Le Borgne, S., Revah, S., 2011. Production of poly-β-hydroxybutyrate (PHB) by Methylobacterium organophilum isolated from a methanotrophic consortium in a two-phase partition bioreactor. J. Hazard. Mater. 190, 876–882. https://doi.org/10.1016/j.jhazmat.2011.04.011

SEGUNDO ARTÍCULO DE INVESTIGACIÓN

CARTA DE ACEPTACIÓN

Dear Dr. Tec-Campos,

We are delighted to accept for publication in *npj Systems Biology and Applications* the manuscript "Linking metabolic phenotypes to pathogenic traits among '*Candidatus* Liberibacter asiaticus' and its hosts" by Karsten Zengler, Cristal Zuniga, Beth Peacock, Bo Liang, Greg McCollum, Sonia Irigoyen, Diego Tec-Campos, Clarisse Marotz, Nien-Chen Weng, Alejandro Zepeda, Georgios Vidalakis, Kranthi Mandadi, and James Borneman, on which you were a contributing author.

By choosing an open access journal for your research, you are ensuring that your article can be freely accessed by anyone, immediately on publication.

In addition, as part of the Springer Nature SharedIt initiative, a stable, shortened URL will be available to you on publication. As your article is open access, readers will be able to download and print the PDF and access the full-text HTML as usual. The link has additional benefits over uploading or emailing a static PDF of your article. As readers will be directed to your article on the journal website, all of the views of your article will be captured on your article metrics page on the journal website, enabling a fuller picture of the impact and reach of your research. You can also take advantage of annotation functionality, so you can add comments and annotations to your article before you share it, aiding collaboration.
As soon as your article is published, you can generate your shareable link by entering the DOI of your article is here: http://authors.springernature.com/share

Yours sincerely,

Nicolas Fanget,

PhD Managing Editor

Linking metabolic phenotypes to pathogenic traits among '*Candidatus* Liberibacter asiaticus' and its hosts

Cristal Zuñiga[1], Beth Peacock[2*], Bo Liang[1,3*], Greg McCollum[4], Sonia C. Irigoyen[5], Diego Tec-Campos[1,6], Clarisse Marotz[1], Nien-Chen Weng[1], Alejandro Zepeda[6], Georgios Vidalakis[2], 5 Kranthi K. Mandadi[5,7], James Borneman[2,#], Karsten Zengler[1,8,9,#]

1Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA

2Department of Microbiology and Plant Pathology, University of California, Riverside, 900 University Ave, Riverside, CA 92521

3State Key Laboratory of Bioreactor Engineering and Institute of Applied Chemistry, East China University of

Science and Technology, Shanghai, P.R. China

4USDA, ARS, US Horticultural Research Laboratory, 2001 S. Rock Road, Ft. Pierce, FL 34945

5Texas A&M AgriLife Research and Extension Center, Texas A&M University System, Weslaco, Texas, USA

6Facultad de Ingeniería Química, Universidad Autónoma de Yucatán, Campus de Ciencias Exactas e Ingenierías, Mérida 97203, Yucatán, México

7Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, USA

8Department of Bioengineering, University of California, San Diego, La Jolla CA 92093-0412, USA

9Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 920930403, USA

*Equal contribution

Correspondence to James Borneman (borneman@ucr.edu) or Karsten Zengler (kzengler@ucsd.edu).

Candidatus Liberibacter asiaticus (*C*Las) has been associated with Huanglongbing, a lethal vector-borne disease affecting citrus crops worldwide. While comparative genomics has provided preliminary insights into the metabolic capabilities of this uncultured microorganism, a comprehensive functional characterization is currently lacking. Here we reconstructed and manually curated genome-scale metabolic models for the six *C*Las strains A4, FL17, gxpsy, Ishi1, psy62, and YCPsy, in addition to a model of the closest related culturable microorganism, *L. crescens* BT-1. Predictions about nutrient requirements and changes in growth phenotypes of *C*Las were confirmed using *in vitro* hairy root-based assays, while the *L. crescens* BT-1 model was validated using cultivation assays. Host-dependent metabolic phenotypes were revealed using expression data obtained from *C*Las-infected citrus trees and from the *C*Las-harboring psyllid *Diaphorina citri* Kuwayama. These results identified conserved and unique metabolic traits as well as strain-specific interactions between *C*Las and its hosts, laying the foundation for the development of model-driven Huanglongbing management strategies.

Candidatus Liberibacter asiaticus (*C*Las) has been associated with Huanglongbing (HLB), or citrus greening, a devastating vector-borne disease causing millions of dollars of agricultural damages every year[1]. *C*Las species infect the phloem of some plants in the family *Rutaceae* (e.g. citrus, *Murraya paniculata*) and *Solanaceae* (e.g. potato)[2]. HLB causes poor vegetative growth, fruit drop, diminished fruit quality, and tree decline[3–7]. *C*Las infections have been documented across most citrus-producing areas in Asia, the Americas and Africa[5,8], and are projected to spread further[9,10]. *C*Las is naturally spread in citrus through a psyllid host, *Diaphorina citri* Kuwayama[10]. The basic HLB management scheme is based on the use of HLB-free nursery stock, inoculum reduction by removal of HLB-affected trees and insecticide treatments for control of psyllid populations[11]. In addition, various combinations of citrus rootstocks and interstocks[11], cocktails of antibiotics[12] and small molecule bacterial inhibitors[13], or brassinosteroids[14], as well as thermotherapy and nanoemulsion technology[15] have been deployed. However, none of these options has been proven to be very successful, economically viable, or environmentally sustainable, making HLB a major threat to the citrus industry worldwide.

Mathematical models have been critical in developing treatment options for infectious diseases[16,17]. These models could provide useful information for best practices to treat or prevent HLB. However, few detailed models of HLB currently exist[18]. To identify novel alternatives for combatting HLB, detailed knowledge about the metabolic dependencies and capabilities of the pathogen (*C*Las) is required. While *C*Las was identified as the likely infectious agent responsible for HLB in 1994 using molecular methods[19], *C*Las has never been consistently cultivated axenically *in vitro*, limiting our ability to functionally characterize this pathogen. On the other hand, *Liberibacter crescens*, the closest culturable relative to *C*Las, was isolated and cultured *in vitro* from the phloem sap of defoliating mountain papaya in Puerto Rico[20,21]. Advances in metagenomic sequencing have enabled the assembly of genomes from several *C*Las strains obtained from HLB-infected citrus[22]. Genome sequences are the primary input used to reconstruct genome-scale metabolic models. These models have been successfully validated as a systems biology framework and deployed for a variety of uses. For example, models have been utilized for elucidating fundamental metabolic processes[23–25], optimizing culture media and growth conditions[26,27], and have been essential for metabolic engineering efforts[28]. These metabolic models are genome-scale knowledge databases, which contain manually curated annotation related to gene-protein-reaction associations for all possible metabolic reactions inside a cell. Reconstructed models have been used to understand and channel the metabolism of different pathogenic and non-pathogenic microorganisms[17,29] as well as to contextualize metabolic states based on omics data[24]. Here we reconstructed genome-scale models for seven *Liberibacter* strains and evaluated their physiology and metabolic response. Furthermore, we used *in vivo* expression data to determine strain-specific interactions of *C*Las while hosted by the psyllid *Diaphorina citri* Kuwayama or *Citrus spp*.

Results

Genome characteristics and model properties

Metabolic models were reconstructed based on complete genomes of the *C*Las strains gxpsy, Ishi-1, psy62, and almost complete genomes obtained by shotgun sequencing of strains A4, FL17, and YCPsy. Strains A4, FL17, Ishi-1, psy62, and YCPsy were obtained from citrus, while sequences for gxpsy were obtained from the psyllid. Additionally, we reconstructed a comparative metabolic model using the complete genome sequence of the microorganism *L. crescens* BT-1 (BT-1), the closest culturable microorganism to *C*Las (Supplementary Fig. 1a). All genome sequences were obtained from the PATRIC database[30]. In total, seven genome-scale metabolic models were reconstructed and the genomic and metabolic content of each strain was compared. Fig. 1 details the main characteristics of the genomes and resulting models. The average number of annotated proteins was 1,185 across all *C*Las genomes and 1,422 for BT-1. We calculated the percent protein sequence identity among the seven *Liberibacter* strains. The *C*Las genomes had around 75% identity to *L. crescens* BT-1 (Fig. 1a) and over 98% identity to each other (Fig. 1b).

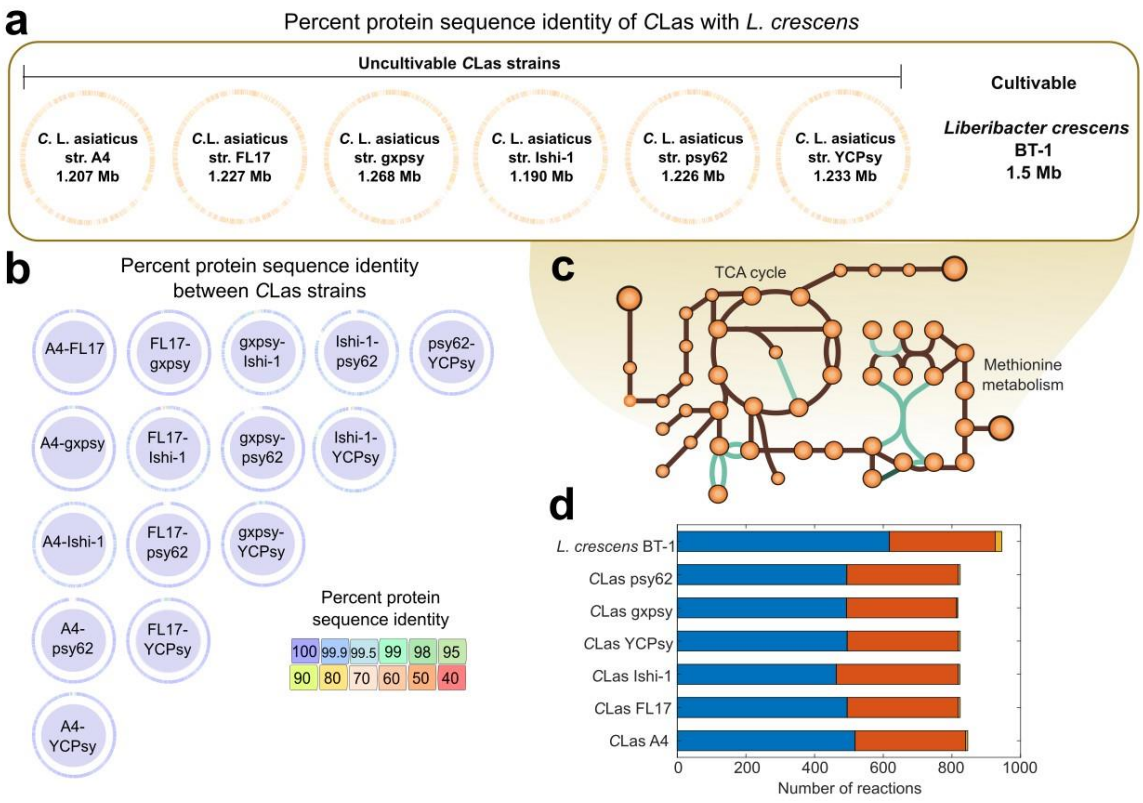


Fig. 1| Properties of genomes and constraint-based metabolic models. a, Percentage of protein sequence identity estimated using whole genome sequences of *Candidatus* Liberibacter asiaticus (*C*Las) strains and *L. crescens* BT-1. The identity between each *C*Las strain and BT-1 varied from 60-70%. **b**, Protein homology identity among *C*Las strains, which were over 99% identical. **c**, Example of pathway completeness and gap filling for the TCA cycle and methionine metabolism in the *C*Las models. Circles represent metabolites, green represents gap filled reactions, and brown represents gene associated reactions. **d**, Reactions involved in central metabolism including glycolysis/gluconeogenesis, the pentose phosphate pathway, and TCA cycle were categorized as pan reactions (highlighted in blue). The accessory reactions that include modeling reactions and other additional metabolic pathways (e.g. amino acid, carbohydrate, and lipid metabolism) are shown in red. Unique reactions in each model are shown in orange.

A comprehensive organization of all available data and information on *Liberibacter* strains is crucial for overcoming the devastating impact of HLB. To this end, we created metabolic models that are also referred to as computational knowledge-databases, which compile manually curated annotations for each *C*Las strain and *L. crescens* BT-1. After extensive curation of all models, a total of 1,751 protein sequences (417 unique), previously annotated as hypothetical, were associated with a metabolic reaction or

transport reaction in these models. A list of these reactions and their gene-protein-reaction associations is provided in Supplementary Table 1. Manual curation was followed by gap filling (Fig. 1c) and conversion of the reconstructions into mathematical models using the COBRA Toolbox[31]. The final model properties are shown in Table 1. The number of metabolites and reactions shared by all *C*Las models, defined as "Pancapabilities", were 445 and 605, respectively (Fig. 1d). Supplementary Fig. 2 shows a comparison of metabolites and reactions across the reconstructed models. The metabolic model of *L. crescens* BT-1 contained ~15% more metabolic reactions and metabolites than the *C*Las strains, hinting at a broader metabolic capability of *L. crescens*. Around 30% of reactions present only in BT-1 were associated with amino acid metabolism (e.g. methionine, lysine, glycine, serine and threonine metabolism). Another ~30% of those reactions were associated with the cell envelope (e.g. transporters and fatty acids). The rest of the reactions were divided among carbohydrate, glycan, and nucleotide metabolism (see Supplementary Table 2). We found that auxotrophies are *C*Las strain-specific, especially auxotrophies for L-proline, L-serine, and Larginine (Supplementary Fig. 2). All models predicted auxotrophies for vitamins (e.g. riboflavin, biotin, thiamin, choline) and steroids (e.g. pantothenate, taurine, L-carnitine, quinate) (Supplementary Fig. 2).

Validation of the *Liberibacter crescens* BT-1 model and analysis of *Candidatus* Liberibacter asiaticus models. Cruz-Munoz et al.[32] recently reported citrate as the preferred carbon and energy source for *L. crescens* BT-1. The authors reported the growth of *L. crescens* BT-1 on the complex media BM-7 by measuring optical density. They also tested various media compositions (M13, M14, and M15), and developed an optimized defined medium (M15, containing citrate) which improved the growth rate[32]. Using OD measurements, we determined the growth rates of *L. crescens* BT-1 while growing in BM-7, reaching a growth rate of 0.011±0.007 1/h, which results in a 63 h generation time. Defined media compositions, such as M13, M14, and M15 can also support growth of this strain, resulting in growth rates of 0.009±0.0006 1/h, 0.0081±0.0004 1/h, and 0.012±0.003 1/h respectively.

These growth rate phenotypes observed experimentally were reproduced *in silico* by the BT-1 model, obtaining growth rate predictions of 0.011, 0.009, 0.011, and 0.015 1/h for the culture media BG-7, M13, M14, and M15, respectively. Complete data is shown in Supplementary Table 3. Growth rate predictions using M15 media with varying citrate concentrations are shown in Fig. 2a and corroborate experimental results[32]. To determine the impact of amino acids and intermediaries of the TCA cycle on growth rate, we interrogated the model for metabolites most affecting the growth rate of *L. crescens*. We found that citrate, aspartate, and serine have an interwoven growth effect. Simulations performed while varying serine, aspartate, and citrate uptake rates are shown in Fig. 2b, highlighting a proportional increase among predicted growth rates and constrained uptake rates. Predicted trends were confirmed experimentally for citrate and serine individually and in combination (Fig. 2c), however increased aspartate additions reduced the growth of BT-1 experimentally.

*C*Las models were evaluated using different culture media compositions to unravel the most important metabolites contributing to *C*Las growth. The culture media compositions tested for *L. crescens* BT-1 were used as constraints to simulate *C*Las growth as shown in Supplementary Table 3. Simulations using single carbon sources (uptake rate of 15 mmol/gDWh) demonstrated that none of the *C*Las strains were able to grow, suggesting that co-metabolism with the host could play an important role for these bacteria. Modeling results showed that arginine, glucose, glutamate, glutamine, proline, ornithine, citrate, and alpha-ketoglutarate could each support the growth of *L. crescens* BT-1 individually. However, we found that *C*Las strains were highly dependent on co-metabolism to stimulate growth, necessitating the combination of two or more carbon sources at the same time (e.g. glucose and glycine, or aspartate, or serine, or succinate). The growth rate improvement due to the addition of single metabolites in the culture media containing glucose was evaluated for each metabolite (Supplementary Table 3). Co-metabolism simulations were performed by assessing the growth rate predictions while varying the uptake rates of serine, aspartate, and citrate in a continuous gradient from 0 to 45 mmol/gDW/h. Experimental results suggest that citrate is the main driver of *L. crescens* growth, followed by serine and aspartate (Fig. 2c).

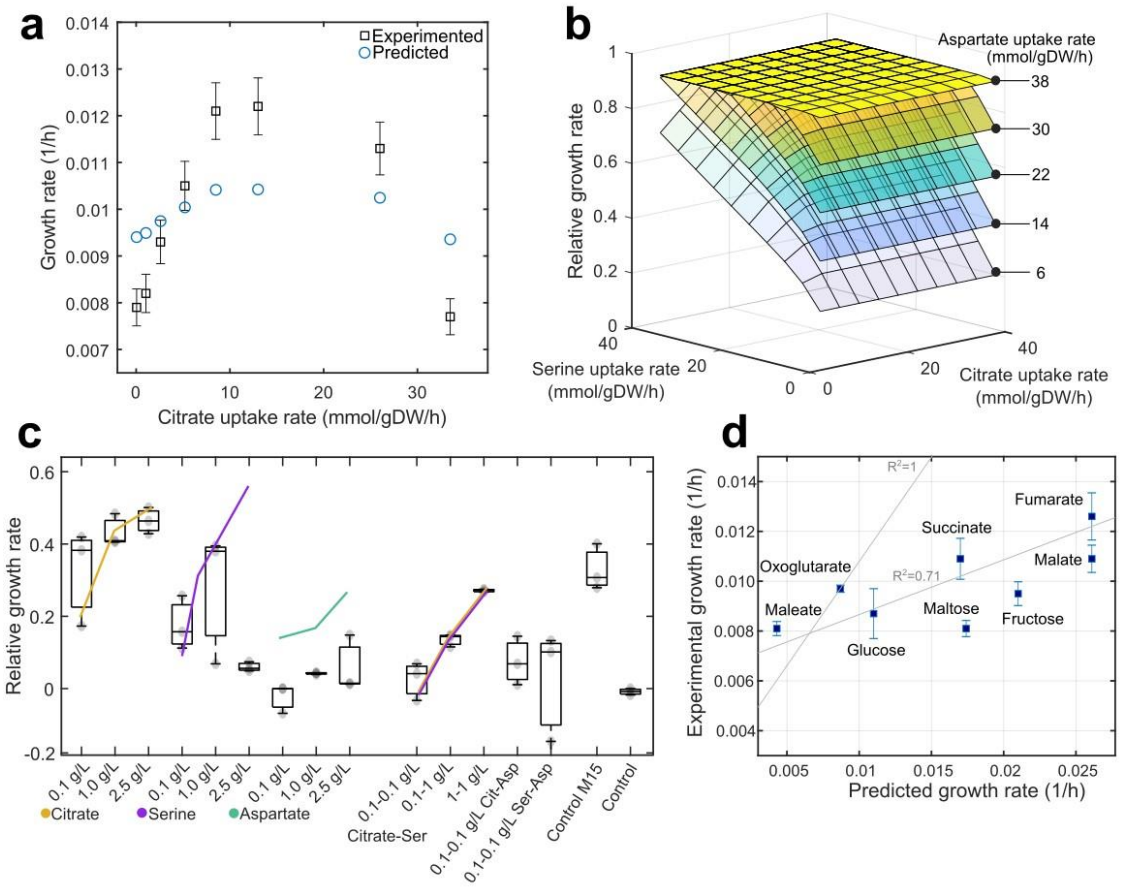


Fig. 2| Validation of the metabolic model of *Liberibacter crescens* BT-1. a, Simulated growth rate results using the culture medium M15 while varying the citrate concentration (blue circles). Squares with error bars represent the growth rate calculated using OD measurements. Citrate constraints used for simulation as well as experimental and predicted growth rates are given in Supplementary Table 3. **b**, Relative growth rate predictions across varying citrate, serine, and aspartate uptake rates. Growth rate in M15 was used as control **c**, Experimentally determined growth phenotypes. Box plots show *L. crescens* BT-1 growth under 14 treatments using several concentrations of citrate, serine, and aspartate and their interaction. Experiments were carried out using three independent replicates. Growth rates were calculated using 25 consecutive time points (sampled every 12 hours) measured over the course of 12 days. The culture medium M15 and the base media (without nutrients) were used as controls. **d**, Comparison among predicted and experimental growth rates of BT-1 while using the culture media M14 (Supplementary Table 3) in combination with glucose, fructose, fumarate, malate, maleate, maltose, oxoglutarate, and succinate as additional carbon sources. Predicted and observed growth phenotypes were correlated up to 71% (Pearson correlation, p-value<0.0041, n=3). A theoretical R^2=1 is also shown in the figure for comparison purposes.

The role of each metabolite in the metabolic model was evaluated across all models (Supplementary Fig. 3). We found that all *C*Las strains have similar simulated growth rates across all media compositions (i.e. BM-7, M13, M14, and M15). However, when the effect of each metabolite was analyzed independently, we found that media composition affected *C*Las growth rates differentially (Supplementary Fig. 3), suggesting strain-specific phenotypes. Also, the growth rates predicted for *L. crescens* BT-1 differed from those predicted for the six *C*Las strains (Supplementary Fig. 3).

The average metabolite connectivity was evaluated in the seven metabolic models. This connectivity network highlighted distinct differences among the models for the six *C*Las strains (Fig. 3a). For example, proline was connected (based on bubble size) to six reactions in the gxspy model, but to only five reactions in the other *C*Las models, while the opposite was observed for methionine and malate (Supplementary Table 4). The connectivity network enabled us to estimate the contribution of each metabolite to growth rate, and thereby calculate the essentiality fraction (relative change to growth rate). Fig. 3a summarizes modeling predictions of the individual carbon sources for the metabolic models in different culture media compositions, highlighting the differences in predicted growth rate by the presence of specific metabolites (shaded areas in Fig. 3a). Highly connected metabolites, such as glutamine, glutamate, serine, and alpha-ketoglutarate, had high essentiality fractions, around 0.5-0.7 in both the *C*Las and *L. crescens* BT-1 models. However, the only metabolite with the same essentiality fraction as glucose was glycine for *L. crescens* BT-1. The essentiality fraction of metabolites, such as succinate, fumarate, citrate, urea, tryptophan, arginine, and riboflavin, were dramatically lower for *L. crescens* compared to the *C*Las strains. The metabolites predicted to increase growth rate for each of the four media compositions are shown in a Venn diagram (Fig. 3b). Glucose and nine amino acids improved the simulated *C*Las growth rates in all culture media, whereas alphaketoglutarate (akg) and urea were predicted to improve growth rate only for the BM-7 media. Overall, modeling predictions showed that serine, malate, fumarate, and aspartate will improve the growth rate of *C*Las strains in culture medium M13, M14, and, M15. On the other hand, predicted metabolites limiting the growth rate are nicotinate, pantothenate, riboflavin, and aminobenzoate when BM-7 culture media was used.

While these results hint at specific metabolites that could improve *C*Las growth rates, they also suggest that phenotypic outcomes depend on the media composition to which each metabolic model is being subjected. The heatmap in Fig. 3c shows the predicted number of metabolites produced when biomass production is optimized. For microorganisms such as *C*Las gxpsy and Ishi-1 the media composition can strongly affect the ability of a strain to synthesize metabolites, varying from 60 to over 220 metabolites. The M15 media composition resulted in maximal metabolite production for every model, except for *C*Las gxpsy. The culture medium BM-7 resulted in a similar number of

produced metabolites for all seven models. We also evaluated the growth phenotype under eight different carbon sources assayed experimentally, predicting accurately the growth rate for oxoglutarate, glucose, and malate and an overall growth increase for fructose, fumarate, malate, maltose, and succinate (Fig 2d). The BT-1 model could explain up to 71% (Pearson correlation, p-value<0.0041) of the changes in the observed growth rate for the different carbon sources[32].

To test this experimentally with *C*Las, a cocktail of predicted metabolites was added to *in vitro* citrus hairy root cultures infected with *C*Las. Culture media B5, which is used to propagate the hairy root cultures, was supplemented with three concentrations (0.1, 1, and 5 mM) of an amino acid cocktail containing glycine, serine, proline, aspartate, glutamine, and glutamate. Measurements of *C*Las in the hairy root cultures obtained over the course of seven days showed that the treatment of 1 mM significantly improved *C*Las growth rate compared to 0.1 and 5 mM treatments as well as controls (F=5.99, Prob>F 0.01, df=14) (Fig. 3d).

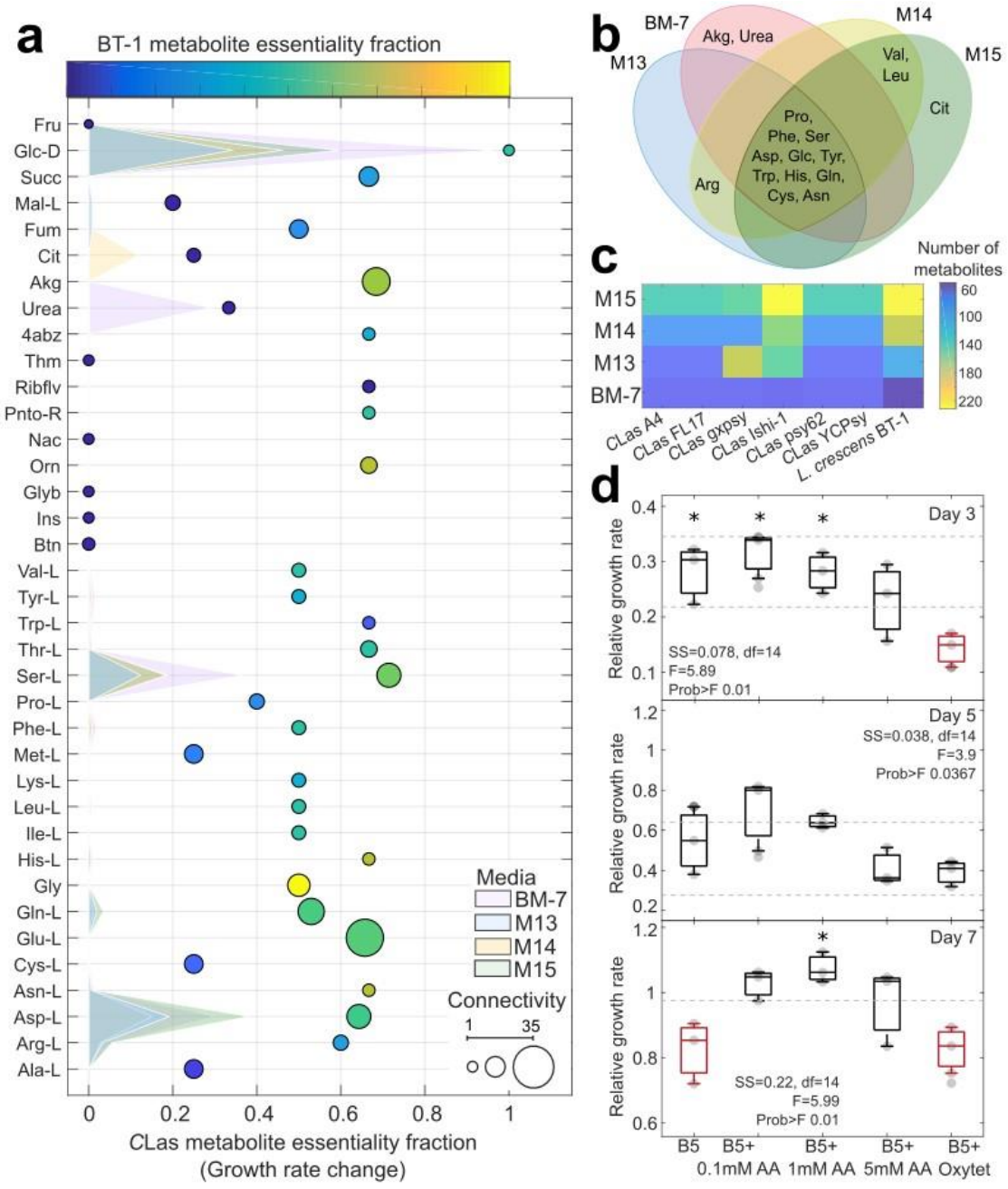


Fig. 3| Model evaluation by culture media. a, Metabolic connectivity and essentiality. The essentiality fraction
(change in growth rate) was determined by identifying the percentage of reactions in which each metabolite in the culture media participates. Metabolite essentiality in *Candidatus* Liberibacter asiaticus strains and *Liberibacter crescens* BT-1 are shown by bubble location and bubble color (non-essential in dark blue to essential in yellow), respectively. Bubble size indicates connectivity. Shaded surfaces represent the metabolite contribution to the growth rate by culture media. **b**, Metabolites predicted to improve *C*Las growth rates are compared across media compositions in the Venn

diagram. **c**, Biosynthetic capacity under maximal *Liberibacter* biomass growth rate by culture media. The color scale represents the number of metabolites predicted to be synthesized. **d**, *In vitro* measurements of *C*Las growth rates over time in hairy root-assays. Panels show *C*Las titers for day 3, 5, and 7.

Experimental measurements were normalized to one for day 0. Each panel shows the one-way ANOVA for the five assayed treatments (B5, culture medium suitable for the hairy root system; B5+0.1 mM AA, the hairy root system using B5 medium in addition with 0.1 mM of a cocktail of glycine, serine, proline, aspartate, glutamate, and glutamine; subsequently the amino acid cocktail was added in concentrations 1 mM and 5 mM for the treatments, B5+1 mM AA; B5 and B5+5 mM AA; the treatment B5+Oxytet contained 500 mg/L of the antibiotic oxytetracycline). Measurements were obtained using at least four and maximum five replicates of independent samples, each sample was measured four times (0, 3, 5, and 7 days). The ANOVA function tests the hypothesis that the samples (4-5 total) are drawn from populations with the same mean against the alternative hypothesis that the population means are not all the same. Standard ANOVA stats are given in each panel. Boxplots marked with * are significantly different from the red boxplot(s) of the same panel.

Host-dependent constraints reveal activation of *Candidatus* Liberibacter asiaticus pathways associated with pathogenic phenotypes. After validating the *L. crescens* BT-1 genome-scale metabolic model, we evaluated the response of *C*Las models using host-specific RNAsequencing data. Expression data was collected from phloem-enriched samples (referred to as phloem in the rest of the paper) from three Citrus cultivars (Valencia orange, Washington navel orange, and Tango mandarin) and from the alimentary canals of *Diaphorina citri* (psyllid). Using RNA-sequencing reads we generated strain-specific counts that were used to constrain boundaries of reactions in each model. Supplementary Table 5 shows statistics about data preprocessing of RNA-seq data such as number of raw reads and total counts generated after trimming reads aligned to the citrus or psyllid genomes. Supplementary Fig. 4 and 5 show the expression profiles for all *C*Las strains and the differentially expressed genes using a cut off of pvalue<0.05 (t-test). As expected, gene expression fold change between phloem and psyllid samples was normally distributed across genes. The maximum fold change observed was 28. Fig. 4a shows the distribution of the flux ratio used to constrain the *C*Las models from each host, as well as the fold change between samples from the two hosts obtained from RNA-sequencing data. Data from the individual phloem and psyllid samples were combined into two separate datasets that were used to constrain the models, giving insight into specific metabolic traits operated by *C*Las under each host. Constrained models are provided in Dataset 1.

Host-specific simulations performed using the *C*Las models (A4, FL17, gxpsy, Ishi-1, psy62, and YCPsy) resulted in a growth rate decrease of up to 68% and 74% for phloem

and psyllid, respectively, in comparison to media-constrained conditions in which the predicted and experimental growth rate was around 0.013±0.0009 1/h. Phloem-constraint flux distributions revealed increased amino acid metabolism, fatty acid metabolism, gluconeogenesis, nitrogen metabolism, one-carbon pool of folate, nicotinamide and nicotinate metabolism (Fig. 4b). Network evaluation showed that these pathways are interconnected by metabolites, such as formate, glycine, and 5,10-methylenetetrahydrofolate, linking these subsystems with possible pathogenic traits.

On the other hand, psyllid-constraint flux distributions revealed increased amino acid metabolism, cell envelope, glycolysis, pyruvate metabolism, and TCA cycle (Fig. 4b). The reactions associated with the cell envelope catalyze the synthesis of membrane lipids (enoyl reductase, EC 1.3.1.9) and oligosaccharides (DTDP-4-dehydrorhamnose 3,5-epimerase, EC 5.1.3.13, DTDPglucose 4,6-dehydratase, EC 4.2.1.46, and Glucose-1-phosphate thymidylyltransferase, EC 2.7.7.24). These oligosaccharide enzymes also participate in the metabolism of nucleotide sugars, streptomycin biosynthesis, and polyketide sugar biosynthesis. Polyketide biosynthesis is closely related to the additional availability of sugars and organic acids (e.g. pyruvate, succinate, malate), products of *C*Las metabolism. Fig. 4c shows the predicted flux distributions when BT-1 and *C*Las models are differentially constrained (media and expression data of phloem or psyllid). The overall flux distributions across all constraints showed that subsystems including glycolysis, transporters, TCA cycle, pyruvate metabolism, and nitrogen metabolism were highly activated, opposite to riboflavin metabolism, steroid biosynthesis, and cysteine and methionine metabolism. Furthermore, the pentose phosphate pathway was predicted to be active only under host-constraints while glycine, serine, and threonine metabolism reduced in flux. Broadly, *C*Las models constrained with phloem expression data clustered with models constrained with culture media M15 and *C*Las models constrained with psyllid expression data clustered with the flux distribution of BT-1 (Fig. 4c). These findings give insight into potential metabolic stages that facilitate *C*Las cultivation *in vitro* when obtained from the psyllid host.

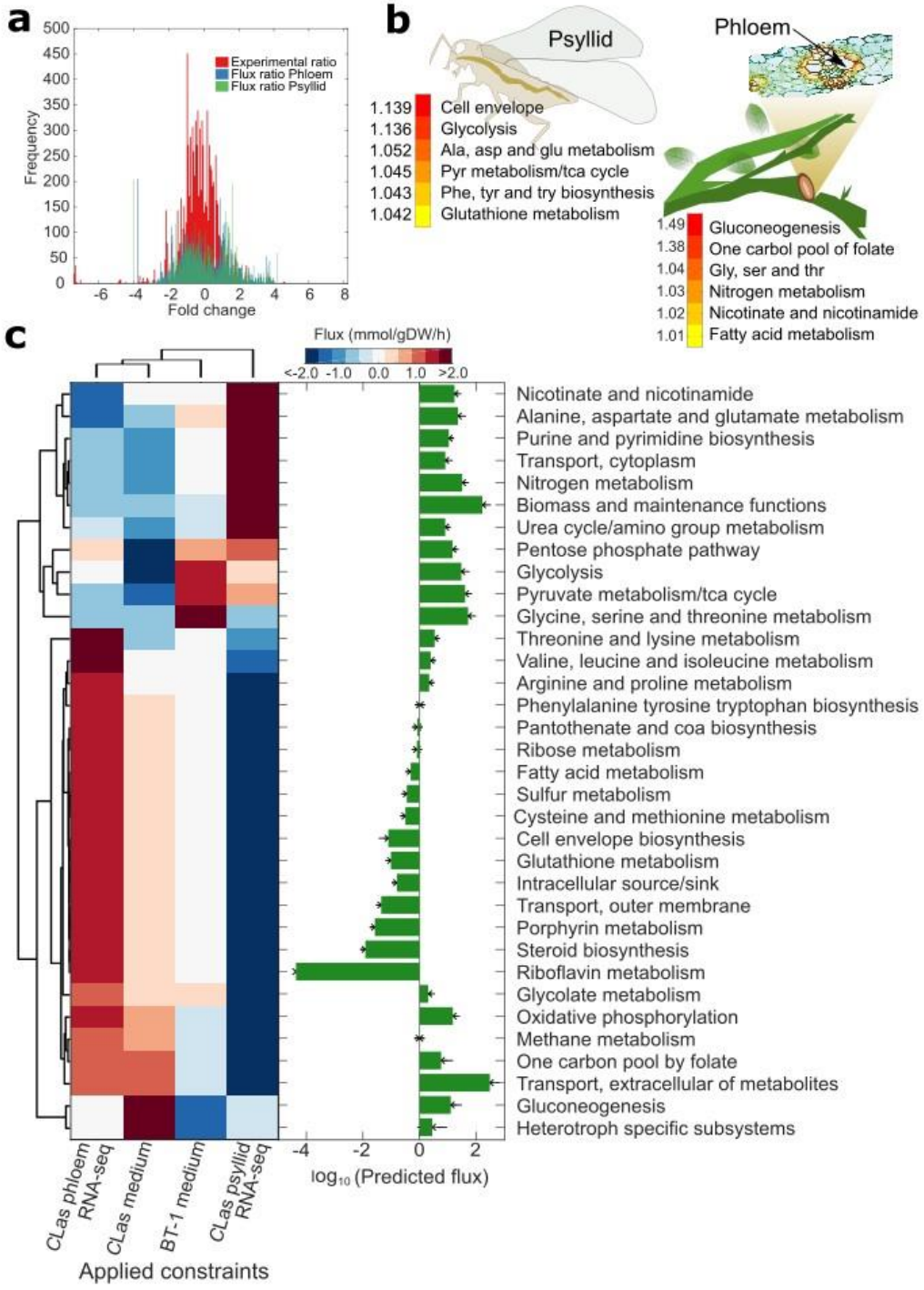


Fig. 4| Model-driven analysis of *Candidatus* Liberibacter asiaticus (CLas) RNA-sequencing data. a, The histogram shows the distribution of the average fold change of *C*Las expression for samples obtained from the psyllid alimentary canals and citrus phloem. In red the fold change between phloem and psyllid samples is shown using the entire RNA-sequencing dataset. Plotted in blue and green is the average ratio of

predicted flux distribution from the *C*Las model (A4, FL17, gxpsy, Ishi-1, psy62, and YCPsy) using M-15 culture medium and RNAsequencing data from psyllid and phloem samples, respectively. **b**, Heatmap highlighting *C*Las subsystems with highest flux activity in psyllid and phloem samples. Numbers represent the fold change increase of predicted flux. **c**, Analysis of predicted flux distributions from constraining the *C*Las models with RNA-sequencing data (phloem and psyllid), or culture medium M15 (Medium) compared to the predicted flux distribution for *L. crescens* BT-1 using M15 medium. Barplot shows the total flux carried by pathway by all the simulated flux distributions for all *C*Las and *L. crescens* strains. Correlation matrix showed that flux distribution cluster by constraints conditions independently of *C*Las strain (Supplementary Fig. 6), enabling to average the flux distribution by strain.

Predicting genetic targets for HLB management. Identification of potential *C*Las essential genes can lead to the development of HLB management strategies by designing molecules that specifically block or inhibit these gene products. We simulated single gene knockouts, changing the reaction bounds of all reaction(s) associated to each gene, while maintaining the hostdependent constraints for the rest of the network. Fig. 5 shows the phenotypic changes when media, psyllid RNA-seq data and phloem RNA-seq data constraints were applied to each of the six *C*Las models and the BT-1 model. We found that host-dependent constraints not only affect growth phenotypes in different environments, but also affect gene essentiality by strain. The *C*Las strain YCPsy was the most sensitive, with 94 essential genes in comparison with the strains gxpsy, Ishi-1, psy62, FL17 and A4, which had 93, 91, 90, 89, and 87 essential genes, respectively (see Supplementary Fig. 7). Overall, the number of essential genes increased around 27±2% because of the host-constraints imposed on the models compared to media constraints (Fig. 5a, green bars). Most of the essential genes common across the six *C*Las strains are involved in purine and pyrimidine metabolism, panthothenate and CoA biosynthesis, fatty acid metabolism and gluconeogenesis (Supplementary Fig. 7). However, when we evaluated the unique differences by subsystem across the six *C*Las strains, we found that genes involved in fatty acid metabolism, gluconeogenesis, glycine, isoleucine, leucine, serine, threonine, and valine metabolism, the TCA cycle, transport reactions, and urea cycle are the most sensitive and provide potential targets for the development of HLB management strategies (Fig. 5b). A full list of genes by subsystem is given in Supplementary Table 6-8.

Previously, a study of random transposon mutagenesis of *L. crescens* suggested that 105 metabolic genes were essential[33]. Genome-scale models contain 71 of those genes and predict that 18 of those genes are essential, 12 reduce the growth rate, and 41 are not essential (Supplementary Table 8). We then compared the 18 essential genes that were identified both experimentally and *in silico* with genes overexpressed during *C*Las infection to pinpoint potential targets for HLB mitigation. During *C*Las infection of *C. sinensis* the enzymes DTMP kinase (EC 2.7.4.9), inorganic diphosphatase (3.6.1.1), coproporphyrinogen oxidase (EC 1.3.3.3), and protoporphyrinogen oxidase (EC 1.3.3.4) were overexpressed (t-test, p-value<0.05, fold change>3) and identified as essential.

Additionally, we compared all predicted essential genes (91) with genes overexpressed (t-test, p-value<0.05 and fold change>10, n=3) in the citrus cultivars Valencia and Washington navel orange (*C. sinensis* L. Osb.) and Tango mandarin (*C. reticulata* Blanco) and the *C*Las enzymes phosphoglycerate mutase (EC 5.4.2.12), dihydroorotic acid (menaquinone-8) (EC 1.3.5.2), ribonucleoside-diphosphate reductase (UDP) (glutaredoxin) (EC 1.17.4.1), and glutaredoxin reductase (EC 1.20.4.1) were selected. Together, these results suggest eight distinct enzymes, whose inhibition could reduce *C*Las pathogenicity. The full dataset of metabolic reactions that are potential genetic targets in the *C*Las strains studied here are shown in Supplementary Table 9 and Supplementary Fig. 6.

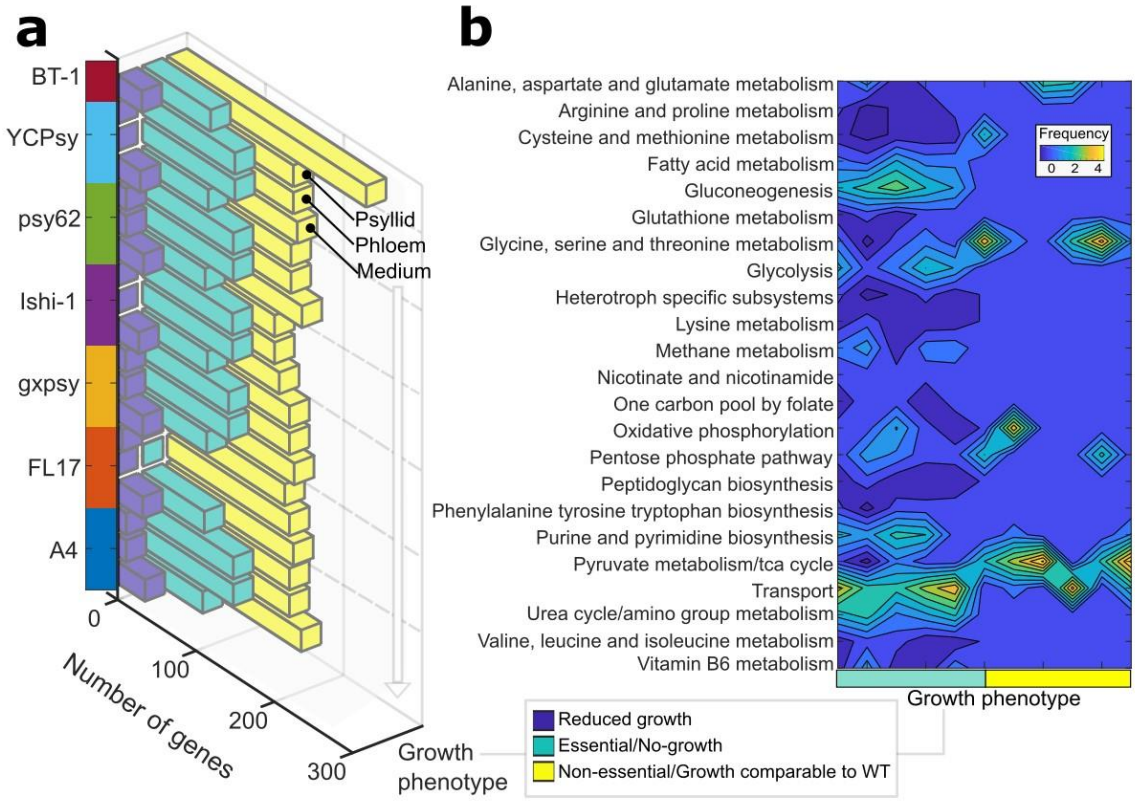


Fig. 5| Predicted gene essentiality by strain and host. a, Breakdown of phenotypes of all the genes in the metabolic models. Gene essentiality predictions are shown for all six *C*Las strains and BT-1. Phenotypes are binned into three categories based on predicted growth rate after individual gene knock-out; no growth (essential genes - green), reduced growth rate (blue), and comparable growth rate to wild-type conditions (non-essential - yellow). The gene essentiality results are plotted for each strain under the three different constraint datasets (psyllid RNAsequencing data, psyllid; phloem RNA-sequencing data, phloem; and culture medium M15, medium). Note that *L. crescens* BT-1 only contains predictions for the culture medium, because it is not a pathogen found in either the psyllid or the citrus phloem. **b**, The contour plot shows the number of *C*Las

genes that have unique phenotypes across the three constraint datasets, meaning that some genes that are non-essential in culture medium became essential when constrained by host conditions. We found that genes reducing the growth rate were consistent among growth conditions, while categorizing genes as essential or non-essential depended on the host. The number of genes that changed between essential and non-essential categorization under different host conditions is binned by subsystem, where yellow means maximum six genes changed in the subsystem, and dark blue is no genes changed in the subsystem. The complete dataset is provided in the Supplementary Table 6 and 7.

Discussion

Constraint-based modeling allowed us to elucidate metabolic changes in *Candidatus* Liberibacter asiaticus (*C*Las) during growth in the psyllid host *Diaphorina citri* Kuwayama and infection of the plant host *Citrus spp*. This work represents the first systems biology modeling approach to understand the metabolic role of *C*Las, the putative vector-borne causal pathogen of HLB, in citrus infection. We generated high-quality, manually curated genome-scale metabolic models of the six *C*Las strains A4, FL-17, gxpsy, Ishi-1, psy62, and YCPsy (Fig. 1). All models combine genomic and biochemical information with available literature resources to date. Manually curated models are characterized by an unprecedented quality in annotation[17,28,29], since they dramatically reduce the amount of possible misannotation caused by automated tools. In metabolic models, annotation is referred to as gene-protein-reaction associations. Compared to the genome annotations, approximately 24-28% of the gene-protein-reaction associations in the models were improved during the manual curation process, increasing the accuracy of predicted metabolic phenotypes (Supplementary Table 1).

Metabolic models are broadly used because they can simulate the metabolism of organisms with minimal experimental data, such as substrate uptake rates[31]. When such data is not available, it can be generalized using experimental data from closely related organisms[34]. For this purpose, we reconstructed a model of *L. crescens* BT-1, a closely related, culturable microorganism. We generated constraints to simulate growth phenotypes based on BT-1 experimental data[32]. The BT-1 model was validated by accurately predicting growth rates across four culture media compositions (i.e. BM-7, M13, M14, and M15) and multiple substrates (e.g. fumarate, glucose, oxoglutarate) (Fig. 2). In confirmation of these findings, citrate was recently discovered to improve the growth rate of BT-1 experimentally[32]. In addition, we experimentally confirmed our serine and other amino-acid predictions in *L. crescens* and the *C*Las-hairy root assays.

After successful validation of the BT-1 model we performed simulations to understand *C*Las metabolism. We found an interwoven effect of media composition on phenotypic traits, such as growth rate and metabolic production capabilities, which identified citrate and amino acids, such as glycine, serine, proline, aspartate, glutamine, and glutamate, as metabolites with a significant effect on *C*Las and *L. crescens* BT-1 growth (Fig. 2d,3d). Furthermore, it has been observed that metabolites, such as glycine, serine, citrate, glycine, glutamic acid, inositol, and malate, significantly change their concentration during *C*Las habitation in the psyllid host[35] and citrate, histidine, phenylalanine, and sucrose during infection of *C. sinensis*[36].

Genes essential for *L. crescens* BT-1 growth *in vitro* that are absent in *C*Las may be responsible for the failure of maintaining *C*Las strains in culture[37]. The lack of these genes suggests that *C*Las acquires aromatic amino acids, vitamins, saccharides, and fatty acids

from their hosts, as previously shown in other microbial communities[27]. We identified over 109 metabolic reactions that are present in *L. crescens* BT-1 but missing across all *C*Las strains (Supplementary Table 2). Previous studies have also suggested that *C*Las species lost the ability to synthesize proline, phenylalanine, tryptophan, cysteine, tyrosine, and histidine in addition to other translation components that may compromise regulatory systems[33,38]. We confirmed all of these auxotrophies and found that the proline, aspartate, arginine and serine auxotrophies are *C*Las strain-specific. Additionally, we predicted auxotrophies for steroids, cofactors and vitamins such as biotin, carnitine, choline, coniferol, riboflavin, and thiamin (Supplementary Fig. 2).

Using genome-scale metabolic models, we focused on understanding the metabolic behavior of *C*Las when it inhabits its two hosts. Application of modeling constraints based on *C*Las expression data enabled simulation and identification of metabolic changes at various functional stages, for example when *C*Las inhabits the psyllid or the plant. *In vivo* data (i.e. metagenomics and metatranscriptomics) are reliable sources of information for modeling uncultivable microorganisms. Host-specific (psyllid or plant expression data) constrained models predicted growth rates ~70% slower than media-constrained models. The predicted growth rate in the psyllid was higher than in the plant, as was previously observed experimentally[39]. These findings suggest different behaviors of *C*Las are dependent on its host (Fig. 4). *C*Las grows faster while inhabiting the psyllid, activating pathways related to nucleotide sugar metabolism, streptomycin biosynthesis, polyketide sugar unit biosynthesis, and cell envelope synthesis. Among these pathways, enzymes related to cell wall oligosaccharide enzymes were identified by screening all predicted flux distributions. It has been observed that *C*Las uses these enzymes to synthesize polysaccharides and thrive under different environments, especially in the presence of competitive bacterial biological agents[40–42]. These results suggest that in the psyllid host, *C*Las may activate the synthesis of antibiotics and antimicrobial precursors to compete with endogenous bacteria in the psyllid gut. On the other hand, in the citrus phloem, *C*Las may activate pathways that counteract plant defense mechanisms, such as the production of reactive oxygen species by NADPH oxidase[43,44], or the synthesis of antimicrobial peptides and long chain fatty acids[45,46] by activating reactions that depletes intermediaries of these toxic metabolites (e.g. orotic acid dehydrogenase, L-aspartate and glycolate oxidases). We also found that fatty acid metabolism was highly activated in *C*Las from citrus phloem samples, including enoyl-acyl reductase which has been associated with antibiotic resistance[47].

Significant progress has been made toward understanding the interactions between *C*Las and its hosts, and systems biology and omics tools can help to further unravel metabolic mechanisms associated with HLB initiation and progression as well as identify targets in *C*Las that can be used to develop HLB management strategies. Our results are consistent with and expand on prior findings. For example, other studies in *L. crescen*s have shown that supplementation of amino acids to the culture media increases growth

rates[33]. Gene essentiality simulations (Fig. 5) agree with previous findings, revealing ABC transporters, cell envelope biosynthesis, and fatty acid metabolism to be crucial subsystems for *C*Las[33,46,48]. Additionally, we found genetic targets in metabolic pathways whose inhibition may block the growth of *C*Las, thus preventing spread of this destructive disease (Supplementary Fig. 7). The systems biology tools presented here allow for the simulation of thousands of conditions, by applying environmental and/or genetic constraints, which reveal the vulnerabilities of *C*Las across various environments and improve our ability to guide future research and management efforts to combat this pathogen.

Methods

Draft model reconstruction and manual curation. Reconstructions are biochemically and genomically structured networks that contain information about associations among genes, reaction stoichiometry, and reaction reversibility. Here, we used a semi-automated process to reconstruct high-quality metabolic models, which comprises four fundamental steps: i) creation of an automated draft reconstruction, ii) draft refinement by manual curation, iii) conversion from reconstruction to mathematical model, and iv) model evaluation.

Semi-automatic reconstruction methods reduce building time, while maintaining high-quality architecture and prediction accuracy[49]. This method results in draft models which require refinement through manual curation. Draft models are generated based on protein-homology comparison between each protein sequence in the genome of the target microorganism (e.g. *C*Las) and the protein sequence of a manually curated reference model(s).

The reference models used here were chosen from the BiGG Database[50]. Supplementary Fig. 1 shows the phylogenetic relationships between *Liberibacter* strains and bacteria with available reference models in the repository. *Pseudomonas putida* KT2440, *i*JN746[51] was the closest related microorganism to *Liberibacter*, followed by *Yersinia pestis* CO92, *i*PC815[52]. The model of *Escherichia coli* str. K-12 substr. MG1655, *i*ML1515[53], was also used during the draft generation because it is the most extensively curated model to date. Table 1 shows the genome IDs of the protein sequences of *C*Las strains A4, FL17, gxpsy, psy62, YCPsy and *L. crescens* BT-1, which were used as input to The COBRA[54] and RAVEN Toolboxes[55] for MATLAB (The MathWorks Inc., Natick, MA).

Table 1. Properties of the genome-scale metabolic models

Microorganism	Genome ID	Model ID	Genes	Reactions	Metabolites
Candidatus Liberibacter asiaticus A4	34021.4	A4	283	840	837
FL17	34021.11	FL17	272	818	807
gxpsy	1174529.3	gxpsy	276	815	807
Ishi-1	931202.3	Ishi-1	253	818	802
psy62	537021.9	psy62	285	818	807
YCPsy	34021.12	YCPsy	279	814	805
Liberibacter crescens BT-1	1215343.11	BT-1	372	892	887

Each metabolic reaction in the reconstructed models was manually curated for their correct geneprotein-reaction association (GPR) using protein BLAST[56] to compare protein sequences between each strain of *Liberibacter* in the multi-strain model with sequences of *E. coli*, *P. putida*, and *Y. pestis* using UniProtKB/Swiss-Prot databases[57]. Transporter protein sequences were identified and compared using the TCBD database[58]. Metabolic reactions where no gene association could be found underwent another round of curation, where literature was reviewed to find evidence for the presence/absence of these proteins. Reactions with no supportive literature or matching sequences were included in the model for gap filling to ensure the completeness of relevant pathways[34]. The manual curation process was followed by model evaluation and validation. The reconstructions were analyzed for connectivity, mass and charge balance and converted into a functional mathematical model for simulation using The COBRA Toolbox[31]. Metabolic models were shared following the standard protocols for computational analysis[59].

Constraints and growth simulations. The seven *C*Las and BT-1 metabolic model reconstructions were constrained identically using the culture media BM-7, M13, M14, and M15[32]. All media compositions were simulated by setting a lower bound of −100 (allowing unlimited uptake) on the exchange reactions for Co^{2+}, Fe^{2+}, H^+, H_2O, Na^+, NH_4, PO_4, SO_4. Supplementary Table 3 shows the media compositions and applied constraints for each culture media. Growth simulations were performed using the flux balance analysis procedure[31]. Constraints on biomass composition were imported from the reference model of *P. putida* KT2440, *i*JN746[51].

The model topology was evaluated following the constraint-based modeling standard protocol[34]. The degree of metabolite connectivity (D) was determined by estimating the participation of metabolites in all reactions into the model (1). The matrix S is a feature in constraint-based modeling and its size is determined by the number of metabolites (rows) and reactions (columns) in the model. Reaction essentiality by metabolite was calculated by scanning the matrix S across all reactions. For the reactions in which each metabolite was found to participate the boundaries were set to zero and compound growth rate was estimated ($\mu_{comp,metabolite}$). The essentiality fraction was determined by the ratio between compound growth rate and the growth rate determined without any modification to the boundaries (wild type growth rate, μ_{WT}).

$$Dconnectivity,i = \sum Sbin,i,:$$

$$(A_{comp}): A_{comp} = S_{bin} * S_{bin}{}^{T} \quad (1)$$

$$\gamma_{benefitial,i=\frac{\mu_{comp,i}}{\mu_{WT}}}$$

Phenotypic experimental data

Liberibacter crescens **cultivation.** M15 media consists of $CaCl_2$•$2H_2O$ (1,320 mg/L), $MgCl_2$ (1,068.2 mg/L), $MgSO_4$•$7H_2O$ (2,778 mg/L), KCl (2,240 mg/L), NaH_2PO_4•H_2O (1,007 mg/L), L-alanine (447.24 mg/L), L-arginine-HCl (1,777 mg/L), L-asparagine monohydrate (1,075.45 mg/L), L-cysteine-2HCl (56.38 mg/L), L-glutamic acid (1,502.2 mg/L), glycine (859.51 mg/L), L-histidine hydrochloride monohydrate (2,366.11 mg/L), L-isoleucine (687.36 mg/L), L-leucine (592.89 mg/L), L-lysine-HCl (1,464.85 mg/L), L-methionine (678.9 mg/L), L-phenylalanine (789.62 mg/L), L-proline (940.61 mg/L), L-threonine (459.8 mg/L), L-tryptophan (373.73 mg/L), L-tyrosine disodium salt (391.37 mg/L), L-valine (644.31 mg/L), betaine monohydrate (0.36 mg/L), DL-ornithine hydrochloride (293.39 mg/L), methionine sulfoxide (18.2 mg/L), Dbiotin (0.1 mg/L), choline chloride (1,000 mg/L), folic acid (0.2 mg/L), myo-inositol (0.2 mg/L), niacin (0.2 mg/L), D-calcium pantothenate (0.2 mg/L), para-aminobenzoic acid (0.2 mg/L), pyridoxine-HCl (0.2 mg/L), riboflavin (0.2 mg/L), thiamine-HCl (0.2 mg/L), L-aspartic acid (2,500 mg/L), DL-serine (2,500 mg/L), L-glutamine (358.04 mg/L), and citric acid (2,500 mg/L)[32]. All ingredients were combined with the exception of tyrosine, which was first dissolved in 1M HCl before being added. Once all ingredients were dissolved, the medium was adjusted to pH 5.92 with 5M KOH and filter sterilized. Other derivatives of M15 (the different treatment media types) were prepared in the same way, but with differing concentrations of the components being examined (i.e. L-aspartic acid, DL-serine, L-glutamine, and citric acid). The concentrations of all other components were kept the same as the original M15 recipe. M15basic media was made with minimal (0.1

mg/L) amounts of the treatment components (i.e. citrate, serine, and aspartate or citrate, serine, and glutamine) but was otherwise kept the same as the original M15 recipe.

Liberibacter crescens (type strain BT-1^T, 5ATCC BAA-2481^{T}5DSM T 26877) was used for all experiments[21]. Glycerol stocks of *L. crescens* strain BT-1 in BM-7 complex media were used to inoculate M15 media, which was then shaken at 150 rpm and 28°C for 3-5 days to grow sufficient quantities for the experiments. Bacteria were pelleted via centrifugation at 6,000 rcf for 10 minutes, re-suspended in M15-basic medium, which does not contain the treatment components (citrate, serine, and aspartate or citrate, serine, and glutamine) and shaken at 150 rpm and 28°C for 1 hour to remove any large traces of the treatment components. OD_{600} was measured, and bacteria were re-pelleted using the same conditions described above. Pelleted bacteria were re-suspended in sterile DI water and used to inoculate treatment media for growth to stationary phase: OD_{600} = 0.8. Treatment media tubes were grown in 5 ml volumes in 16x100mm tubes at 150 rpm and 28°C. Growth was measured every 12-24 hours for 300 hours using a Spectronic-20 (Milton Roy, Houston, TX) spectrophotometer and OD_{600}.

***C*Las-Citrus hairy root culturing and *in vitro* assays.** The *ex vivo* *C*Las-citrus hairy root cultures were generated using methods described previously[60], with *C*Las-infected sour orange tissues (*Citrus x aurantium* L.) as ex-plant/inoculum source for *C*Las. Briefly, quantitative polymerase chain reaction (qPCR) validated CLas containing sour orange were identified and 510 cm shoots were excised for hairy root induction. The cut-end of the ex-plant was inoculated with fresh culture of *Rhizobium rhizogenes* (American Type Culture Collection strain 15834, OD 0.5) under gentle vacuum infiltration (~200 kPa). *R. rhizogenes* is a soil bacterium that naturally transforms plant cells to induce hairy roots at the point of contact by reprogramming plant hormone signaling [61]. In citrus, hairy root formation typically occurs in approximately 90 days, and because of the vascular connectivity between the shoot ex-plant, CLas naturally migrates into the newly formed hairy roots. Presence of *C*Las in the hairy root cultures was further confirmed by qPCR, using *C*Las-specific primers as described below. To determine the effect of amino-acids on *C*Las titers, *in vitro* assays were set up using the validated *C*Las-citrus hairy roots[60]. Briefly, the *C*Las-citrus hairy roots were surface sterilized with 70% ethanol and 2.5% sodium hypochlorite for five minutes followed by six washes with sterile water. Approximately 100 mg of *C*Las-citrus hairy roots were added to a multi-well culture plates and supplemented with B5+amino acid cocktail (glycine, serine, proline, aspartate, glutamine, and glutamate) concentrations (0, 0.1, 1 and 5 µM). An oxytetracycline (500ppm = 500mg/L) treatment was included as a *C*Las-inhibitor control for the *in vitro* assay. Four to five independent biological replicates were included for all treatments. The samples were vacuum infiltrated at 200 kPa for ~15 minutes to facilitate penetration of the media and nutrients into the hairy root cultures. The assay plates were placed on an orbital shaker at 40 rpm at~25°C and in dark. Fresh B5 medium was replaced at 3 and 5 days after incubation. Samples were collected at 0, 3, 5 and 7 days after treatment and flash-frozen in liquid nitrogen and stored at -80°C until further use.

DNA extraction and qPCR analysis. All control and treated *C*Las-citrus hairy root cultures were lyophilized and homogenized in a MiniG 1600 (Spex Sample Prep) homogenizer at 1500xrpm for 30 seconds, with a single steel bead (2 times, re-freezing samples at -80°C in between). Total DNA extraction was carried out according to Almeyda et al. (2001)[62]. qPCR reactions were carried out in a CFX-384 Real-Time PCR Detection System (BioRad, Hercules, CA) with 25ng of DNA as template, using Sso Advanced Universal SYBR Green Supermix (BioRad, Hercules, CA), and the following primers for citrus GAPC2 (CsiGAPC2-F 5'-TCTTGCCTGCTTTGAATGGA-3'and CsiGAPC2-R 5'-TGTGAGGTCAACCACTGCGACAT-3') and for the β-subunit od *nrd*B gene from *C*Las, RNR (RNRf 5'-CATGCTCCATGAAGCTACCC-3' and RNRr 5'-GGAGCATTTAACCCCACGAA-3')[63]. The reactions were carried out under the following conditions: Initial denaturation 95°C for 3 minutes, followed by 95°C for 15 s and 55°C for 30 s for 40 cycles, and a final extension at 65°C for 5 s. Relative *C*Las titers were estimated using the $^{\Delta\Delta}$Ct method [64]. Briefly, the *C*Las Ct was first normalized to the housekeeping gene (GAPC2) to account for DNA template differences, and then to the 0 days Ct which was set to 1 (or 100%). Growth rates at each time point were calculated using the initial *C*Las titer as a reference point. The minimal and maximal data were discarded before the ANOVA analysis. Calculations were performed using The Preprocessing Data and The Statistics and Machine Learning Toolboxes of MATLAB (The MathWorks Inc.).

Expression data. RNA sequencing data collected from environmental samples was used to constrain the *C*Las models. The samples were obtained from the phloem-enriched samples from different citrus cultivars and from Asian citrus psyllid (ACP) alimentary canals as described below. For each growth condition, the storage and consumption of starch, calculated using experimental data, were taken into account (Supplementary Table 6).

RNA extraction from citrus. Samples were harvested from 12 *C*Las-infected citrus trees grown in a greenhouse at the U.S. Horticultural Research Facility in Fort Pierce, Florida. Three trees were selected each from three different *Citrus* cultivars: Valencia orange (*Citrus sinensis* [L.] Osbeck) on Swingle citrumelo (*C. paradisi* Macf. X Poncirus trifoliate [L.] Raf.) rootstock, Tango mandarin (*Citrus reticulata* Blanco) on Sour orange rootstock, and Washington navel orange (*Citrus sinensis* [L.] Osbeck) on Sour orange rootstock (Supplementary Table 6). One to two years prior to sampling, the greenhouse trees were exposed to *C*Las-positive ACP for varying lengths of time between one week and one month, and *C*Las infection was verified using qPCR at or near the time of harvest.

Four pieces of budwood that were roughly one year old and ~15 cm long were harvested from each tree and immediately placed on ice. Within 15 minutes, they were sampled from as follows: budwood was removed from ice and sprayed with CVS brand Health Alcohol Free Liquid Bandage Spray (CVS, Woonsocket, RI) to prevent surface contamination. After drying for three minutes, the bark was peeled from each piece and the inside surface located away from the cut ends was quickly scraped with a razor blade twice – first to remove surface contamination and potential xylem and second to collect

a phloem-enriched sample. Samples were immediately placed in PowerBead Tubes filled with Solution MBL and the Phenolic Separation Solution from an RNeasy PowerPlant Kit (Qiagen, Valencia, CA), which were held in a CoolRack (BioCision, San Rafael, CA) on dry ice. RNA was extracted using the RNeasy PowerPlant Kit following the kit protocol with two minutes of bead beating, eluted in 50 µl of RNase-free water, and stored at -80°C for library preparation.

RNA extraction from Asian citrus psyllids. Approximately 100 adult ACP were collected from *C*Las-exposed colonies maintained at the U.S. Horticultural Research Facility in Fort Pierce, Florida using an aspirator. Alimentary canals were dissected from ACP in a weight-boat containing Solution PM1 from an RNeasy Power Microbiome Kit (Qiagen, Valencia, CA) placed on ice, and deposited in two PowerBead Tubes containing Solution PM1 (50 canals each) held in a CoolRack (BioCision) on dry ice. RNA was extracted using the RNeasy PowerMicrobiome Kit following the kit protocol with one minute of bead beating, eluted in 50 µl of RNase-free water, and stored at -80°C for library preparation.

Library preparation and sequencing. Alimentary canal sample cDNA libraries were prepared using the ScriptSeq Complete Gold Kit (Yeast) (Illumina, San Diego, CA), following kit protocols and performing ribo-depletion. Citrus sample cDNA libraries were prepared using the ScriptSeq Complete Kit (Plant Leaf) (Illumina, San Diego, CA), again using provided protocols and ribo-depletion. In both cases, ScriptSeq Index PCR Primers (Illumina, San Diego, CA) were used for barcoding samples. RNA Sequencing was performed using Illumina's HiSeq2500 platform. Raw RNA reads were trimmed using TrimGalore (version 0.4.4) including adapter removal and quality control: low-quality ends from reads (Phred score < 20) were trimmed and reads less than 20 bp were discarded. Next, read quality was checked using FastQC (version 0.11.7). To discard host and 16S rRNA reads, *C. maxima* (Burm.) Merr. genome (NCBI_Assembly: GCA_002006925.1) and bacterial 16S rRNA sequences (SILVA database: https://www.arb-silva.de/) were chosen as reference templates. Valid reads were aligned to reference templates using bowtie2 (version 2.3.4.1) with parameters set by the flag verysensitive. Unmatched reads were picked out and converted to fastq format using samtools (version 1.8) and bam2fastq (http://www.hudsonalpha.org/gsl/information/software/bam2fastq), respectively. To count the FPKM (fragment per kilobase per million mapped reads), reads were mapped to 7 Liberibacter strains: *C*Las strains A4 (GCF_000590865.2), FL17 (GCF_000820625.1), psy62 (GCF_000023765.2), YCPsy (GCF_001296945.1), gxpsy (GCF_000346595.1), Ishi-1 (GCF_000829355.1) and *L. crescens* BT-1 (GCF_000325745.1).

Acknowledgments

This work was supported in part by funds from National Science Foundation under Grant No. 1332344 and CBET-1804187, and the U.S. Department of Energy (DOE), Office of Science, Office of Biological & Environmental Research under Award DE-SC0012658,

the University of California Riverside (UCR) Agricultural Experiment Station Mission Funding Program, a UCR Seed Grant, the University of California Office of the President via a grant from the Multicampus Research Programs and Initiatives (MRP-19-601384), the USDA National Institute of Food and Agriculture (NIFA) Hatch Projects CA-R-PPA-7774-H, 233744, and USDA-NIFACDRE (2018-70016-28198 and 2019-70016-29066). We gratefully acknowledge funding from the Trial Ecosystem Advancement for Microbiome Science and the Microbial Community Analysis and Functional Evaluation in Soils Programs at Lawrence Berkeley National Laboratory by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research Award DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory. Cristal Zuniga was in part supported by Mexican National Research Council, CONACYT, fellowship No. 237897. We thank Livia Zaramela for assistance during the analysis of expression data.

Authors contributions

C.Z., J.B., and K.Z. designed research; B.P., G.V., G.M. and S.I. generated expression data and performed growth experiments; C.Z., N.W., D.T., A.Z., and B.L. generated the models; C.Z., B.L., and C. M. analyzed data, and C.Z. and C.M. wrote the paper with input of all co-authors**.**

Declaration of interests

The authors declare no conflict of interest.

Supplemental Information

Dataset 1. Models were constrained using the traditional culture medium BG-7 and the optimized culture media M13, M14 and M15. Experimental growth measurements showed that citrate benefits the growth of *L. crescens* as predicted by the model (Fig.2). All sequencing reads were deposited in the Sequence Read Archive under BioProject PRJNA509215, with specific numbers listed in Supplementary Table 5. Additionally, all supplemental materials are available at https://github.com/cristalzucsd/Liberibacte

References

1. Soliman, T., Mourits, M. C. M., Oude Lansink, A. G. J. M. & van der Werf, W. Economic justification for quarantine status - the case study of ‘ Candidatus *Liberibacter solanacearum*’ in the European Union. *Plant Pathol.* **62,** 1106–1113 (2013).

2. Walter, A. J., Duan, Y. & Hall, D. G. Titers of ‘*Ca*. Liberibacter asiaticus’ in *Murraya paniculata* and *Murraya*-reared *Diaphorina citri* Are Much Lower than in *Citrus* and *Citrus*-reared Psyllids. *HortScience* **47,** 1449–1452 (2012).

3. Albrecht, U. & Bowman, K. D. Gene expression in *Citrus sinensis* (L.) Osbeck following infection with the bacterial pathogen *Candidatus* Liberibacter asiaticus causing Huanglongbing in Florida. *Plant Sci.* **175,** 291–306 (2008).

4. Cevallos-Cevallos, J. M., Futch, D. B., Shilts, T., Folimonova, S. Y. & Reyes-DeCorcuera, J. I. GC–MS metabolomic differentiation of selected citrus varieties with different sensitivity to citrus huanglongbing. *Plant Physiol. Biochem.* **53,** 69–76 (2012).

5. Gottwald, T. R. Current epidemiological understanding of citrus Huanglongbing. *Annu. Rev. Phytopathol.* **48,** 119–139 (2010).

6. Kim, J.-S., Sagaram, U. S., Burns, J. K., Li, J.-L. & Wang, N. Response of sweet orange (Citrus sinensis) to ‘Candidatus Liberibacter asiaticus’ infection: Microscopy and microarray analyses. *Phytopathology* **99,** 50–57 (2009).

7. Zhao, H. *et al.* Small RNA profiling reveals phosphorus deficiency as a contributing factor in symptom expression for citrus Huanglongbing disease. *Mol. Plant* **6,** 301–310 (2013).

8. da Graça, J. V. *et al.* Huanglongbing: An overview of a complex pathosystem ravaging the world’s citrus. *J. Integr. Plant Biol.* **58,** 373–387 (2016).

9. Narouei-Khandan, H. A., Halbert, S. E., Worner, S. P. & van Bruggen, A. H. C. Global climate suitability of citrus huanglongbing and its vector, the Asian citrus psyllid, using two correlative species distribution modeling approaches, with emphasis on the USA. *Eur. J. Plant Pathol.* **144,** 655–670 (2016).

10. Bové, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *J. plant Pathol.* **88,** 7–37 (2006).

11. Belasque, J. *et al.* Lessons from huanglongbing management in São Paulo state, Brazil. *J. Plant Pathol.* **92,** 285–302 (2010).

12. Puttamuk, T. *et al.* Genetic diversity of Candidatus *Liberibacter asiaticus* based on two hypervariable effector genes in Thailand. *PLoS One* **9,** e112968 (2014).

13. Akula, N., Trivedi, P., Han, F. Q. & Wang, N. Identification of small molecule inhibitors against SecA of Candidatus *Liberibacter asiaticus* by structure based design. *Eur. J. Med. Chem.* **54,** 919–924 (2012).

14. Canales, E. *et al.* 'Candidatus *Liberibacter asiaticus*', causal agent of citrus Huanglongbing, is reduced by treatment with brassinosteroids. *PLoS One* **11,** e0146223 (2016).

15. Munir, S. *et al.* Huanglongbing control: perhaps the end of the beginning. *Microb. Ecol.* **76,** 192–204 (2018).

16. Ward, L. *et al.* Microbial community dynamics in Inferno Crater Lake, a thermally fluctuating geothermal spring. *ISME J.* **i,** 1–10 (2017).

17. Seif, Y. *et al.* A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLOS Comput. Biol.* **15,** e1006644 (2019).

18. Taylor, R. A., Mordecai, E. A., Gilligan, C. A., Rohr, J. R. & Johnson, L. R. Mathematical models are a powerful method to understand and control the spread of Huanglongbing. *PeerJ* **4,** e2642 (2016).

19. Jagoueix, S., Bove, J.-M. & Garnier, M. The phloem-limited bacterium of greening disease of citrus is a member of the subdivision of the Proteobacteria. *Int. J. Syst. Bacteriol.* **44,** 379–386 (1994).

20. Davis, M. J., Mondal, S. N., Chen, H., Rogers, M. E. & Brlansky, R. H. Co-cultivation of '*Candidatus* Liberibacter asiaticus' with Actinobacteria from Citrus with Huanglongbing. *Plant Dis.* **92,** 1547–1550 (2008).

21. Fagen, J. R. *et al. Liberibacter crescens* gen. nov., sp. nov., the first cultured member of the genus Liberibacter. *Int. J. Syst. Evol. Microbiol.* **64,** 2461–2466 (2014).

22. Tyler, H. L., Roesch, L. F. W., Gowda, S., Dawson, W. O. & Triplett, E. W. Confirmation of the sequence of '*Candidatus* Liberibacter asiaticus' and assessment of microbial diversity in Huanglongbing-infected citrus phloem using a metagenomic approach. *Mol. Plant-Microbe Interact.* **22,** 1624–1634 (2009).

23. Zuñiga, C. *et al.* Predicting dynamic metabolic demands in the photosynthetic eukaryote *Chlorella vulgaris*. *Plant Physiol.* **176,** 450–462 (2018).

24. Al-Bassam, M. M. *et al.* Optimization of carbon and energy utilization through differential translational efficiency. *Nat. Commun.* **9,** 4474 (2018).

25. Du, B., Zielinski, D. C., Monk, J. M. & Palsson, B. O. Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. *Proc. Natl. Acad. Sci.* **115,** 11339–11344 (2018).

26. Li, C.-T. *et al.* Utilizing genome-scale models to optimize nutrient supply for sustained algal growth and lipid productivity. *npj Syst. Biol. Appl.* **5,** 33 (2019).

27. Zuñiga, C. *et al.* Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. *Nat. Microbiol.* 1–8 (2019).

28. Tibocha-Bonilla, J. D., Zuñiga, C., Godoy-Silva, R. D. & Zengler, K. Advances in metabolic modeling of oleaginous microalgae. *Biotechnol. Biofuels* **11,** 241 (2018).

29. Zuñiga, C. *et al.* Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Physiol.* **172,** 589–602 (2016).

30. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42,** D581–D591 (2014).

31. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraintbased models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6,** 1290–1307 (2011).

32. Cruz-Munoz, M. *et al.* Development of chemically defined media reveals citrate as preferred carbon source for *Liberibacter* growth. *Front. Microbiol.* **9,** (2018).

33. Lai, K.-K., Davis-Richardson, A. G., Dias, R. & Triplett, E. W. Identification of the genes required for the culture of *Liberibacter crescens*, the closest cultured relative of the Liberibacter plant pathogens. *Front. Microbiol.* **7,** (2016).

34. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5,** 93–121 (2010).

35. Killiny, N., Nehela, Y., Hijaz, F. & Vincent, C. I. A plant pathogenic bacterium exploits the tricarboxylic acid cycle metabolic pathway of its insect vector. *Virulence* **9,** 99–109 (2018).

36. Slisz, A. M., Breksa, A. P., Mishchuk, D. O., McCollum, G. & Slupsky, C. M. Metabolomic analysis of citrus infection by '*Candidatus* Liberibacter' reveals insight into pathogenicity. *J. Proteome Res.* **11,** 4223–4230 (2012).

37. Parker, J. K. *et al.* Viability of ‘ *Candidatus* Liberibacter asiaticus’ prolonged by addition of citrus juice to culture medium. *Phytopathology* **104,** 15–26 (2014).

38. Mendonça, L., Zambolim, J. & Badel, L. Bacterial citrus diseases: major threats and recent progress. *J. Bacteriol. Mycol. Open Access* **5,** (2017).

39. Arratia-Castro, A. A. *et al.* Detection and quantification of ‘ *Candidatus* Phytoplasma asteris’ and ‘ *Candidatus* Liberibacter asiaticus’ at early and late stages of Huanglongbing disease development. *Can. J. Plant Pathol.* **38,** 411–421 (2016).

40. Jiang, L. Response of Candidatus Liberibacter asiaticus-Infected citrus plants to Bacillus amyloliquefaciens GJ1. *Austin J. Plant Biol.* **4,** (2018).

41. Jiang, L., Gao, Z., Li, Y., Wang, S. & Dong, Y. Crystal structures and kinetic properties of enoyl-acyl carrier protein reductase I from *Candidatus* Liberibacter asiaticus. *Protein Sci.* **23,** 366–377 (2014).

42. Nehela, Y. & Killiny, N. *Candidatus* Liberibacter asiaticus and its vector, *Diaphorina citri*, augments the TCA cycle of their host via the GABA shunt and polyamines pathway. *Mol. Plant-Microbe Interact.* MPMI-09-18-0238-R (2018). doi:10.1094/MPMI-09-180238-R

43. Fluhr, R. in *Reactive Oxygen Species in Plant–Pathogen Interactions* 1–23 (2009). doi:10.1007/978-3-642-00390-5_1

44. Pitino, M., Armstrong, C. M. & Duan, Y. Molecular mechanisms behind the accumulation of ATP and H_2O_2 in citrus plants in response to ‘*Candidatus* Liberibacter asiaticus’ infection. *Hortic. Res.* **4,** 17040 (2017).

45. Shah, J. Lipids, lipases, and lipid-modifying enzymes in plant disease resistance. *Annu. Rev. Phytopathol.* **43,** 229–260 (2005).

46. Suh, J. H., Niu, Y. S., Wang, Z., Gmitter, F. G. & Wang, Y. Metabolic analysis reveals altered long-chain fatty acid metabolism in the host by Huanglongbing disease. *J. Agric. Food Chem.* **66,** 1296–1304 (2018).

47. Yao, J. & Rock, C. O. Resistance mechanisms and the future of bacterial enoyl-acyl carrier protein Reductase (FabI) antibiotics. *Cold Spring Harb. Perspect. Med.* **6,** a027045 (2016).

48. Montesinos, E. Antimicrobial peptides and plant disease control. *FEMS Microbiol. Lett.* **270,** 1–11 (2007).

49. Zuñiga, C., Zaramela, L. & Zengler, K. Elucidation of complexity and prediction of interactions in microbial communities. *Microb. Biotechnol.* **10,** 1500–1522 (2017).

50. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44,** D515–D522 (2016).

51. Nogales, J., Palsson, B. Ø. & Thiele, I. A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* **2,** 79 (2008).

52. Charusanti, P. *et al.* An experimentally-supported genome-scale metabolic network reconstruction for *Yersinia pestis* CO92. *BMC Syst. Biol.* **5,** 163 (2011).

53. Monk, J. M. *et al. i*ML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35,** 904–908 (2017).

54. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraintbased models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6,** 1290–1307 (2011).

55. Agren, R. *et al.* The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* **9,** e1002980 (2013).

56. Schaffer, A. A. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29,** 2994–3005 (2001).

57. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204– D212 (2014).

58. Saier, M. H., Reddy, V. S., Tamang, D. G. & Västermark, A. The transporter classification database. *Nucleic Acids Res.* **42,** D251-8 (2014).

59. Rule, A. *et al.* Ten simple rules for writing and sharing computational analyses. *PLOS Comput. Biol.* **15,** e1007007 (2019).

60. Mandadi, K., Irigoyen, S. C. & Mirkov., T. E. Methods, compositions, and systems for culturing and characterizing fastidious plant microbes. *U.S. Pat. Appl.* 1–10 (2017).

61. Chilton, M.-D. *et al. Agrobacterium rhizogenes* inserts T-DNA into the genomes of the host plant root cells. *Nature* **295,** 432–434 (1982).

62. León Almeyda, I. H., Rocha Peña, M. A. ., Piña Razo, J. . & Martínez Soriano, J. P. The use of polymerase chain reaction and molecular hybridization for detection of phytoplasmas in different plant species in México. *Rev. Mex. Fitopatol.* **19,** 1–9 (2001).

63. Zheng, Z. *et al.* Unusual five copies and dual forms of nrdB in '*Candidatus* Liberibacter asiaticus': Biological implications and PCR detection application. *Sci. Rep.* **6,** 39020 (2016).

64. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **25,** 402–408 (2001).

CONCLUSIONES

El uso de algoritmos y estrategias semiautomáticas permitieron obtener en este trabajo seis modelos metabólicos de un consorcio microbiano nitrificante presente en aguas residuales en un período de tiempo relativamente corto, comparado con estrategias manuales y manteniendo la calidad en las predicciones fenotípicas. La optimización de los parámetros para la construcción inicial de los modelos metabólicos utilizando como pilar la herramienta BLAST cumplió un rol fundamental dentro del proceso de desarrollo de los modelos de forma semiautomática, proporcionando modelos bases funcionales de los microorganismos identificados en el metagenoma del consorcio nitrificante. Además, las estrategias asociadas al refinamiento de los modelos bases diseñadas en el presente trabajo aumentaron considerablemente la calidad en las asociaciones genes-proteínas-reacciones en cada modelo generado. Los modelos metabólicos resultantes permitieron identificar la capacidad metabólica de cada microorganismo modelado, ya que se realiza un análisis a nivel genoma de todas las reacciones metabólicas involucradas, y, por lo tanto, determinar con alta fidelidad (mayor al 90%) procesos metabólicos internos, así como producción e intercambio de metabolitos de interés, al evaluar las predicciones realizadas con información experimental presente en la literatura. Por último, se evaluaron las capacidades metabólicas para intercambiar metabolitos de cada uno de los seis microorganismos, de tal forma que se identificó adecuadamente las principales rutas metabólicas de cada microorganismo en su participación en condiciones experimentales de tratamiento de aguas residuales.

Perspectivas

Creemos que los modelos desarrollados en el presente trabajo proporcionan un paso valioso en el camino hacia una mejor caracterización de estos importantes microorganismos aislados como parte de una comunidad microbiana. El diseño y la evaluación de diferentes comunidades microbianas provenientes del consorcio microbiano nitrificante a partir de los seis microorganismos estudiados en el presente trabajo desde enfoques de bioinformática y biología de sistemas podría mejorar el entendimiento que tenemos actualmente de los sistemas biológicos y su uso en el tratamiento de aguas residuales