# MODELLING RHIZOBIAL METABOLISM IN RHIZOBIUM-LEGUME INTERACTIONS




Carolin Christine Margaret Schulte

Keble College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Trinity 2021

I declare that this thesis is my own work and that this work has not been submitted for a degree at any other university. I would like to acknowledge the contributions of all authors involved in the work presented in Chapter 3 and Chapter 4. RNA-Seq data presented in Chapter 3 were acquired and analysed by Vinoy Ramachandran. Metabolic flux analysis in Chapter 4 was performed by Khushboo Borah and proteome data were acquired and analysed by Jason Terpolilli and Gerhard Saalbach.

# Acknowledgements

When I came to Oxford almost four years ago, I could never have imagined how many incredible people I would be lucky enough to meet during my time as a DPhil student. Even though I cannot individually acknowledge everyone who has made my DPhil journey so special, please know that I am forever grateful to every single one of you!

First and foremost, my most sincere thanks and gratitude go to my supervisors, Antonis and Phil. I know I haven't always been the easiest student to supervise, so I deeply appreciate your patience and support and the freedom you gave me to explore my various research interests. Antonis, thank you for agreeing to supervise the shy and stressed first-year student who walked into your office three years ago, I have always valued your advice, both on research questions and on life philosophy. Phil, your sustained passion for your research is truly inspiring. At the start of my DPhil, I would not have thought that I could ever develop an interest in metabolism, but I ended up nothing short of enthusiastic about the intricacies of redox balance in bacteroids.

Harry, thank you for convincing me to join the Control Group. You are one of the smartest and most hard-working people I know, and you've been an inspiration to work alongside. I have also been very fortunate to get additional input from many brilliant scientists as part of research visits, and I would like to thank Antonella Succurro, Bas Teusink and Daan de Groot for the time they spent discussing metabolic modelling of bacteroids with me. To the team at Sound Agriculture, thank you for making my time in California so enjoyable, you know I'll be back as soon as I can!

I would not have been able to do my DPhil at Oxford without the generous support of the Clarendon Fund and the Keble College De Breyne Scholarship, and I extend my sincerest thanks to the donors.

I would like to acknowledge the amazing staff at the Departments of Engineering Science and Plant Sciences. Thank you, Maria, for our early-morning conversations, you always managed to cheer me up! I am very grateful to all the members of the Rhizosphere Lab and the Control Group who I got to know over the course of my DPhil journey. Special thanks to Marcela for helping me through the first few months of my DPhil as well as Carmen and Raphael for their help and advice on all things metabolism. Alison, I have no idea how things in the lab would work without you, thanks for always being on top of organising everything! Javi, I am never quite sure if I should be mad at you for talking me into taking on so much work for the MCR, but I have had the best time sharing an office with you!

I would not have survived my DPhil without the continuous support of my wonderful housemates. Josh, Michael, Pippa and Brandon, thanks for putting up with me (and my sometimes excessive cleaning habits), I will miss you all so much! Similarly, my time in Oxford would not have been the same without the countless hours spent with the Inulox team. Josh, Niels, Andreas and Joel – I will forever remember those nights spent working on business plans (and Excel spreadsheets). I am incredibly proud of how far we have come, and I have learned so much on this rollercoaster ride with you. I really hope you'll be free to meet up at Franco Manca when I come to visit the UK!

Felix, Niels, Alex, Xanita and Matt - you are amazing, and the wonderful time we had in Croatia was one of my highlights of the last four years.

I dedicate this thesis to my parents. Mum and Dad, knowing that you're always there for me no matter what means the world to me.

# Table of contents

Acknowledgements …… 2

Table of contents …… 4

Abstract …… 5

1. Introduction …… 7

    1.1 Context …… 7

    1.2 Motivation …… 8

    1.3 Aims …… 10

    1.4 Thesis outline …… 11

    1.5 Scientific contributions …… 13

2. Background: How rhizobia adapt to the nodule environment …… 14

3. Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum* …… 46

4. Metabolic control of nitrogen fixation in rhizobium-legume symbioses….. 135

5. Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses…… 181

6. Discussion and outlook…… 231

7. Bibliography…… 239

# Abstract

In this work, metabolic modelling approaches were combined with experimental datasets to elucidate rhizobial metabolism in rhizobium-legume interactions. First, a genome-scale metabolic model for *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) was reconstructed and validated. Integration of transcriptome, proteome, metabolome and gene essentiality data was performed to obtain models specific to rhizosphere bacteria, undifferentiated nodule bacteria and nitrogen-fixing bacteroids. Analysis of these models revealed the importance of leucine, polyphosphate and AICAR synthesis in the rhizosphere and indicated that *myo*-inositol catabolism is specifically active in undifferentiated nodule bacteria. The bacteroid model predicted active synthesis of histidine as well as uptake of a pentose sugar in addition to dicarboxylates, which are the main carbon sources for bacteroids.

Next, a reduced model of bacteroid metabolism was obtained using experimental datasets for Rlv3841. Computational methods that do not require the specification of nutrient uptake rates were applied to investigate properties of bacteroid metabolism based on a minimum number of assumptions. Analysing the effect of oxygen availability on central carbon and nitrogen metabolism suggested that both carbon polymer synthesis and amino acid secretion by bacteroids function to facilitate carbon and redox balance under the microaerobic conditions required for nitrogen fixation. The model further predicted that low oxygen levels restrict flux through the decarboxylating tricarboxylic acid cycle arm, which would limit ammonia assimilation by bacteroids and thus constitute a fundamental prerequisite for a productive symbiosis. Model predictions were supported by experimental measurements of oxygen uptake rates and $NADH/NAD^+$ ratios as well as metabolic flux analysis.

Finally, a comparison of bacteroid metabolism for different rhizobial strains was performed to evaluate differences and similarities of symbioses between *Rhizobium* and *Sinorhizobium* strains and various host plants. Genome-scale metabolic models for six different rhizobium-legume symbioses were reconstructed and reconciled, and bacteroid models were obtained through integration of gene expression data. The models were able to correctly identify strain- and symbiosis-specific differences in the pathways that are used for pyruvate synthesis. In addition, significantly lower metabolic activity and symbiotic efficiency were predicted for reversibly differentiated bacteroids compared to terminally differentiated bacteroids, which is consistent with experimental evidence.

# 1. Introduction

## 1.1 Context

Rhizobia are common soil-dwelling bacteria that can enter complex symbiotic relationships with certain legume plants. In these symbioses, rhizobia inhabiting nodules on the plant root convert atmospheric $N_2$ gas into ammonia, which is secreted to the plant in exchange for carbon sources[1]. This biological nitrogen fixation is a sustainable alternative to the application of nitrogenous fertilizers, which are energy-intensive to produce and cause a plethora of environmental issues, such as eutrophication and release of potent greenhouse gases[2,3].

Symbiosis formation is a multi-step process: following a complex exchange of signalling molecules with the plant, rhizobia are entrapped by root hairs and grow down a plant-derived infection thread until they are endocytosed by a plant cell in the developing nodule and differentiate into their nitrogen-fixing form called bacteroid[4,5]. Bacteroids are physiologically very different from free-living rhizobia since they are non-dividing and perform energy-intensive nitrogen fixation at nanomolar oxygen levels in close interdependence with their plant host[1,5,6].

With increasing availability of whole-genome sequences, metabolic modelling has become a popular tool for investigating host-symbiont relationships[7–9]. Metabolic models are knowledge bases containing stoichiometric descriptions of the reactions that can be catalysed by the enzymes encoded in an organism's genome[10–12]. As metabolism happens on a very fast timescale compared to cellular growth, it can often be assumed to be at steady state, where no accumulation or depletion of intracellular metabolites occurs. Metabolites that the organism exchanges with its environment are then defined and flux through an objective reaction describing the organism's assumed

evolutionary goal is maximised to obtain a flux distribution, an approach called flux balance analysis (FBA)[10,13]. Metabolic models were initially reconstructed for well-investigated bacteria such as *Escherichia coli* and focused on describing growth under defined conditions, mainly single-strain cultivations in laboratory media[14]. Advances in both computational and experimental methods have generated a vast array of toolboxes to efficiently perform model reconstruction and analysis[15–18], which have significantly expanded the scope of metabolic network reconstructions. Metabolic models can now be reconstructed for eukaryotic microorganisms[19,20], human metabolism[21], and host-symbiont interactions as found in the human gut microbiome or rhizobium-legume symbioses[8,9,22–24].

## 1.2 Motivation

In order to optimally exploit biological nitrogen fixation for agricultural practices, the efficiency of natural symbioses needs to be improved[25] and synthetic symbioses with important crops such as wheat or corn need to be engineered using synthetic biology approaches[26–28]. In addition, rhizobial strains that are effective when applied in the field need to be competitive, i.e., they have to be able to nodulate the target crop in the presence of other soil bacteria that are potentially less efficient at fixing nitrogen[29,30]. A detailed understanding of the metabolic processes at the different stages of symbiosis formation is required to support research efforts in all of these areas. Studies with isolated mutant strains have generated major insights into the metabolism of bacteroids as well as metabolic characteristics that are important for competitive growth in the rhizosphere and root colonization. However, bacteroids isolated from root nodules are fragile and nitrogenase activity of nodules ceases

quickly when they are separated from the plant root, which has complicated direct measurements of metabolic fluxes in bacteroids[6,31]. To complement and direct targeted investigation of individual metabolic processes, high-throughput experimental techniques for determining gene expression, protein abundances or gene essentiality are increasingly applied to rhizobial strains both in free-living conditions and at different stages of interaction with host and non-host plants[32–39]. Metabolic models provide a powerful tool for interpreting these data and enable a systems-level analysis of metabolic processes without the difficulty of experimental effort required for metabolic flux analysis based on $^{13}C$ labelling[40–43]. While several metabolic modelling studies of rhizobia have been published[9,22,44–46], the scope and level of curation of these reconstructions vary substantially. An additional limitation is the common use of objective functions that are lumped reactions comprising the secretion of ammonia and amino acids as well as production of storage polymers. This approach introduces a stoichiometric coupling between several metabolic processes which is not supported by experimental data. Finally, a significant challenge in modelling rhizobial metabolism in rhizobium-legume interactions is the incomplete knowledge about which metabolites are exchanged with the plant[1,5]. Further modelling studies that move beyond traditional FBA methods and integrate experimental datasets to improve the accuracy of model predictions would therefore be beneficial to improve the current understanding of rhizobial metabolism and guide the design of experimental studies.

## 1.3 Aims

In view of the limitations of experimental approaches described above, this work aims to improve the current understanding of rhizobium-legume interactions at various stages of the symbiosis by combining highly curated metabolic models for rhizobia with various omics datasets, such as transcriptomics, proteomics, and metabolomics. All modelling studies performed as part of this work aim to circumvent the limitations of most currently available bacteroid models by minimizing the number of assumptions about nutrient uptake and investigating alternatives to the use of objective functions that are not supported by experimental evidence.

Since bacteroids are arguably both the most challenging and the most important stage of the symbiosis to investigate, special consideration is given to developing a detailed model of bacteroid metabolism. In particular, several aspects of central carbon and nitrogen metabolism in bacteroids have so far been investigated separately in experimental studies, and a comprehensive model of the interconnection of these pathways is developed to facilitate an integrated interpretation of experimental results. Finally, because it is known that substantial strain- and host plant-specific differences between symbioses exist, unified metabolic models for different rhizobial strains are used to elucidate some of these differences.

Overall, this work aims to generate novel insights and frameworks for understanding rhizobial metabolism that will accelerate experimental efforts to engineer more efficient and/or synthetic rhizobium-legume interactions.

## 1.4 Thesis outline

The main content of this thesis is organised into scientific research papers, which have been published or submitted, or are in preparation as indicated.

Chapter 2 consists of the minireview "How rhizobia adapt to the nodule environment"[5]. This chapter provides a general overview of symbiosis formation in rhizobium-legume interactions. Guided by a meta-analysis of transcriptome and proteome datasets for various rhizobial species, key characteristics of bacteroid physiology are described. The review further explains the current knowledge of bacteroid metabolism and nutrient exchanges with the plant and summarises metabolic modelling studies of rhizobial species that have been published to date.

In Chapter 3, I present a genome-scale metabolic model for *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841), a model strain for investigating rhizobium-legume symbioses. Following extensive curation, I integrate transcriptomics, proteomics, metabolomics and gene essentiality data with the model to elucidate nutrient uptake and metabolic characteristics of Rlv3841 in the rhizosphere, as a free-living nodule bacterium, and as a nitrogen-fixing bacteroid. The manuscript corresponding to this chapter has been submitted[47] and is currently under review.

Chapter 4 contains a detailed investigation of key constraints governing rhizobial metabolism during nitrogen fixation. Based on a reduced model for bacteroid metabolism in Rlv3841, I investigate the role of carbon polymer synthesis and amino acid secretion by bacteroids in the microaerobic conditions required for nitrogen

fixation. I further compare the consequences of using dicarboxylates rather than sugars as a carbon source to support symbiotic nitrogen fixation, in particular with regard to secretion of fixed nitrogen to the plant and electron supply to the nitrogenase enzyme. Model predictions are validated through experimental determination of oxygen uptake rates and $NADH/NAD^+$ ratios and further corroborated by metabolic flux analysis, which was performed by collaborators. This chapter has been published[39].

In Chapter 5, I perform homology-based reconstruction of genome-scale models for three *Rhizobium* strains and one *Sinorhizobium* strain. The models are curated and reconciled to enable direct comparison of model predictions. By integrating transcriptome data with the genome-scale models, I obtain bacteroid models for six different rhizobium-legume symbioses. Model predictions are compared to outline shared metabolic properties and differences between symbioses, in particular highlighting the characteristics that distinguish terminally differentiated bacteroids from reversibly differentiated bacteroids.

In Chapter 6, I summarise the main findings of this thesis and provide an outlook for future studies.

## 1.5 Scientific contributions

The following publications are the main scientific output of my doctoral research and are presented as part of this thesis:

- R. Ledermann*, C. C. M. Schulte*, P. S. Poole, How rhizobia adapt to the nodule environment. *Journal of Bacteriology* **203** (12), e00539 – 20 (2021). https://doi.org/10.1128/JB.00539-20

  *equal contribution

- C. C. M. Schulte*, K. Borah* *et al.*, Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Science Advances* **7** (31), eabh2433 (2021). https://doi.org/10.1126/sciadv.abh2433

  *equal contribution

- C. C. M. Schulte *et al.* Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum, submitted.*

  Preprint: https://doi.org/10.1101/2021.07.28.454262

- C.C. M. Schulte *et al.* Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses, *in preparation*.

I have further contributed to the following studies, which are not presented in this thesis:

- E. Alexis, C. C. M. Schulte *et al.*, Biomolecular mechanisms for signal differentiation, *under revision.*

  Preprint: https://doi.org/10.1101/2021.04.29.441952

- N. Delalez*, A. Sootla*, R. Habgood*, C. C. M. Schulte *et al.* Design of a synthetic sRNA-based feedback filter module, *in preparation*.

  *equal contribution

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Bacteriology®

MINIREVIEW

## 2. Background

# How Rhizobia Adapt to the Nodule Environment

Raphael Ledermann,[a] Carolin C. M. Schulte,[a,b] Philip S. Poole[a]

[a]University of Oxford, Department of Plant Sciences, Oxford, United Kingdom
[b]University of Oxford, Department of Engineering Science, Oxford, United Kingdom

Raphael Ledermann and Carolin C. M. Schulte contributed equally to this work. Author order was determined by alphabetical order of surnames.

**ABSTRACT** Rhizobia are a phylogenetically diverse group of soil bacteria that engage in mutualistic interactions with legume plants. Although specifics of the symbioses differ between strains and plants, all symbioses ultimately result in the formation of specialized root nodule organs that host the nitrogen-fixing microsymbionts called bacteroids. Inside nodules, bacteroids encounter unique conditions that necessitate the global reprogramming of physiological processes and the rerouting of their metabolism. Decades of research have addressed these questions using genetics, omics approaches, and, more recently, computational modeling. Here, we discuss the common adaptations of rhizobia to the nodule environment that define the core principles of bacteroid functioning. All bacteroids are growth arrested and perform energy-intensive nitrogen fixation fueled by plant-provided $C_4$-dicarboxylates at nanomolar oxygen levels. At the same time, bacteroids are subject to host control and sanctioning that ultimately determine their fitness and have fundamental importance for the evolution of a stable mutualistic relationship.

**KEYWORDS** rhizobium-legume symbiosis, nitrogen fixation, root nodule, microaerobiosis, growth arrest, metabolism, modeling, host sanctioning

Plants are dependent on nutrients such as phosphorus, nitrogen, and sulfur to be present in a bioavailable form, which often limits plant growth. To facilitate nutrient uptake, mutualistic symbioses between plants and root-associated microorganisms have evolved. The symbiosis with arbuscular mycorrhiza (AM) originates back to the Devonian period more than 400 million years ago (1). The same signaling pathway that facilitated AM symbioses has since been adopted for other symbionts and is shared among all intracellular root symbionts (2, 3). This includes various root nodule symbioses (4, 5) that evolved around 110 million years ago (6), where the microsymbionts are either actinomycetes or rhizobia. These diazotrophic bacteria fix atmospheric $N_2$ gas into bioavailable ammonia. Biological nitrogen fixation accounts for approximately 50% of the bioavailable nitrogen on earth (7) and is of great importance for sustainable agriculture and the protection of susceptible ecosystems from leached fertilizers (8, 9).

The lack of genetic tools for the microsymbiont has hampered the investigation of actinorhizal symbioses (10), making rhizobium-legume symbioses the most researched and best-understood interactions so far (11). The symbioses between rhizobia (belonging to alpha- and betaproteobacteria [12]) and legumes (Fabaceae) exist with a variety of different infection mechanisms and nodule morphologies (13–15). Although bacterial entry via cracks in the root epidermis and Nod factor-independent infection processes exist (16–19), the best-studied mechanism is root invasion through root hairs via infection threads (ITs). Most major crop legumes and many model plants such as *Medicago* spp., *Pisum sativum* (pea), *Lotus japonicus*, *Phaseolus vulgaris* (bean), and *Glycine max* (soybean) follow this type of infection (20–26). Soil-dwelling rhizobia sense flavonoids exuded by the plant (27), and compatible flavonoids trigger the production of Nod factors (lipochitooligosaccharides) (for reviews, see references 28 and 29). Nod

**Citation** Ledermann R, Schulte CCM, Poole PS. 2021. How rhizobia adapt to the nodule environment. J Bacteriol 203:e00539-20. https://doi.org/10.1128/JB.00539-20.

**Editor** William Margolin, McGovern Medical School

**Copyright** © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Raphael Ledermann, raphael.ledermann@plants.ox.ac.uk.

**Accepted manuscript posted online** 1 February 2021

**Published** 20 May 2021



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

factor signaling leads to the induction of nodule primordia in the root cortex and causes the root hairs to curl their tips, resulting in the characteristic shape called "shepherd's crook." Bacteria attached to root hairs may be entrapped by this curling and form a microcolony from which they enter the root hair via an IT that is formed by the inverted growth of the plant cell wall and membrane. Since the microcolony is derived from one or very few founder cells, ITs are a major bottleneck for bacteria during invasion (30, 31).

Inside ITs, bacteria are thought to move mainly by cell division and potentially some gliding motility (32, 33). Eventually, the ITs reach the preformed nodule primordia, and rhizobia are taken up into individual plant cells by endocytosis, enclosed by a symbiosome membrane of plant origin. There, they multiply until the plant cell is densely packed with bacteria, which differentiate into their nitrogen-fixing form called a bacteroid. Nodules usually have one of two distinctive morphologies, depending on the host plant (34, 35): (i) determinate nodules, lacking a persistent meristem, therefore ceasing to grow at some point and forming spherical nodules, or (ii) indeterminate nodules with a persistent meristem, resulting in indefinite growth, branching, or lobe formation of the nodule, giving rise to elongated or irregular shapes. The detailed infection processes and differences between determinate and indeterminate nodules have been extensively discussed elsewhere (36–39).

## COMPETITIVENESS

Host plants can influence the composition of the root microbiome (40, 41), but rhizobia must compete with various other bacteria for root colonization and with other compatible rhizobial strains for nodulation. Competitiveness is a complex trait: different abiotic factors such as soil pH (42–44) or nutrient availability (45) have an impact on the outcome of symbioses. Biotic factors also affect competitiveness, for example, the host-symbiont pairing (46–48), plant- or strain-intrinsic factors such as type VI secretion systems (49, 50), exopolysaccharide (EPS) production (51), and catabolic capacity in rhizobia for various substrates such as *myo*-inositol (52–54), glycerol (55), arabinose, protocatechuate (56), rhamnose (57), homoserine (58), or erythritol (59). During the transition from a soil-dwelling bacterium to a nitrogen-fixing bacteroid, rhizobia encounter drastically changing conditions, but it has not always been addressed at which stage of symbiosis formation a mutant strain is affected. Competitiveness in the rhizosphere is largely determined by metabolic functions and motility (48). Transcriptomics and mutant studies have shown that *Rhizobium leguminosarum* depends on a variety of substrates such as $C_4$-dicarboxylates, $C_2$-organic acids, and aromatic compounds for growth in the rhizosphere. However, sugar and sugar alcohol transport systems were similarly induced, showing that metabolism in the rhizosphere is highly complex (60). It is well documented that strains and mutants can be competitive in rhizosphere colonization but nonetheless lack competitiveness for nodulation (52, 61, 62), highlighting the need for changing adaptations throughout the infection process.

Insertion sequencing (INSeq) experiments using saturated random transposon mutant libraries identified essential genes in *R. leguminosarum* during four distinct stages of the symbiosis: rhizosphere growth, root attachment, nodule bacteria (viable bacteria within nodules), and differentiated bacteroids (63). In these experiments, bacteria were under selective pressure by pea plants, and mutants lacking competitiveness were lost at the affected and subsequent stages. Interestingly, while several genes were associated with competitiveness in the rhizosphere and root attachment, they were not necessarily involved in competitive nodulation, further indicating that persistence on the root surface is a distinct trait. In total, 390 genes were associated with competitive nodulation. Among these, 146 were already needed in the rhizosphere, 33 were needed in the root-attached stage, and 211 were needed in ITs and predifferentiated bacteroids. Functions related to competitive nodulation from the rhizosphere onward include lipopolysaccharide (LPS) and EPS biosynthesis, the NtrBC system and the $\sigma^{54}$ factor RpoN, and the stringent response. LPS and EPS modifications have a known role in evading the plant immune response during IT formation (64–70), and the stringent response has been shown to be involved at multiple stages during the *Sinorhizobium-Medicago* symbiosis (71–73). During later infection stages, allophanate, aldehyde, erythritol metabolism,



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

and glycogen synthesis play important roles. Interestingly, in agreement with previous screens (48), almost all mutations causing auxotrophies lacked competitiveness, indicating that *de novo* synthesis of essential building blocks is crucial even when they are exogenously present in root exudates (74, 75).

The large number of genes associated with nodulation found outside the symbiotic plasmids in *R. leguminosarum* (63) is consistent with the findings for the chimeric *Ralstonia solanacearum/Cupriavidus taiwanensis-Mimosa pudica* experimental evolution of a symbiosis (76). Although the initial chimera was unable to form nodules, nodulation was achieved and competitiveness was gradually increased with repeated inoculation cycles on *M. pudica* seedlings. Strikingly, the ability to nodulate was directly linked with competitiveness (77). After initial infection was achieved by a loss of pathogenicity, a major improvement in competitiveness was seen when additional mutations in global transcriptional regulators enhanced intracellular colonization and persistence (78, 79), further highlighting that the ability to competitively nodulate legume plants heavily relies on the global rewiring of bacterial gene expression outside the acquired classical symbiosis genes (*nod*, *nif*, *fix*, etc.) found on mobile symbiotic plasmids or islands (80–84).

Notably, competitiveness for initial nodulation (but not persistence) is not affected by a strain's efficiency in nitrogen fixation (85). Under field conditions, indigenous rhizobia with an inferior fixation capacity may outcompete the "elite" rhizobial strains inoculated on the legume seeds, resulting in suboptimal yields (86, 87). The dependence on environmental parameters and the overall complexity based on symbiotic and general-function genes (88) make competitive strains with high nitrogen fixation efficiencies difficult to identify. A recently reported barcoded and $P_{nifH}$-driven superfolder green fluorescent protein (sfGFP)-based toolkit allows the simultaneous high-throughput analysis of a strain's competitiveness and nitrogen fixation rate (89). It enabled the identification of highly effective strains by measuring fixation rates of individual strains in single nodules. This approach facilitates the parallelized inoculation of multiple strains and the simultaneous assessment of their nodulation competitiveness and nitrogen fixation ability.

## ADAPTATIONS TO NITROGEN FIXATION

Root nodules are *de novo* plant organs that share root- and stem-typical traits (90) and are formed only in the presence of compatible rhizobia. Their specialized function—to host intracellular bacteria for nitrogen fixation—makes the conditions inside nodules unique. Nodules are immunocompromised compartments with an immune response that differs from that of the rest of the plant, but they are also confined and insulated from the root, allowing them to host large numbers of intracellular bacteria (91). Despite some distinguishing properties specific to different rhizobia and their plant hosts, all rhizobium-legume symbioses share some universal characteristics. With the goal of highlighting these core changes defining the transition from free-living bacteria to nitrogen-fixing bacteroids, we performed a meta-analysis of 18 transcriptome and 1 proteome data sets published in 13 different studies on various symbioses (92–104) (Table 1). All genes/proteins were classified according to the Clusters of Orthologous Genes (COG) system (105) to enable interspecies comparisons.

Overall, most of the subsystems contained more downregulated than upregulated genes, indicating a reduced physiological complexity in bacteroids (Fig. 1). Notable exceptions to the general downregulation were "energy production and conversion" and "inorganic ion transport and metabolism" (see Table S1 in the supplemental material). Upregulated features in these subsystems comprised nitrogenase components (*nif* genes) and electron transfer proteins (including *fix* genes) such as ferredoxins, which act as electron donors for nitrogenase (106, 107) (Table S2). Free-living diazotrophs such as *Klebsiella* regulate *nif* gene expression in response to nitrogen starvation and low oxygen concentrations (108), whereas the nitrogen status in many bacteroids plays either no or only a minor role in *nif* gene activation, which is mainly induced by low oxygen concentrations (see Adaptations to Microaerobic Conditions, below). All rhizobia regulate *nif* and *fix* gene expression using similar sets of oxygen-responsive pathways (FixLJ or hFixL FxkR, FixK/Fnr,



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

**TABLE 1** Data sets used in the meta-analysis[a]

| Species (rhizobium strain-plant host) | Reference data set | Bacteroids terminally differentiated | Reference |
|---|---|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841-*Pisum sativum* | Bacteria grown in MM (mid-exponential phase) | Yes | 98 |
| *Rhizobium leguminosarum* bv. *viciae* 3841-*Vicia cracca* | Bacteria grown in MM (mid-exponential phase) | Yes | 98 |
| *Rhizobium leguminosarum* bv. *viciae* A34-*Pisum sativum* | Bacteria grown in MM (mid- to late exponential phase) | Yes | 99 |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292-*Phaseolus vulgaris* | Bacteria grown in MM (mid- to late exponential phase) | No | 99 |
| *Rhizobium etli* CFN42-*Phaseolus vulgaris* | Bacteria grown in MM (early exponential phase) | No | 92 |
| *Rhizobium etli* CFN42-*Phaseolus vulgaris* | Bacteria grown in MM (stationary phase) | No | 92 |
| *Sinorhizobium meliloti* 2011-*Medicago truncatula* | FIId: distal fraction of ZII; contains cells during infection and differentiation | Yes | 102 |
| *Sinorhizobium meliloti* 1021-*Medicago truncatula* | Bacteria grown in CM (mid-exponential phase) | Yes | 95 |
| *Sinorhizobium meliloti* 1021-*Medicago sativa* cv. Europe | Bacteria grown in MM (mid-exponential phase) | Yes | 104 |
| *Sinorhizobium medicae-Medicago truncatula* | FI: nonfixing zone of the nodule | Yes | 97 |
| *Sinorhizobium* sp. strain NGR234-*Vigna unguiculata* | Bacteria grown in CM (mid-exponential phase) | No | 101 |
| *Sinorhizobium* sp. NGR234-*Leucaena leucocephala* | Bacteria grown in CM (mid-exponential phase) | No | 101 |
| *Bradyrhizobium* sp. ORS285-*Aeschynomene afraspera* | Bacteria grown in CM (mid-exponential phase) | Yes | 100 |
| *Bradyrhizobium* sp. ORS285-*Aeschynomene indica* | Bacteria grown in CM (mid-exponential phase) | Yes | 100 |
| *Bradyrhizobium japonicum* USDA110-*Glycine max* | Bacteria grown in CM (mid-exponential phase) | No | 96 |
| *Bradyrhizobium japonicum* USDA110-*Vigna unguiculata* | Bacteria grown in CM (mid-exponential phase) | No | 103 |
| *Bradyrhizobium japonicum* USDA110-*Macroptilium atropurpureum* | Bacteria grown in CM (mid-exponential phase) | No | 103 |
| *Azorhizobium caulinodans* ORS571-*Sesbania rostrata* | Bacteria grown in MM (mid- to late exponential phase) | No | 94 |
| *Mesorhizobium huakuii* 7653R-*Astragalus sinicus* | Bacteria grown in CM (late exponential phase) | Yes | 93 |

[a]Transcriptome data were used in all cases apart from the *Sinorhizobium medicae* study, which is a proteome data set. All data sets compare bacteroids with the reference conditions described in the second column. MM, minimal medium; CM, complex medium.

and NifA). The pathways differ in their sensitivities to oxygen (109), leading to staggered activation and thus a gradual adaptation of the invading rhizobia to the microoxic nodule environment (110). However, each species uses them in a slightly different way, and the interconnections between the pathways differ as well, a topic that has been discussed elsewhere (111–115). In addition to *nif* genes, the expression of iron and/or molybdenum transporters was increased across the various data sets, consistent with the requirement for these metals as cofactors of the nitrogenase enzyme (116).

General upregulation was further found for chaperones, indicating a response to stress factors in nodules and restructuring of the bacteroid proteome. This agrees with the established importance of chaperones in bacteroids. Nodules formed by a *clpB* mutant of *Mesorhizobium ciceri* contained only a few bacteroids (117), and GroESL chaperones were found to be essential for symbiosis (118, 119) and to interact with nodule-specific cysteine-rich (NCR) peptides (120) (see Adaptations to a Nongrowing State, below). Further examples of stress responses include glutathione *S*-transferases, which were consistently upregulated in bacteroids. Mutants unable to produce glutathione, an important antioxidant molecule, showed reduced symbiotic efficiency (121, 122). Similarly, enzymes involved in the detoxification of reactive oxygen species, such as superoxide dismutases and catalases, were commonly upregulated, in agreement with the presence of reactive oxygen species due to auto-oxidation of leghemoglobin and ferredoxin (123).

## ADAPTATIONS TO MICROAEROBIC CONDITIONS

The iron-sulfur clusters of the nitrogenase enzyme are highly susceptible to molecular oxygen (124), and the nodule cortex therefore contains an oxygen diffusion barrier. This limits the amount of free oxygen within the nodule, creating a microoxic



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

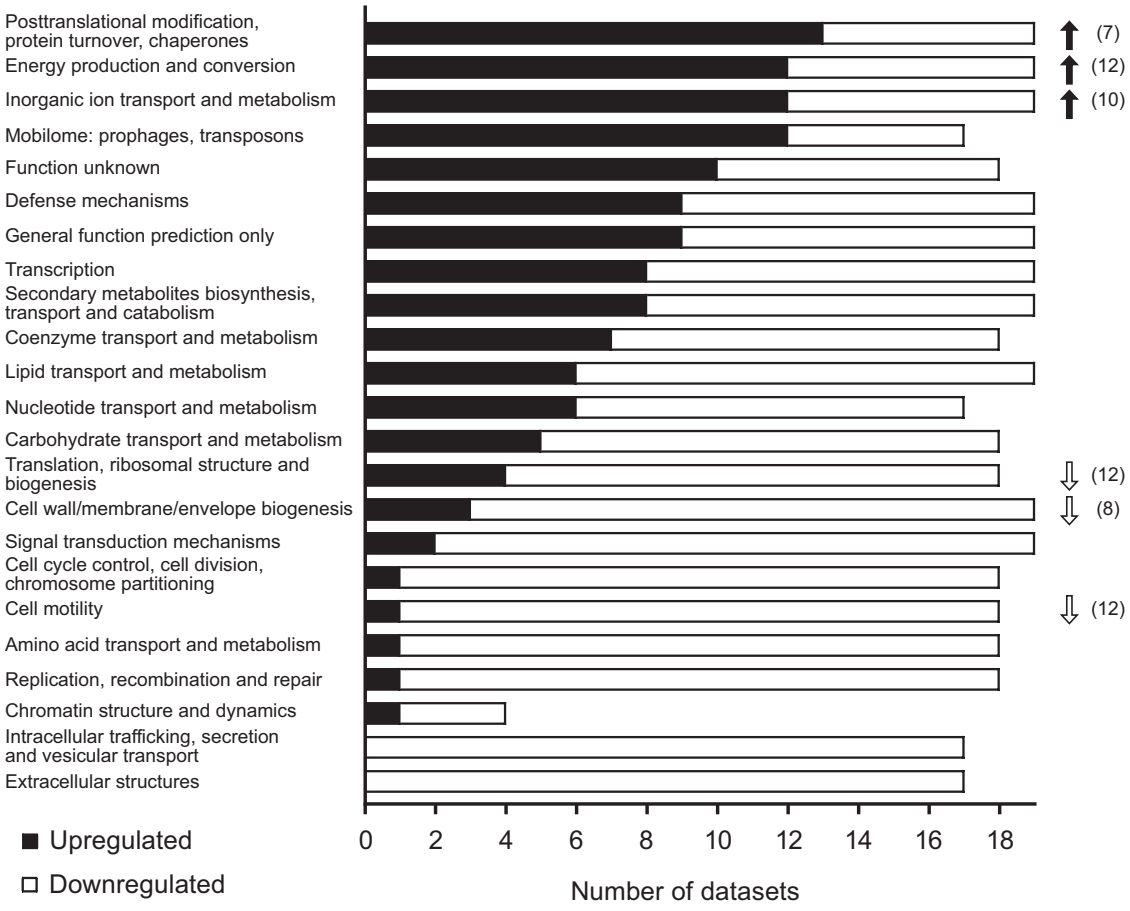



**FIG 1** Central features of bacteroids determined from genome scale data sets. A meta-analysis of 18 transcriptome and 1 proteome data sets for bacteroids of various rhizobial strains was performed. The bar graph shows the number of data sets in which a COG category contained more upregulated than downregulated (black) or more downregulated than upregulated (white) genes/proteins. Where the total number is <19, some data sets contained either no differentially expressed genes/proteins or equal numbers of up- and downregulated genes/proteins in this category. Arrows indicate the 3 categories that were most often significantly enriched within upregulated (black) or downregulated (white) genes/proteins, with numbers in parentheses indicating the number of data sets for which significant up/downregulation was observed. *P* values were determined with a hypergeometric test followed by Bonferroni multiple-test correction, and a *P* value of <0.05 was considered significant. For the same analysis separated according to terminally differentiated and not terminally differentiated bacteroids, see Fig. S1 in the supplemental material.

environment of around 11 nM free oxygen (112, 125). For comparison, the freely diffused oxygen concentration in water at atmospheric oxygen levels is around 255 $\mu$M. Nitrogen fixation is an energy-intensive process, and energy in biological systems is most efficiently generated by respiratory chains with molecular oxygen as the final electron acceptor (126). To circumvent this apparent paradox, where oxygen is needed to energize nitrogen fixation and nitrogenase is sensitive to oxygen, rhizobia possess high-affinity terminal oxidases that allow respiration at low oxygen concentrations. Rhizobia can usually express a complex set of different respiratory oxygen reductases, reflecting their diverse lifestyles (127–130). Biochemical and genetic data show that a cytochrome *c* oxidase, encoded by *fixNOQP*, is a high-affinity $cbb_3$-type oxidase responsible for respiration under microaerobic conditions in rhizobia. Biochemical assays revealed a $K_m$ value of 4 to 7 nM for oxygen bound to membranes of anaerobically grown bradyrhizobia, corresponding to the $cbb_3$-type cytochrome oxidase FixNOQP (131). These values are compatible with the low oxygen concentrations in nodules. Mutants of *fixNOQP* in *Bradyrhizobium japonicum* had only marginal nitrogenase activity (132, 133), whereas multiple copies of *fixNOQP* are present in the genomes of *Mesorhizobium loti*, *Sinorhizobium meliloti*, *Rhizobium etli*, and *R. leguminosarum*. In *S. meliloti* and *R. leguminosarum*, the two copies are functionally redundant, but



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

only one copy is functional during symbiosis in *R. etli* (134–136). Mutation of the (single-copy) $cbb_3$-type oxidase in *Azorhizobium caulinodans* resulted in only a 50% reduction of nitrogenase activity in symbiosis (137). The remaining activity could be abolished by creating a double mutant with a *bd*-type quinol oxidase. This type of oxidase is crucial for nitrogen fixation in the free-living diazotrophs *Klebsiella oxytoca* (138) and *Azotobacter vinelandii* (139). Recently, genome sequences of betaproteobacterial rhizobia showed that *Paraburkholderia* strains lack *fixNOQP* but contain *bd*-type quinol oxidase genes. However, functional studies are missing (140). In accordance with the described adaptations to the microaerobic nodule environment, several cytochromes and high-affinity cytochrome oxidases showed upregulation in our meta-analysis. Interestingly, subunits of the ATP synthase were mostly downregulated, indicating that the overall energy demand in bacteroids is lower than that in exponentially growing cells.

Nodules have a characteristic red color due to the accumulation of leghemoglobins. These plant-derived proteins are found exclusively in the plant cytoplasm and provide buffering capacity for free oxygen as well as facilitating the diffusion of oxygen toward symbiosomes (141–144). Mutation or silencing of leghemoglobin biosynthesis resulted in inefficient symbioses (145, 146). The free oxygen concentration was, however, only marginally higher in mutant than in wild-type nodules. Although there was a slight increase in oxygen available for respiration, the ATP/ADP ratio dropped compared to the wild type (146). Thus, the main function of leghemoglobins is not to keep free oxygen concentrations low (which is achieved by the oxygen diffusion barrier) but rather to enhance the oxygen supply to symbiosomes, which is limiting because of the diffusion barrier. In this function, leghemoglobins are equivalent to myoglobins in animals.

## ADAPTATIONS TO A NONGROWING STATE

In addition to changes induced by microoxic conditions, bacteroids are characterized by their nongrowing state. Growth arrest in free-living bacteria is usually induced by unfavorable conditions such as starvation, oxygen depletion, or antibiotic substances. Accordingly, the main adaptations to such conditions focus on persistence, i.e., maintaining low levels of proton motive force, synthesizing integral cell building blocks, slowing metabolism, and catabolizing nonessential endogenous compounds, which enables bacteria to stay viable for months or years (147).

Commonly downregulated traits in the analyzed data sets that are shared with growth-arrested cells (92) include translation (ribosomal proteins), cell envelope biogenesis (outer membrane proteins, peptidoglycan, and exopolysaccharide synthesis), intracellular trafficking, cell cycle control and DNA replication, signal transduction, motility, and the $F_oF_1$-type ATPase (see Table S3 in the supplemental material). The ribosomal silencing factor RsfS, which inhibits the assembly of the 30S and 50S ribosomal subunits (148, 149), was generally upregulated, indicating additional downregulation of protein biosynthesis. Furthermore, ClpA, a subunit of the ClpAP protease, known for its role in cell cycle control and stationary-phase adaptation (150–152) but also potentially part of FixK regulation (153), was commonly upregulated in bacteroids. These findings suggest that a major adaptation to the nodule environment for bacteroids is related to growth arrest. Strikingly, the experimental evolution experiments described above found mutations in *efpR* to be associated with adaptation to intracellular colonization (78). EfpR was found to act as a positive regulator of several 30S ribosomal protein subunits and as a negative regulator of the adapter protein ClpS (78), which redirects protein aggregates to ClpAP (154), mimicking the bacteroid growth arrest adaptation.

In contrast to free-living growth-arrested bacteria, however, bacteroids are not nutrient starved, are well adapted for respiration under microoxic conditions, and retain high levels of metabolic activity to sustain nitrogen fixation. It thus remains unknown how and why bacteroids stop dividing. Some legumes induce terminal differentiation and chromosomal endoreduplication in bacteroids, meaning that the bacteroid is unable to dedifferentiate into a free-living bacterium (155, 156). Generally, hosts with determinate nodules (e.g., *Lotus*, soybean, and bean) harbor bacteroids that are similar



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

to free-living bacteria, and differentiation is not terminal (U morphotype) (157). Legumes of the inverted-repeat-lacking clade and some *Aeschynomene* spp. contain enlarged and elongated bacteroids (E morphotype) that sometimes branch and become Y-shaped or irregularly shaped. Many legumes of the Dalbergioid cluster (e.g., peanut and *Aeschynomene* spp.) cause bacteroids to become enlarged and spherical (S morphotype). Rhizobia capable of nodulating different host plants will adopt the bacteroid morphotype of the respective host (31, 158, 159). Enlargement of the bacteroids in the case of E and S morphotypes is accompanied by endoreduplication. This differentiation leads to up to 24C (chromosomes) in E-morphotype *S. meliloti* bacteroids in symbiosis with *Medicago* (155). Similarly, *Bradyrhizobium* sp. strain ORS285 bacteroids reach 7C in *Aeschynomene afraspera* (E morphotype) and 16C in *Aeschynomene indica* (S morphotype) (160).

The key factor for bacterial endoreduplication is NCR peptides. These consist of 60 to 90 amino acids and display similarities with plant defensins (161, 162). NCR peptides are abundant in hosts harboring swollen bacteroids (around 700 are known in *Medicago truncatula*) but evolved independently in different clades (155, 156, 163). They are exclusively expressed in bacteroid-containing cells and are targeted to the bacteroid (164). Plant mutants with defects in the NCR peptide secretory pathway form ineffective nodules (164–166). Some NCR peptides have antimicrobial activities (167, 168), and rhizobia with increased sensitivity to NCR peptides form ineffective symbioses because they are rapidly killed after release from ITs into plant cells (160, 169–172). At the right dosage, however, NCR peptides will promote endoreduplication, even in free-living cultures (173), although the mechanisms by which they interfere with the bacterial cell cycle are not yet fully understood (167). The mode of action of some NCR peptides has been identified, for example, inhibition of FtsZ (the structural component of the Z ring involved in cell division) and ribosomal proteins (120). Moreover, rhizobia are successively exposed to different NCR peptides, suggesting that they act in an orchestrated manner, controlling various steps of rhizobial physiology (173–175). NCR peptide-mediated endoreduplication and terminal differentiation might contribute to the growth-arrested phenotype of bacteroids. However, NCR peptides are absent in hosts harboring unswollen (U morphotype) bacteroids (176), and endoreduplication and swelling of bacteroids happen only after bacteroids stop dividing (159), indicating the presence of other mechanisms that are yet to be discovered.

## METABOLIC ADAPTATIONS

As a result of the adaptations required for symbiotic nitrogen fixation, bacteroid metabolism is unique in several respects: there is a strong downregulation of most biosynthetic functions compared to growth in the free-living state, while specialized enzymes for nitrogen fixation are highly induced (98, 177). In contrast to free-living bacteria in stationary phase, bacteroid metabolism is highly active due to the energy requirement of nitrogen fixation despite the microoxic environment. In addition, bacteroid metabolism is closely interdependent with the plant host because of bidirectional nutrient exchanges between the symbiotic partners (Fig. 2) (116, 178).

An important implication of nutrient provision by the plant is that downregulation in biosynthetic pathways can occur either because a compound is required only in small quantities in bacteroids or because it is provided by the plant. Multiple auxotrophic rhizobial mutants with a $Fix^+$ phenotype are known, indicating that plants provide the respective compounds to their bacteroids (179). A striking illustration of this concept is symbiotic auxotrophy: mutants lacking the general amino acid permeases Aap and Bra in *R. leguminosarum* were severely impaired in their symbiotic efficiency, indicating a requirement for plant supply of amino acids (180). While free-living rhizobia are capable of synthesizing all proteinogenic amino acids, downregulation in branched-chain amino acid synthesis occurs in bacteroids, making them dependent on the supply of branched-chain amino acids by their host plant (181, 182). A similar downregulation was observed in our meta-analysis for serine biosynthesis. Indeed, a serine auxotrophic mutant of *R. etli* was fully proficient in symbiosis (183). Yet cysteine biosynthesis, which requires serine as a precursor, was one of the few upregulated



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

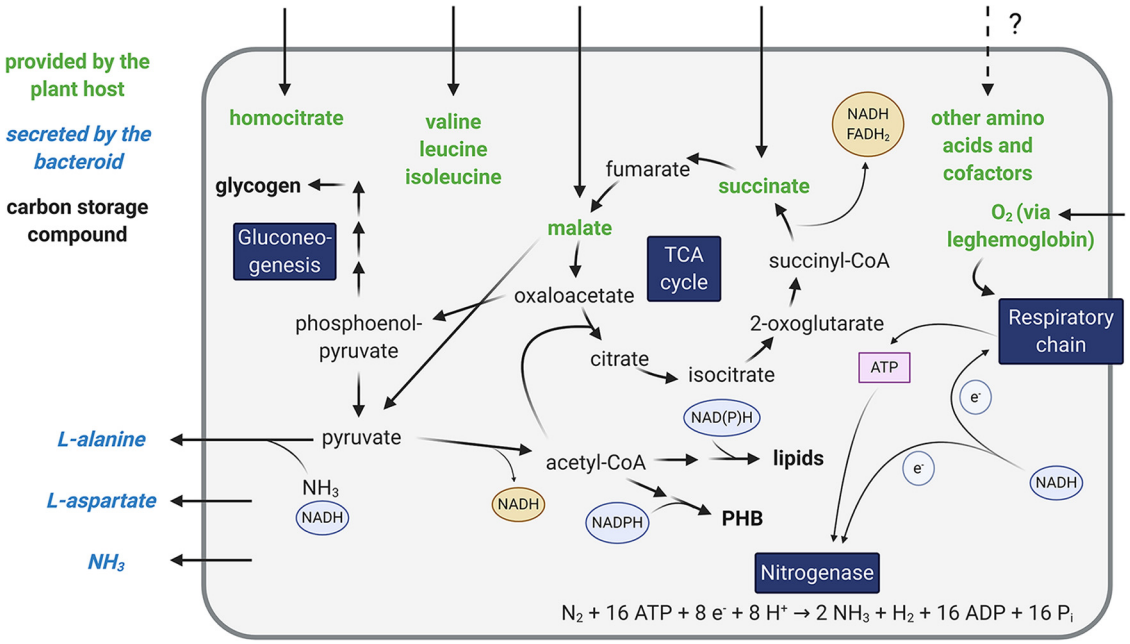



**FIG 2** Central metabolic and transport reactions in bacteroids. Shown is a schematic map of major metabolic pathways and transport reactions in bacteroids as well as important flows of electrons and ATP. This figure was created with BioRender.

pathways for amino acid biosynthesis. This upregulation is likely related to the production of iron-sulfur clusters for the nitrogenase enzyme, which starts with cysteine as a sulfur donor (184). It is therefore possible that plant hosts supply serine to bacteroids to assist with cysteine biosynthesis for nitrogenase assembly. Experimental evidence also supports the provision of amino acids by the plant host that are nonessential for bacteroid functioning. In pea, enzymes catabolizing $\gamma$-aminobutyric acid (GABA) are highly induced in bacteroids, and labeling studies indicated that GABA is supplied by the plant, but mutants of *R. leguminosarum* unable to catabolize GABA differentiated into functional bacteroids (185).

Apart from amino acids, a different example of plant-derived metabolites essential for nitrogenase activity is homocitrate, a component of the iron-molybdenum cofactor. In contrast to free-living diazotrophs, most symbiotic rhizobia are unable to synthesize homocitrate, and a homocitrate synthase mutant (Fen1) of *L. japonicus* was incapable of forming effective nodules (186). These observations illustrate the intricate control that plants exercise over their bacterial symbiont.

Legumes provide bacteroids with $C_4$-dicarboxylates as the main carbon source, primarily malate and succinate (187–189). Catabolism of $C_4$-dicarboxylates requires their conversion to both oxaloacetate and acetyl coenzyme A (acetyl-CoA), which are condensed to form citrate in the first step of the tricarboxylic acid (TCA) cycle. There are two pathways for generating acetyl-CoA from TCA cycle intermediates: (i) malate is oxidized to pyruvate by malic enzyme (ME), or (ii) phosphoenolpyruvate carboxykinase (PCK) converts oxaloacetate into phosphoenolpyruvate, which is converted into pyruvate by pyruvate kinase (PK). In both cases, acetyl-CoA is obtained from pyruvate via the pyruvate dehydrogenase complex. While all bacteroids are required to convert TCA cycle intermediates into acetyl-CoA, the synthesis routes differ. Bacteroids of *S. meliloti* and *A. caulinodans* exclusively use ME (190, 191), whereas *Sinorhizobium fredii* and *R. leguminosarum* utilize both ME and PCK/PK (191, 192). In addition, the activity of PCK is important for bacteroids to sustain gluconeogenesis, which is essential for the production of precursors for various metabolites (63, 193). The TCA cycle is the main pathway for dicarboxylate metabolism in bacteroids, and at least some TCA cycle enzymes were upregulated in the transcriptome/proteome data sets across different rhizobial species.

Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

However, studies with individual mutants have produced contradictory evidence on which parts of the TCA cycle are required for nitrogen fixation. Mutants in different TCA cycle enzymes were Fix$^-$ in *S. meliloti*, *S. fredii*, *Rhizobium tropici*, and *R. leguminosarum* (194–198), but several TCA cycle mutants in *B. japonicum* still formed an efficient symbiosis (199–202). This may be explained by flux through anaplerotic reactions or the existence of uncharacterized isozymes (203). For example, an aconitase mutant of *B. japonicum* retained almost 30% of the aconitase activity (201). Additional mutant studies with careful evaluation of both enzyme activities and redirection of metabolic fluxes are therefore required to elucidate the essentiality of flux through different parts of the TCA cycle.

Dicarboxylate catabolism in the TCA cycle generates large amounts of reduced electron carriers (NADH and reduced flavin adenine dinucleotide [$FADH_2$]) required for ATP synthesis and nitrogen fixation (116, 178). Due to the low oxygen levels in nodules, flux through the TCA cycle may become disadvantageous because the oxidation of electron carriers in the respiratory chain is restricted, and the accumulation of reducing equivalents inhibits pyruvate dehydrogenase and $\alpha$-ketoglutarate dehydrogenase (204–206). Bacteroids could therefore channel carbon into polymers such as polyhydroxybutyrate (PHB), glycogen, and lipids (106, 207–209). Accumulation of these polymers is observed in bacteroids, although the nature of the compounds differs between symbioses (106, 210). In *B. japonicum*, the regulator of PHB synthesis, PhaR, has been shown to repress various gene targets, notably $FixK_2$, a transcriptional regulator for adaptation to microoxic conditions (211, 212). This suggests a certain degree of coordination between PHB metabolism and symbiotic adaptation, at least in some rhizobial strains.

A fundamental requirement for rhizobium-legume symbioses is the secretion of fixed nitrogen to the plant instead of its assimilation by the bacteroid. Ammonia is the main secretion product, but the secretion of alanine and aspartate has also been reported (213–216). The glutamine synthetase-glutamate oxoglutarate aminotransferase (GS-GOGAT) system, which assimilates ammonia into glutamate in free-living rhizobia, appears to be largely downregulated in bacteroids (116, 217). Rhizobia usually possess two glutamine synthetases (GSI/GlnA and GSII/GlnII), with some strains encoding a third enzyme (GSIII/GlnT). In general, only GSI activity is present at low levels in bacteroids, and the deletion of any of the GS genes does not impair the symbiosis (217, 218). Mutant studies have further shown GOGAT to be dispensable in *S. meliloti* (219, 220), *R. etli* (221), and *R. leguminosarum* (222). Overall, these results indicate no need for ammonia assimilation by the bacteroid, which may be due to amino acid provision by the plant or reduced protein biosynthesis as a result of growth arrest.

A recent study showed that the nitrogen-related phosphotransferase system in *R. leguminosarum* controls central carbon metabolism in response to carbon and nitrogen availability, and mutation of this system caused an ineffective symbiosis (223). This highlights the complex interconnection between carbon and nitrogen metabolism in bacteroids, which must be finely tuned to enable a functional symbiosis. The complex regulatory systems underlying rhizobial nitrogen metabolism have been comprehensively discussed elsewhere (179, 217, 224).

## MODELING BACTEROID METABOLISM

Bacteroid metabolism is difficult to investigate experimentally due to the challenge of measuring nutrient exchange with the plant host, the multiple lifestyle adaptations required during symbiosis formation (63), and the fragile nature of isolated bacteroids (178). In addition to various omics techniques, constraint-based metabolic models have therefore become a popular tool for investigating rhizobium-legume symbioses (225). Metabolic models are knowledgebases of the reactions catalyzed by enzymes encoded in an organism's genome. By imposing an objective function and assuming a steady state of all metabolic reactions, flux distributions are calculated using flux balance analysis (FBA) (226, 227). Models with different scopes have been reconstructed



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

for various rhizobial species (see Table S4 in the supplemental material), and studies have progressed from describing bacteroid metabolism (228–231) to integrating bacterial and plant models as well as models of nodule zone-specific metabolism (232, 233). Rhizobial genomes are large and may include secondary replicons encoding niche-specific functions (60, 234). Metabolic models specific to the nitrogen-fixing state must therefore be constrained using transcriptome and proteome data sets (228) and/or appropriate objective functions (231, 234). Since direct measurement of reaction rates and exchange fluxes with the plant is infeasible, FBA studies have mostly used objective functions composed of generally accepted features of bacteroid metabolism, such as nitrogen fixation, amino acid secretion, and the production of carbon polymers (229, 230, 235).

All models published to date are consistent with experimental evidence for the use of the TCA cycle and oxidative phosphorylation as the main pathways fueling bacteroid metabolism. Similar to the disparities in experimental results, the operation of both a full and a partial TCA cycle during nitrogen fixation has been predicted. This may be due to strain-specific differences as well as the choice of constraints, in particular carbon and oxygen uptake rates. Synthesis of carbon polymers as well as amino acid exchange with the plant have also been investigated *in silico*. An increase in symbiotic efficiency was predicted for the deletion of glycogen or PHB synthesis in *R. etli* (229), which agrees with the enhanced symbiotic properties of a glycogen synthase mutant of *R. tropici* (236) but not with the phenotypes of mutants for PHB and glycogen synthases in *R. leguminosarum* and *S. meliloti* (207, 209). Most, but not all, models assume extensive amino acid cycling between the plant and bacteroid. This is based on early experimental studies suggesting glutamate supply by the plant (180), which, however, was later shown to be nonessential for bacteroids (181). With regard to amino acids secreted by the bacteroid, alanine has been predicted to be the main nitrogenous compound produced at low oxygen levels since it serves as a sink for both carbon and electrons (232).

A recent study of an integrated model for plant and bacteroid metabolism further addressed the question of why $C_4$-dicarboxylates are supplied to bacteroids instead of sucrose, which is abundant in nodules. Rhizobia differentiating into bacteroids were predicted to catabolize sugars (233), which is supported by studies of mutants in dicarboxylate transport, which are able to nodulate and differentiate into bacteroids but fail to fix nitrogen (187, 188). Modeling results reported previously (233) indicated that sucrose catabolism in fully differentiated bacteroids would achieve higher yields of fixed nitrogen and that proton supply by the plant is required to support the catabolism of $C_4$-dicarboxylates. The provision of $C_4$-dicarboxylates was further predicted to be driven by oxygen limitation of the plant mitochondria (233). Experimental evidence has been found for oxygen limitation of both plant (106) and bacteroid (237, 238) fractions, highlighting this as an outstanding question for future studies.

It is important to note that carbon polymer synthesis and amino acid secretion are commonly included in the bacteroid objective function in addition to ammonia export. Since flux distributions calculated by FBA must produce all compounds that are part of the objective function, carbon and nitrogen allocations in metabolic models are constrained by forcing the stoichiometric production of these compounds. This was circumvented by an integrated model of plant and bacteroid metabolism and optimization for plant biomass production (233). In future studies, the application of comprehensive and unbiased modeling approaches, such as elementary flux mode enumeration (239), may be insightful to improve and expand our current understanding of bacteroid metabolism.

Apart from the direct characterization of bacteroid metabolism, the underlying gene-protein-reaction relationships in metabolic models are useful for guiding targeted investigation of gene essentiality. Seeing as symbiosis formation is a multistep process requiring various metabolic capabilities at different stages (63), it is difficult to distinguish between genes essential for infection and those for essential nitrogen fixation itself. Discrepancies



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

between experimental and *in silico* gene essentiality can thus unravel metabolic requirements at different stages of symbiosis formation.

## PLANT CONTROL AND METABOLIC SANCTIONING

Both computational and experimental studies have demonstrated how bacteroid metabolism can be modulated in response to nutrient supply by the plant, in particular carbon and oxygen (240, 241). Plant control over bacteroid metabolism is essential due to the high fitness cost of nodule formation and the maintenance of bacteroids. The plant host must therefore be capable of monitoring symbiotic performance and responding accordingly, which becomes even more important when considering that the efficiency of the symbiont cannot be evaluated prior to nodulation (85, 242). Mathematical models demonstrated that a substantial investment in nitrogen fixation will be advantageous for rhizobia only if the plant adjusts its resource allocation according to symbiotic performance (243). To maximize their fitness, reproductive bacteroids can divert some of the plant-supplied carbon into polymers and catabolize those after dedifferentiation into the free-living state. For nonreproductive bacteroids, the undifferentiated population in the nodule benefits from the carbon supply and will be released into the soil after nodule senescence (244). Thus, the rhizobium-legume symbioses seem prone to selecting for cheaters that fix no or little nitrogen. In fact, field isolates vary greatly in their nitrogen fixation efficiencies (89). Consequently, legumes must have the means to identify and punish inefficient strains to maximize the nitrogen return for their carbon investment. It is well established that $Fix^-$ strains do not persist within mature nodules (245), and when nodules are coinhabited with a fixing strain, the $Fix^-$ strain is sanctioned rapidly in a cell-autonomous way (246). Sanctioning results in a fitness increase of good symbionts since higher numbers will be released from nodules. Under experimental conditions, efficient symbionts will outcompete $Fix^-$ strains in a few plant generations (247, 248) since plants readily select the most efficient strain (249), even when the initial cell numbers are lower than those of the less efficient strain. Experimental data on carbon and oxygen supply to cheating strains (85, 240, 241) show that plants sense and integrate the nitrogen output from bacteroids and readjust carbon investment (250) or apply oxygen sanctioning in response to poorly performing strains.

## FINAL CONCLUSIONS AND FUTURE PERSPECTIVES

The bacteroid stage in the rhizobial life cycle features a unique combination of traits. It is characterized by a nongrowing state in an oxygen-depleted environment. Despite the downregulation of many biosynthetic pathways, bacteroids maintain a highly active and specialized metabolism driven by the catabolism of $C_4$-dicarboxylates. Rhizobia are equipped with specialized functions to cope with this environment, and global adjustments of cellular and metabolic processes must take place to ensure a successful transition to nitrogen-fixing bacteroids. Failure to do so is punished by the host plant and results in the reduced fitness of a rhizobial strain. Host plants impose control over their symbionts at multiple levels. Nutrient supply, mostly in the form of malate and succinate, can be reduced; the oxygen supply can be limited; or, in some cases, plants can target bacteroids directly with antimicrobial peptides to force them into terminal differentiation.

Despite major advances in our knowledge, the complex interactions in rhizobium-legume symbioses retain many unanswered questions. Modulation of the plant immune response by both the host and the symbiont to accommodate thousands of intracellular prokaryotes per cell, what it shares with the better-understood immune response during infection, and its interplay with host sanctioning and pathogens remain exciting research areas despite substantial published work (245, 251–253). Similarly, our understanding of nutrient supply to bacteroids beyond malate and branched-chain amino acids is fragmented, obscuring the stoichiometry of metabolic fluxes that drive nitrogen fixation. The rhizobium-legume symbioses are prime examples of biological systems whose complexity necessitates the combination of experimental and computational approaches. *In silico* modeling is a rapidly advancing field that will greatly



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

contribute to our understanding of the symbioses and aid in advancing toward engineering synthetic symbioses.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.5 MB.

## ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (PostDoc.Mobility fellowship number 183901 to R.L.) and the Biotechnology and Biological Sciences Research Council (BBSRC) (grant number BB/M011224/1). C.C.M.S. is supported by the Clarendon Fund (Oxford University Press) and the Keble College De Breyne scholarship.

## REFERENCES

1. Brundrett MC. 2002. Coevolution of roots and mycorrhizas of land plants. New Phytol 154:275–304. https://doi.org/10.1046/j.1469-8137.2002.00397.x.
2. Radhakrishnan GV, Keller J, Rich MK, Vernie T, Mbadinga Mbadinga DL, Vigneron N, Cottret L, Clemente HS, Libourel C, Cheema J, Linde A-M, Eklund DM, Cheng S, Wong GKS, Lagercrantz U, Li F-W, Oldroyd GED, Delaux P-M. 2020. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. Nat Plants 6:280–289. https://doi.org/10.1038/s41477-020-0613-7.
3. Gherbi H, Markmann K, Svistoonoff S, Estevan J, Autran D, Giczey G, Auguy F, Peret B, Laplaze L, Franche C, Parniske M, Bogusz D. 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankia* bacteria. Proc Natl Acad Sci U S A 105:4928–4932. https://doi.org/10.1073/pnas.0710618105.
4. Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci U S A 92:2647–2651. https://doi.org/10.1073/pnas.92.7.2647.
5. Delaux PM, Radhakrishnan G, Oldroyd G. 2015. Tracing the evolutionary path to nitrogen-fixing crops. Curr Opin Plant Biol 26:95–99. https://doi.org/10.1016/j.pbi.2015.06.003.
6. Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, Roswanjaya YP, Kohlen W, Pujic P, Battenberg K, Alloisio N, Liang Y, Hilhorst H, Salgado MG, Hocher V, Gherbi H, Svistoonoff S, Doyle JJ, He S, Xu Y, Xu S, Qu J, Gao Q, Fang X, Fu Y, Normand P, Berry AM, Wall LG, Ane J-M, Pawlowski K, Xu X, Yang H, Spannagl M, Mayer KFX, Wong GK-S, Parniske M, Delaux P-M, Cheng S. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science 361:eaat1743. https://doi.org/10.1126/science.aat1743.
7. Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311:1–18. https://doi.org/10.1007/s11104-008-9668-3.
8. Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM. 2010. Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol 52:61–76. https://doi.org/10.1111/j.1744-7909.2010.00899.x.
9. Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, Vitousek P, Leach A, Bouwman AF, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M, Voss M. 2013. The global nitrogen cycle in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 368:20130164. https://doi.org/10.1098/rstb.2013.0164.
10. Pesce C, Oshone R, Hurst SG, IV, Kleiner VA, Tisa LS. 2019. Stable transformation of the actinobacteria *Frankia* spp. Appl Environ Microbiol 85:e00957-19. https://doi.org/10.1128/AEM.00957-19.
11. Ward MH. 1887. On the tubular swellings on the root of *Vicia faba*. Philos Trans R Soc Lond B 178:539–562.
12. Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many *Rhizobium* recipes? Trends Microbiol 17:458–466. https://doi.org/10.1016/j.tim.2009.07.004.
13. Hadri A-E, Spaink HP, Bisseling T, Brewin NJ. 1998. Diversity of root nodulation and rhizobial infection processes, p 347–360. *In* Spaink HP, Kondorosi A, Hooykaas PJJ (ed), The Rhizobiaceae. Springer, Dordrecht, The Netherlands. https://doi.org/10.1007/978-94-011-5060-6_18.
14. Sprent JI, Ardley J, James EK. 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. New Phytol 215:40–56. https://doi.org/10.1111/nph.14474.
15. Sprent JI. 2009. Legume nodulation: a global perspective, p 1–33. Wiley-Blackwell, West Sussex, United Kingdom. https://doi.org/10.1002/9781444316384.ch1.
16. Alazard D, Duhoux E. 1990. Development of stem nodules in a tropical forage legume, *Aeschynomene afraspera*. J Exp Bot 41:1199–1206. https://doi.org/10.1093/jxb/41.9.1199.
17. Chaintreuil C, Arrighi JF, Giraud E, Miche L, Moulin L, Dreyfus B, Munive-Hernandez JA, Villegas-Hernandez MD, Bena G. 2013. Evolution of symbiosis in the legume genus *Aeschynomene*. New Phytol 200:1247–1259. https://doi.org/10.1111/nph.12424.
18. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermeglio A, Medigue C, Sadowsky M. 2007. Legumes symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. Science 316:1307–1312. https://doi.org/10.1126/science.1139548.
19. Okazaki S, Kaneko T, Sato S, Saeki K. 2013. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proc Natl Acad Sci U S A 110:17131–17136. https://doi.org/10.1073/pnas.1302360110.
20. Turgeon BG, Bauer WD. 1985. Ultrastructure of infection-thread development during the infection of soybean by *Rhizobium japonicum*. Planta 163:328–349. https://doi.org/10.1007/BF00395142.
21. Calvert HE, Pence MK, Pierce M, Malik NSA, Bauer WD. 1984. Anatomical analysis of the development and distribution of *Rhizobium* infections in soybean roots. Can J Bot 62:2375–2384. https://doi.org/10.1139/b84-324.
22. Heron DS, Pueppke SG. 1984. Mode of infection, nodulation specificity, and indigenous plasmids of 11 fast-growing *Rhizobium japonicum* strains. J Bacteriol 160:1061–1066. https://doi.org/10.1128/JB.160.3.1061-1066.1984.
23. Pueppke SG. 1983. *Rhizobium* infection threads in root hairs of *Glycine max* (L.) Merr., *Glycine soja* Sieb. & Zucc, and *Vigna unguiculata* (L.) Walp. Can J Microbiol 29:69–76. https://doi.org/10.1139/m83-011.
24. Szczyglowski K, Shaw RS, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, de Bruijn FJ. 1998. Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. Mol Plant Microbe Interact 11:684–697. https://doi.org/10.1094/MPMI.1998.11.7.684.
25. Rae AL, Bonfante-Fasolo P, Brewin NJ. 1992. Structure and growth of infection threads in the legume symbiosis with *Rhizobium leguminosarum*. Plant J 2:385–395. https://doi.org/10.1111/j.1365-313X.1992.00385.x.
26. Gage DJ, Bobo T, Long SR. 1996. Use of green fluorescent protein to visualize the early events of symbiosis between *Rhizobium meliloti* and alfalfa (*Medicago sativa*). J Bacteriol 178:7159–7166. https://doi.org/10.1128/jb.178.24.7159-7166.1996.
27. Cooper JE. 2004. Multiple responses of rhizobia to flavonoids during legume root infection. Adv Bot Res 41:1–62. https://doi.org/10.1016/S0065-2296(04)41001-5.
28. D'Haeze W, Holsters M. 2002. Nod factor structures, responses, and perception during initiation of nodule development. Glycobiology 12:79R–105R. https://doi.org/10.1093/glycob/12.6.79r.
29. Spaink HP. 2000. Root nodulation and infection factors produced by rhizobial bacteria. Annu Rev Microbiol 54:257–288. https://doi.org/10.1146/annurev.micro.54.1.257.



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

30. Gage DJ. 2002. Analysis of infection thread development using Gfp- and DsRed-expressing *Sinorhizobium meliloti*. J Bacteriol 184:7042–7046. https://doi.org/10.1128/jb.184.24.7042-7046.2002.
31. Ledermann R, Bartsch I, Remus-Emsermann MN, Vorholt JA, Fischer HM. 2015. Stable fluorescent and enzymatic tagging of *Bradyrhizobium diazoefficiens* to analyze host-plant infection and colonization. Mol Plant Microbe Interact 28:959–967. https://doi.org/10.1094/MPMI-03-15-0054-TA.
32. Fournier J, Timmers AC, Sieberer BJ, Jauneau A, Chabaud M, Barker DG. 2008. Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. Plant Physiol 148:1985–1995. https://doi.org/10.1104/pp.108.125674.
33. Brewin NJ. 2004. Plant cell wall remodelling in the *Rhizobium*-legume symbiosis. Crit Rev Plant Sci 23:293–316. https://doi.org/10.1080/07352680490480734.
34. Wilson KJ, Anjaiah V, Nambiar PT, Ausubel FM. 1987. Isolation and characterization of symbiotic mutants of *Bradyrhizobium* sp. (*Arachis*) strain NC92: mutants with host-specific defects in nodulation and nitrogen fixation. J Bacteriol 169:2177–2186. https://doi.org/10.1128/jb.169.5.2177-2186.1987.
35. Pueppke SG, Broughton WJ. 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Mol Plant Microbe Interact 12:293–318. https://doi.org/10.1094/MPMI.1999.12.4.293.
36. Newcomb W, Sippell D, Peterson RL. 1979. The early morphogenesis of *Glycine max* and *Pisum sativum* root nodules. Can J Bot 57:2603–2616. https://doi.org/10.1139/b79-309.
37. Newcomb W. 1981. Nodule morphogenesis and differentiation. Int Rev Cytol Suppl 13:247–298.
38. Brewin NJ. 1991. Development of the legume root nodule. Annu Rev Cell Biol 7:191–226. https://doi.org/10.1146/annurev.cb.07.110191.001203.
39. Hirsch AM. 1992. Developmental biology of legume nodulation. New Phytol 122:211–237. https://doi.org/10.1111/j.1469-8137.1992.tb04227.x.
40. Tkacz A, Cheema J, Chandra G, Grant A, Poole PS. 2015. Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. ISME J 9:2349–2359. https://doi.org/10.1038/ismej.2015.41.
41. Zgadzaj R, Garrido-Oter R, Jensen DB, Koprivova A, Schulze-Lefert P, Radutoiu S. 2016. Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. Proc Natl Acad Sci U S A 113:E7996–E8005. https://doi.org/10.1073/pnas.1616564113.
42. Frey SD, Blum LK. 1994. Effect of pH on competition for nodule occupancy by type I and type II strains of *Rhizobium leguminosarum* bv. *phaseoli*. Plant Soil 163:157–164. https://doi.org/10.1007/BF00007964.
43. Angus AA, Lee A, Lum MR, Shehayeb M, Hessabi R, Fujishige NA, Yerrapragada S, Kano S, Song N, Yang P, Estrada de los Santos P, de Faria SM, Dakora FD, Weinstock G, Hirsch AM. 2013. Nodulation and effective nitrogen fixation of *Macroptilium atropurpureum* (siratro) by *Burkholderia tuberum*, a nodulating and plant growth promoting $\beta$-proteobacterium, are influenced by environmental factors. Plant Soil 369:543–562. https://doi.org/10.1007/s11104-013-1590-7.
44. Rathi S, Tak N, Bissa G, Chouhan B, Ojha A, Adhikari D, Barik SK, Satyawada RR, Sprent JI, James EK, Gehlot HS. 2018. Selection of *Bradyrhizobium* or *Ensifer* symbionts by the native Indian caesalpinioid legume *Chamaecrista pumila* depends on soil pH and other edaphic and climatic factors. FEMS Microbiol Ecol 94:fiy180. https://doi.org/10.1093/femsec/fiy180.
45. Elliott GN, Chou JH, Chen WM, Bloemberg GV, Bontemps C, Martinez-Romero E, Velazquez E, Young JP, Sprent JI, James EK. 2009. *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. Environ Microbiol 11:762–778. https://doi.org/10.1111/j.1462-2920.2008.01799.x.
46. Lardi M, de Campos SB, Purtschert G, Eberl L, Pessi G. 2017. Competition experiments for legume infection identify *Burkholderia phymatum* as a highly competitive $\beta$-rhizobium. Front Microbiol 8:1527. https://doi.org/10.3389/fmicb.2017.01527.
47. Boivin S, Ait Lahmidi N, Sherlock D, Bonhomme M, Dijon D, Heulin-Gotty K, Le-Quere A, Pervent M, Tauzin M, Carlsson G, Jensen E, Journet EP, Lopez-Bellido R, Seidenglanz M, Marinkovic J, Colella S, Brunel B, Young P, Lepetit M. 2020. Host-specific competitiveness to form nodules in Rhizobium leguminosarum symbiovar viciae. New Phytol 226:555–568. https://doi.org/10.1111/nph.16392.
48. Salas ME, Lozano MJ, Lopez JL, Draghi WO, Serrania J, Torres Tejerizo GA, Albicoro FJ, Nilsson JF, Pistorio M, Del Papa MF, Parisi G, Becker A, Lagares A. 2017. Specificity traits consistent with legume-rhizobia coevolution displayed by *Ensifer meliloti* rhizosphere colonization. Environ Microbiol 19:3423–3438. https://doi.org/10.1111/1462-2920.13820.
49. Russell AB, LeRoux M, Hathazi K, Agnello DM, Ishikawa T, Wiggins PA, Wai SN, Mougous JD. 2013. Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. Nature 496:508–512. https://doi.org/10.1038/nature12074.
50. de Campos SB, Lardi M, Gandolfi A, Eberl L, Pessi G. 2017. Mutations in two *Paraburkholderia phymatum* type VI secretion systems cause reduced fitness in interbacterial competition. Front Microbiol 8:2473. https://doi.org/10.3389/fmicb.2017.02473.
51. Geddes BA, Gonzalez JE, Oresnik IJ. 2014. Exopolysaccharide production in response to medium acidification is correlated with an increase in competition for nodule occupancy. Mol Plant Microbe Interact 27:1307–1317. https://doi.org/10.1094/MPMI-06-14-0168-R.
52. Fry J, Wood M, Poole PS. 2001. Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. Mol Plant Microbe Interact 14:1016–1025. https://doi.org/10.1094/MPMI.2001.14.8.1016.
53. Jiang G, Krishnan AH, Kim YW, Wacek TJ, Krishnan HB. 2001. A functional *myo*-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] Merr.). J Bacteriol 183:2595–2604. https://doi.org/10.1128/JB.183.8.2595-2604.2001.
54. Kohler PR, Zheng JY, Schoffers E, Rossbach S. 2010. Inositol catabolism, a key pathway in *Sinorhizobium meliloti* for competitive host nodulation. Appl Environ Microbiol 76:7972–7980. https://doi.org/10.1128/AEM.01972-10.
55. Ding H, Yip CB, Geddes BA, Oresnik IJ, Hynes MF. 2012. Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. Microbiology (Reading) 158:1369–1378. https://doi.org/10.1099/mic.0.057281-0.
56. Garcia-Fraile P, Seaman JC, Karunakaran R, Edwards A, Poole PS, Downie JA. 2015. Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar *viciae* in the pea rhizosphere. Plant Soil 390:251–264. https://doi.org/10.1007/s11104-015-2389-5.
57. Oresnik IJ, Pacarynuk LA, O'Brien SAP, Yost CK, Hynes MF. 1998. Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. *trifolii*: evidence for a plant-inducible rhamnose locus involved in competition for nodulation. Mol Plant Microbe Interact 11:1175–1185. https://doi.org/10.1094/MPMI.1998.11.12.1175.
58. Vanderlinde EM, Hynes MF, Yost CK. 2014. Homoserine catabolism by *Rhizobium leguminosarum* bv. *viciae* 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. Environ Microbiol 16:205–217. https://doi.org/10.1111/1462-2920.12196.
59. Yost CK, Rath AM, Noel TC, Hynes MF. 2006. Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. Microbiology (Reading) 152:2061–2074. https://doi.org/10.1099/mic.0.28938-0.
60. Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS. 2011. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. Genome Biol 12:R106. https://doi.org/10.1186/gb-2011-12-10-r106.
61. Ledermann R, Bartsch I, Müller B, Wülser J, Fischer HM. 2018. A functional general stress response of *Bradyrhizobium diazoefficiens* is required for early stages of host plant infection. Mol Plant Microbe Interact 31:537–547. https://doi.org/10.1094/MPMI-11-17-0284-R.
62. Moawad HA, Ellis WR, Schmidt EL. 1984. Rhizosphere response as a factor in competition among three serogroups of indigenous *Rhizobium japonicum* for nodulation of field-grown soybeans. Appl Environ Microbiol 47:607–612. https://doi.org/10.1128/AEM.47.4.607-612.1984.
63. Wheatley RM, Ford BL, Li L, Aroney STN, Knights HE, Ledermann R, East AK, Ramachandran VK, Poole PS. 2020. Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. Proc Natl Acad Sci U S A 117:23823–23834. https://doi.org/10.1073/pnas.2009094117.
64. Noel KD, Vandenbosch KA, Kulpaca B. 1986. Mutations in *Rhizobium phaseoli* that lead to arrested development of infection threads. J Bacteriol 168:1392–1401. https://doi.org/10.1128/jb.168.3.1392-1401.1986.
65. Finan TM, Hirsch AM, Leigh JA, Johansen E, Kuldau GA, Deegan S, Walker GC, Signer ER. 1985. Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. Cell 40:869–877. https://doi.org/10.1016/0092-8674(85)90346-0.
66. Cheng HP, Walker GC. 1998. Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. J Bacteriol 180:5183–5191. https://doi.org/10.1128/JB.180.19.5183-5191.1998.
67. Kelly SJ, Muszyński A, Kawaharada Y, Hubber AM, Sullivan JT, Sandal N, Carlson RW, Stougaard J, Ronson CW. 2013. Conditional requirement for




Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

exopolysaccharide in the *Mesorhizobium-Lotus* symbiosis. Mol Plant Microbe Interact 26:319–329. https://doi.org/10.1094/MPMI-09-12-0227-R.
68. Scheidle H, Gross A, Niehaus K. 2005. The lipid A substructure of the *Sinorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants. New Phytol 165:559–565. https://doi.org/10.1111/j.1469-8137.2004.01214.x.
69. Stacey G, So J-S, Roth EL, Lakshmi B, Carlson RW. 1991. A lipopolysaccharide mutant of *Bradyrhizobium japonicum* that uncouples plant from bacterial differentiation. Mol Plant Microbe Interact 4:332–340. https://doi.org/10.1094/mpmi-4-332.
70. Jones KM, Sharopova N, Lohar DP, Zhang JQ, VandenBosch KA, Walker GC. 2008. Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. Proc Natl Acad Sci U S A 105:704–709. https://doi.org/10.1073/pnas.0709338105.
71. Wells DH, Long SR. 2002. The *Sinorhizobium meliloti* stringent response affects multiple aspects of symbiosis. Mol Microbiol 43:1115–1127. https://doi.org/10.1046/j.1365-2958.2002.02826.x.
72. Krol E, Becker A. 2011. ppGpp in Sinorhizobium meliloti: biosynthesis in response to sudden nutritional downshifts and modulation of the transcriptome. Mol Microbiol 81:1233–1254. https://doi.org/10.1111/j.1365-2958.2011.07752.x.
73. Wippel K, Long SR. 2016. Contributions of Sinorhizobium meliloti transcriptional regulator DksA to bacterial growth and efficient symbiosis with Medicago sativa. J Bacteriol 198:1374–1383. https://doi.org/10.1128/JB.00013-16.
74. Gaworzewska ET, Carlile MJ. 1982. Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. J Gen Microbiol 128:1179–1188. https://doi.org/10.1099/00221287-128-6-1179.
75. Haichar FEZ, Santaella C, Heulin T, Achouak W. 2014. Root exudates mediated interactions belowground. Soil Biol Biochem 77:69–80. https://doi.org/10.1016/j.soilbio.2014.06.017.
76. Marchetti M, Capela D, Glew M, Cruveiller S, Chane-Woon-Ming B, Gris C, Timmers T, Poinsot V, Gilbert LB, Heeb P, Medigue C, Batut J, Masson-Boivin C. 2010. Experimental evolution of a plant pathogen into a legume symbiont. PLoS Biol 8:e1000280. https://doi.org/10.1371/journal.pbio.1000280.
77. Marchetti M, Jauneau A, Capela D, Remigi P, Gris C, Batut J, Masson-Boivin C. 2014. Shaping bacterial symbiosis with legumes by experimental evolution. Mol Plant Microbe Interact 27:956–964. https://doi.org/10.1094/MPMI-03-14-0083-R.
78. Capela D, Marchetti M, Clerissi C, Perrier A, Guetta D, Gris C, Valls M, Jauneau A, Cruveiller S, Rocha EPC, Masson-Boivin C. 2017. Recruitment of a lineage-specific virulence regulatory pathway promotes intracellular infection by a plant pathogen experimentally evolved into a legume symbiont. Mol Biol Evol 34:2503–2521. https://doi.org/10.1093/molbev/msx165.
79. Tang M, Bouchez O, Cruveiller S, Masson-Boivin C, Capela D. 2020. Modulation of quorum sensing as an adaptation to nodule cell infection during experimental evolution of legume symbionts. mBio 11:e03129-19. https://doi.org/10.1128/mBio.03129-19.
80. Andrews M, De Meyer S, James EK, Stępkowski T, Hodge S, Simon MF, Young JPW. 2018. Horizontal transfer of symbiosis genes within and between rhizobial genera: occurrence and importance. Genes (Basel) 9:321. https://doi.org/10.3390/genes9070321.
81. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci U S A 92:8985–8989. https://doi.org/10.1073/pnas.92.19.8985.
82. Sullivan JT, Ronson CW. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci U S A 95:5145–5149. https://doi.org/10.1073/pnas.95.9.5145.
83. Haskett TL, Ramsay JP, Bekuma AA, Sullivan JT, O'Hara GW, Terpolilli JJ. 2017. Evolutionary persistence of tripartite integrative and conjugative elements. Plasmid 92:30–36. https://doi.org/10.1016/j.plasmid.2017.06.001.
84. López-Guerrero MG, Ormeño-Orrillo E, Acosta JL, Mendoza-Vargas A, Rogel MA, Ramírez MA, Rosenblueth M, Martínez-Romero J, Martínez-Romero E. 2012. Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. Plasmid 68:149–158. https://doi.org/10.1016/j.plasmid.2012.07.002.
85. Westhoek A, Field E, Rehling F, Mulley G, Webb I, Poole PS, Turnbull LA. 2017. Policing the legume-*Rhizobium* symbiosis: a critical test of partner choice. Sci Rep 7:1419. https://doi.org/10.1038/s41598-017-01634-2.
86. Checcucci A, DiCenzo GC, Bazzicalupo M, Mengoni A. 2017. Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. Front Microbiol 8:2207. https://doi.org/10.3389/fmicb.2017.02207.
87. Dwivedi SL, Sahrawat KL, Upadhyaya HD, Mengoni A, Galardini M, Bazzicalupo M, Biondi EG, Hungria M, Kaschuk G, Blair MW, Ortiz R. 2015. Advances in host plant and *Rhizobium* genomics to enhance symbiotic nitrogen fixation in grain legumes. Adv Agron 129:1–116. https://doi.org/10.1016/bs.agron.2014.09.001.
88. Pastor-Bueis R, Sanchez-Canizares C, James EK, Gonzalez-Andres F. 2019. Formulation of a highly effective inoculant for common bean based on an autochthonous elite strain of *Rhizobium leguminosarum* bv. *phaseoli*, and genomic-based insights into its agronomic performance. Front Microbiol 10:2724. https://doi.org/10.3389/fmicb.2019.02724.
89. Mendoza-Suárez MA, Geddes BA, Sanchez-Canizares C, Ramirez-Gonzalez RH, Kirchhelle C, Jorrin B, Poole PS. 2020. Optimizing *Rhizobium*-legume symbioses by simultaneous measurement of rhizobial competitiveness and $N_2$ fixation in nodules. Proc Natl Acad Sci U S A 117:9822–9831. https://doi.org/10.1073/pnas.1921225117.
90. Couzigou JM, Mondy S, Sahl L, Gourion B, Ratet P. 2013. To be or noot to be: evolutionary tinkering for symbiotic organ identity. Plant Signal Behav 8:e24969. https://doi.org/10.4161/psb.24969.
91. Benezech C, Berrabah F, Jardinaud MF, Le Scornet A, Milhes M, Jiang G, George J, Ratet P, Vailleau F, Gourion B. 2020. *Medicago-Sinorhizobium-Ralstonia* co-infection reveals legume nodules as pathogen confined infection sites developing weak defenses. Curr Biol 30:351–358.e4. https://doi.org/10.1016/j.cub.2019.11.066.
92. Vercruysse M, Fauvart M, Beullens S, Braeken K, Cloots L, Engelen K, Marchal K, Michiels J. 2011. A comparative transcriptome analysis of *Rhizobium etli* bacteroids: specific gene expression during symbiotic non-growth. Mol Plant Microbe Interact 24:1553–1561. https://doi.org/10.1094/MPMI-05-11-0140.
93. Peng J, Hao B, Liu L, Wang S, Ma B, Yang Y, Xie F, Li Y. 2014. RNA-Seq and microarrays analyses reveal global differential transcriptomes of *Mesorhizobium huakuii* 7653R between bacteroids and free-living cells. PLoS One 9:e93626. https://doi.org/10.1371/journal.pone.0093626.
94. Tsukada S, Aono T, Akiba N, Lee KB, Liu CT, Toyazaki H, Oyaizu H. 2009. Comparative genome-wide transcriptional profiling of *Azorhizobium caulinodans* ORS571 grown under free-living and symbiotic conditions. Appl Environ Microbiol 75:5037–5046. https://doi.org/10.1128/AEM.00398-09.
95. Barnett MJ, Toman CJ, Fisher RF, Long SR. 2004. A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. Proc Natl Acad Sci U S A 101:16636–16641. https://doi.org/10.1073/pnas.0407269101.
96. Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H. 2007. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. Mol Plant Microbe Interact 20:1353–1363. https://doi.org/10.1094/MPMI-20-11-1353.
97. Ogden AJ, Gargouri M, Park J, Gang DR, Kahn ML. 2017. Integrated analysis of zone-specific protein and metabolite profiles within nitrogen-fixing *Medicago truncatula-Sinorhizobium medicae* nodules. PLoS One 12:e0180894. https://doi.org/10.1371/journal.pone.0180894.
98. Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS. 2009. Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J Bacteriol 191:4002–4014. https://doi.org/10.1128/JB.00165-09.
99. Green RT, East AK, Karunakaran R, Downie JA, Poole PS. 2019. Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. Microb Genom 5:e000254. https://doi.org/10.1099/mgen.0.000254.
100. Lamouche F, Gully D, Chaumeret A, Nouwen N, Verly C, Pierre O, Sciallano C, Fardoux J, Jeudy C, Szücs A, Mondy S, Salon C, Nagy I, Kereszt A, Dessaux Y, Giraud E, Mergaert P, Alunni B. 2019. Transcriptomic dissection of *Bradyrhizobium* sp. strain ORS285 in symbiosis with *Aeschynomene* spp. inducing different bacteroid morphotypes with contrasted symbiotic efficiency. Environ Microbiol 21:3244–3258. https://doi.org/10.1111/1462-2920.14292.
101. Li Y, Tian CF, Chen WF, Wang L, Sui XH, Chen WX. 2013. High-resolution transcriptomic analyses of *Sinorhizobium* sp. NGR234 bacteroids in determinate nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena leucocephala*. PLoS One 8:e70531. https://doi.org/10.1371/journal.pone.0070531.
102. Roux B, Rodde N, Jardinaud M-F, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, Debellé F, Capela D, de Carvalho-

Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

Niebel F, Gouzy J, Bruand C, Gamas P. 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J 77:817–837. https://doi.org/10.1111/tpj.12442.
103. Koch M, Delmotte N, Rehrauer H, Vorholt JA, Pessi G, Hennecke H. 2010. Rhizobial adaptation to hosts, a new facet in the legume root-nodule symbiosis. Mol Plant Microbe Interact 23:784–790. https://doi.org/10.1094/MPMI-23-6-0784.
104. Capela D, Filipe C, Bobik C, Batut J, Bruand C. 2006. *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. Mol Plant Microbe Interact 19:363–372. https://doi.org/10.1094/MPMI-19-0363.
105. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. 2019. Microbial genome analysis: the COG approach. Brief Bioinform 20:1063–1070. https://doi.org/10.1093/bib/bbx117.
106. Terpolilli JJ, Masakapalli SK, Karunakaran R, Webb IU, Green R, Watmough NJ, Kruger NJ, Ratcliffe RG, Poole PS. 2016. Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. J Bacteriol 198:2864–2875. https://doi.org/10.1128/JB.00451-16.
107. Hauser F, Pessi G, Friberg M, Weber C, Rusca N, Lindemann A, Fischer HM, Hennecke H. 2007. Dissection of the *Bradyrhizobium japonicum* NifA+$s^{54}$ regulon, and identification of a ferredoxin gene (*fdxN*) for symbiotic nitrogen fixation. Mol Genet Genomics 278:255–271. https://doi.org/10.1007/s00438-007-0246-9.
108. Glöer J, Thummer R, Ullrich H, Schmitz RA. 2008. Towards understanding the nitrogen signal transduction for *nif* gene expression in *Klebsiella pneumoniae*. FEBS J 275:6281–6294. https://doi.org/10.1111/j.1742-4658.2008.06752.x.
109. Sciotti MA, Chanfon A, Hennecke H, Fischer HM. 2003. Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in *Bradyrhizobium japonicum*. J Bacteriol 185:5639–5642. https://doi.org/10.1128/JB.185.18.5639-5642.2003.
110. Rutten PJ, Steel H, Hood GA, McMurtry L, Geddes B, Papachristodoulou A, Poole PS. 2020. A multi-sensor system provides spatiotemporal oxygen regulation of gene expression in a *Rhizobium*-legume symbiosis. bioRxiv https://doi.org/10.1101/2020.09.09.289140:2020.09.09.289140.
111. Fischer HM. 1994. Genetic regulation of nitrogen fixation in rhizobia. Microbiol Rev 58:352–386. https://doi.org/10.1128/MMBR.58.3.352-386.1994.
112. Batut J, Boistard P. 1994. Oxygen control in *Rhizobium*. Antonie Van Leeuwenhoek 66:129–150. https://doi.org/10.1007/BF00871636.
113. Terpolilli JJ, Hood GA, Poole PS. 2012. What determines the efficiency of $N_2$-fixing *Rhizobium*-legume symbioses? Adv Microb Physiol 60:325–389. https://doi.org/10.1016/B978-0-12-398264-3.00005-X.
114. Rutten PJ, Poole PS. 2019. Oxygen regulatory mechanisms of nitrogen fixation in rhizobia. Adv Microb Physiol 75:325–389. https://doi.org/10.1016/bs.ampbs.2019.08.001.
115. Sullivan JT, Brown SD, Ronson CW. 2013. The NifA-RpoN regulon of *Mesorhizobium loti* strain R7A and its symbiotic activation by a novel LacI/GalR-family regulator. PLoS One 8:e53762. https://doi.org/10.1371/journal.pone.0053762.
116. Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol 64:781–805. https://doi.org/10.1146/annurev-arplant-050312-120235.
117. Brigido C, Robledo M, Menendez E, Mateos PF, Oliveira S. 2012. A ClpB chaperone knockout mutant of *Mesorhizobium ciceri* shows a delay in the root nodulation of chickpea plants. Mol Plant Microbe Interact 25:1594–1604. https://doi.org/10.1094/MPMI-05-12-0140-R.
118. Fischer HM, Schneider K, Babst M, Hennecke H. 1999. GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum*. Arch Microbiol 171:279–289. https://doi.org/10.1007/s002030050711.
119. Ogawa J, Long SR. 1995. The *Rhizobium meliloti groELc* locus is required for regulation of early *nod* genes by the transcription activator NodD. Genes Dev 9:714–729. https://doi.org/10.1101/gad.9.6.714.
120. Farkas A, Maróti G, Durgő H, Györgypál Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P, Kondorosi É. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. Proc Natl Acad Sci U S A 111:5183–5188. https://doi.org/10.1073/pnas.1404169111.
121. Cheng G, Karunakaran R, East AK, Munoz-Azcarate O, Poole PS. 2017. Glutathione affects the transport activity of *Rhizobium leguminosarum* 3841 and is essential for efficient nodulation. FEMS Microbiol Lett 364:fnx045. https://doi.org/10.1093/femsle/fnx045.
122. Yang L, El Msehli S, Benyamina S, Lambert A, Hopkins J, Cazareth J, Pierre O, Herouart D, Achi-Smiti S, Boncompagni E, Frendo P. 2020. Glutathione deficiency in *Sinorhizobium meliloti* does not impair bacteroid differentiation but induces early senescence in the interaction with *Medicago truncatula*. Front Plant Sci 11:137. https://doi.org/10.3389/fpls.2020.00137.
123. Dalton DA, Post CJ, Langeberg L. 1991. Effects of ambient oxygen and of fixed nitrogen on concentrations of glutathione, ascorbate, and associated enzymes in soybean root nodules. Plant Physiol 96:812–818. https://doi.org/10.1104/pp.96.3.812.
124. Imlay JA. 2006. Iron-sulphur clusters and the problem with oxygen. Mol Microbiol 59:1073–1082. https://doi.org/10.1111/j.1365-2958.2006.05028.x.
125. Appleby CA. 1974. Leghemoglobin, p 521–554. *In* Quispel A (ed), The biology of nitrogen fixation. North-Holland Publishing Company, Amsterdam, The Netherlands.
126. Unden G, Bongaerts J. 1997. Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. Biochim Biophys Acta 1320:217–234. https://doi.org/10.1016/s0005-2728(97)00034-0.
127. Kitts CL, Ludwig RA. 1994. *Azorhizobium caulinodans* respires with at least four terminal oxidases. J Bacteriol 176:886–895. https://doi.org/10.1128/jb.176.3.886-895.1994.
128. Delgado MJ, Bedmar EJ, Downie JA. 1998. Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. Adv Microb Physiol 40:191–231. https://doi.org/10.1016/s0065-2911(08)60132-0.
129. Youard ZA, Abicht HK, Mohorko E, Serventi F, Rigozzi E, Ledermann R, Fischer H-M, Glockshuber R, Hennecke H. 2015. A plethora of terminal oxidases and their biogenesis factors in *Bradyrhizobium japonicum*, p 293–306. *In* de Bruijn FJ (ed), Biological nitrogen fixation, vol 1. John Wiley & Sons, Inc, Hoboken, NJ. https://doi.org/10.1002/9781119053095.ch29.
130. Yurgel SN, Berrocal J, Wilson C, Kahn ML. 2007. Pleiotropic effects of mutations that alter the *Sinorhizobium meliloti* cytochrome *c* respiratory system. Microbiology (Reading) 153:399–410. https://doi.org/10.1099/mic.0.2006/002634-0.
131. Preisig O, Zufferey R, Thony-Meyer L, Appleby CA, Hennecke H. 1996. A high-affinity $cbb_3$-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. J Bacteriol 178:1532–1538. https://doi.org/10.1128/jb.178.6.1532-1538.1996.
132. Preisig O, Anthamatten D, Hennecke H. 1993. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. Proc Natl Acad Sci U S A 90:3309–3313. https://doi.org/10.1073/pnas.90.8.3309.
133. Zufferey R, Preisig O, Hennecke H, Thöny-Meyer L. 1996. Assembly and function of the cytochrome $cbb_3$ oxidase subunits in *Bradyrhizobium japonicum*. J Biol Chem 271:9114–9119. https://doi.org/10.1074/jbc.271.15.9114.
134. Girard L, Brom S, Davalos A, Lopez O, Soberon M, Romero D. 2000. Differential regulation of *fixN*-reiterated genes in *Rhizobium etli* by a novel *fixL-fixK* cascade. Mol Plant Microbe Interact 13:1283–1292. https://doi.org/10.1094/MPMI.2000.13.12.1283.
135. Kahn ML, Batut J, Daveran M-L, Fourmet J. 1993. Structure and regulation of the *fixNOQP* operon from *Rhizobium meliloti*, p 474. *In* Palacios R, Mora J, Newton WE (ed), New horizons in nitrogen fixation. Kluwer Academic Publishers, Dordrecht, The Netherlands.
136. Schlüter A, Patschkowski T, Quandt J, Selinger LB, Weidner S, Kramer M, Zhou L, Hynes MF, Priefer UB. 1997. Functional and regulatory analysis of the two copies of the *fixNOQP* operon of *Rhizobium leguminosarum* strain VF39. Mol Plant Microbe Interact 10:605–616. https://doi.org/10.1094/MPMI.1997.10.5.605.
137. Mandon K, Kaminski PA, Elmerich C. 1994. Functional analysis of the *fixNOQP* region of *Azorhizobium caulinodans*. J Bacteriol 176:2560–2568. https://doi.org/10.1128/jb.176.9.2560-2568.1994.
138. Smith A, Hill S, Anthony C. 1990. The purification, characterization and role of the *d*-type cytochrome oxidase of *Klebsiella pneumoniae* during nitrogen fixation. J Gen Microbiol 136:171–180. https://doi.org/10.1099/00221287-136-1-171.
139. Kelly MJ, Poole RK, Yates MG, Kennedy C. 1990. Cloning and mutagenesis of genes encoding the cytochrome *bd* terminal oxidase complex in *Azotobacter vinelandii*: mutants deficient in the cytochrome *d* complex are unable to fix nitrogen in air. J Bacteriol 172:6010–6019. https://doi.org/10.1128/jb.172.10.6010-6019.1990.
140. De Meyer SE, Briscoe L, Martinez-Hidalgo P, Agapakis CM, de-Los Santos PE, Seshadri R, Reeve W, Weinstock G, O'Hara G, Howieson JG, Hirsch AM. 2016. Symbiotic *Burkholderia* species show diverse arrangements of



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

*nif*/*fix* and *nod* genes and lack typical high-affinity cytochrome $cbb_3$ oxidase genes. Mol Plant Microbe Interact 29:609–619. https://doi.org/10.1094/MPMI-05-16-0091-R.
141. Studer D, Hennecke H, Müller M. 1992. High-pressure freezing of soybean nodules leads to an improved preservation of ultrastructure. Planta 188:155–163. https://doi.org/10.1007/BF00216809.
142. Verma DP, Bal AK. 1976. Intracellular site of synthesis and localization of leghemoglobin in root nodules. Proc Natl Acad Sci U S A 73:3843–3847. https://doi.org/10.1073/pnas.73.11.3843.
143. Robertson JG, Wells B, Bisseling T, Farnden KJF, Johnston AWB. 1984. Immuno-gold localization of leghaemoglobin in cytoplasm in nitrogen-fixing root nodules of pea. Nature 311:254–256. https://doi.org/10.1038/311254a0.
144. Bergersen FJ, Appleby CA. 1981. Leghaemoglobin within bacteroid-enclosing membrane envelopes from soybean root nodules. Planta 152:534–543. https://doi.org/10.1007/BF00380824.
145. Wang L, Rubio MC, Xin X, Zhang B, Fan Q, Wang Q, Ning G, Becana M, Duanmu D. 2019. CRISPR/Cas9 knockout of leghemoglobin genes in *Lotus japonicus* uncovers their synergistic roles in symbiotic nitrogen fixation. New Phytol 224:818–832. https://doi.org/10.1111/nph.16077.
146. Ott T, van Dongen JT, Gunther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi MK. 2005. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Curr Biol 15:531–535. https://doi.org/10.1016/j.cub.2005.01.042.
147. Bergkessel M, Basta DW, Newman DK. 2016. The physiology of growth arrest: uniting molecular and environmental microbiology. Nat Rev Microbiol 14:549–562. https://doi.org/10.1038/nrmicro.2016.107.
148. Jiang M, Sullivan SM, Walker AK, Strahler JR, Andrews PC, Maddock JR. 2007. Identification of novel *Escherichia coli* ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. J Bacteriol 189:3434–3444. https://doi.org/10.1128/JB.00090-07.
149. Häuser R, Pech M, Kijek J, Yamamoto H, Titz B, Naeve F, Tovchigrechko A, Yamamoto K, Szaflarski W, Takeuchi N, Stellberger T, Diefenbacher ME, Nierhaus KH, Uetz P. 2012. RsfA (YbeB) proteins are conserved ribosomal silencing factors. PLoS Genet 8:e1002815. https://doi.org/10.1371/journal.pgen.1002815.
150. Katayama Y, Kasahara A, Kuraishi H, Amano F. 1990. Regulation of activity of an ATP-dependent protease, Clp, by the amount of a subunit, ClpA, in the growth of *Escherichia coli* cells. J Biochem 108:37–41. https://doi.org/10.1093/oxfordjournals.jbchem.a123158.
151. Weichart D, Querfurth N, Dreger M, Hengge-Aronis R. 2003. Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli*. J Bacteriol 185:115–125. https://doi.org/10.1128/jb.185.1.115-125.2003.
152. Liu J, Francis LI, Jonas K, Laub MT, Chien P. 2016. ClpAP is an auxiliary protease for DnaA degradation in *Caulobacter crescentus*. Mol Microbiol 102:1075–1085. https://doi.org/10.1111/mmi.13537.
153. Bonnet M, Stegmann M, Maglica Z, Stiegeler E, Weber-Ban E, Hennecke H, Mesa S. 2013. $FixK_2$, a key regulator in *Bradyrhizobium japonicum*, is a substrate for the protease ClpAP *in vitro*. FEBS Lett 587:88–93. https://doi.org/10.1016/j.febslet.2012.11.014.
154. Román-Hernández G, Hou JY, Grant RA, Sauer RT, Baker TA. 2011. The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. Mol Cell 43:217–228. https://doi.org/10.1016/j.molcel.2011.06.009.
155. Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. Proc Natl Acad Sci U S A 103:5230–5235. https://doi.org/10.1073/pnas.0600912103.
156. Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J, Chaintreuil C, Nguyen P, Gressent F, Da Silva C, Poulain J, Wincker P, Rofidal V, Hem S, Barriere Q, Arrighi JF, Mergaert P, Giraud E. 2015. Convergent evolution of endosymbiont differentiation in *Dalbergioid* and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. Plant Physiol 169:1254–1265. https://doi.org/10.1104/pp.15.00584.
157. Sprent JI, Sutherland JM, de Faria SM, Dilworth MJ, Corby HDL, Becking JH, Materon LA, Drozd JW. 1987. Some aspects of the biology of nitrogen-fixing organisms. Philos Trans R Soc Lond B 317:111–129.
158. Sen D, Weaver RW, Bal AK. 1986. Structure and organization of effective peanut and cowpea root nodules induced by rhizobial strain 32H1. J Exp Bot 37:356–363. https://doi.org/10.1093/jxb/37.3.356.
159. Bonaldi K, Gargani D, Prin Y, Fardoux J, Gully D, Nouwen N, Goormachtig S, Giraud E. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction. Mol Plant Microbe Interact 24:1359–1371. https://doi.org/10.1094/MPMI-04-11-0093.
160. Guefrachi I, Pierre O, Timchenko T, Alunni B, Barriere Q, Czernic P, Villaecija-Aguilar JA, Verly C, Bourge M, Fardoux J, Mars M, Kondorosi E, Giraud E, Mergaert P. 2015. *Bradyrhizobium* BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene* legumes. Mol Plant Microbe Interact 28:1155–1166. https://doi.org/10.1094/MPMI-04-15-0094-R.
161. Alunni B, Kevei Z, Redondo-Nieto M, Kondorosi A, Mergaert P, Kondorosi E. 2007. Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. Mol Plant Microbe Interact 20:1138–1148. https://doi.org/10.1094/MPMI-20-9-1138.
162. Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol 132:161–173. https://doi.org/10.1104/pp.102.018192.
163. Montiel J, Downie JA, Farkas A, Bihari P, Herczeg R, Balint B, Mergaert P, Kereszt A, Kondorosi E. 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. Proc Natl Acad Sci U S A 114:5041–5046. https://doi.org/10.1073/pnas.1704217114.
164. Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaitre B, Alunni B, Bourge M, Kucho K, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E, Mergaert P. 2010. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science 327:1122–1126. https://doi.org/10.1126/science.1184057.
165. Wang D, Griffitts J, Starker C, Fedorova E, Limpens E, Ivanov S, Bisseling T, Long S. 2010. A nodule-specific protein secretory pathway required for nitrogen-fixing symbiosis. Science 327:1126–1129. https://doi.org/10.1126/science.1184096.
166. Pan H, Oztas O, Zhang X, Wu X, Stonoha C, Wang E, Wang B, Wang D. 2016. A symbiotic SNARE protein generated by alternative termination of transcription. Nat Plants 2:15197. https://doi.org/10.1038/nplants.2015.197.
167. Penterman J, Abo RP, De Nisco NJ, Arnold MF, Longhi R, Zanda M, Walker GC. 2014. Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. Proc Natl Acad Sci U S A 111:3561–3566. https://doi.org/10.1073/pnas.1400450111.
168. Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D. 2015. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. Proc Natl Acad Sci U S A 112:15238–15243. https://doi.org/10.1073/pnas.1500123112.
169. Long S, McCune S, Walker GC. 1988. Symbiotic loci of *Rhizobium meliloti* identified by random Tn*phoA* mutagenesis. J Bacteriol 170:4257–4265. https://doi.org/10.1128/jb.170.9.4257-4265.1988.
170. Karunakaran R, Haag AF, East AK, Ramachandran VK, Prell J, James EK, Scocchi M, Ferguson GP, Poole PS. 2010. BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. J Bacteriol 192:2920–2928. https://doi.org/10.1128/JB.00020-10.
171. diCenzo GC, Zamani M, Ludwig HN, Finan TM. 2017. Heterologous complementation reveals a specialized activity for BacA in the *Medicago-Sinorhizobium meliloti* symbiosis. Mol Plant Microbe Interact 30:312–324. https://doi.org/10.1094/MPMI-02-17-0030-R.
172. Glazebrook J, Ichige A, Walker GC. 1993. A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes Dev 7:1485–1497. https://doi.org/10.1101/gad.7.8.1485.
173. Haag AF, Baloban M, Sani M, Kerscher B, Pierre O, Farkas A, Longhi R, Boncompagni E, Herouart D, Dall'angelo S, Kondorosi E, Zanda M, Mergaert P, Ferguson GP. 2011. Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. PLoS Biol 9:e1001169. https://doi.org/10.1371/journal.pbio.1001169.
174. Maunoury N, Redondo-Nieto M, Bourcy M, Van de Velde W, Alunni B, Laporte P, Durand P, Agier N, Marisa L, Vaubert D, Delacroix H, Duc G, Ratet P, Aggerbeck L, Kondorosi E, Mergaert P. 2010. Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two transcriptome-switches. PLoS One 5:e9519. https://doi.org/10.1371/journal.pone.0009519.
175. Nallu S, Silverstein KA, Samac DA, Bucciarelli B, Vance CP, VandenBosch KA. 2013. Regulatory patterns of a large family of defensin-like genes



Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

expressed in nodules of *Medicago truncatula*. PLoS One 8:e60355. https://doi.org/10.1371/journal.pone.0060355.
176. Vinardell JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P, Kondorosi A, Kondorosi E. 2003. Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. Plant Cell 15:2093–2105. https://doi.org/10.1105/tpc.014373.
177. Becker A, Bergès H, Krol E, Bruand C, Rüberg S, Capela D, Lauber E, Meilhoc E, Ampe F, de Bruijn FJ, Fourment J, Francez-Charlot A, Kahn D, Küster H, Liebe C, Pühler A, Weidner S, Batut J. 2004. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. Mol Plant Microbe Interact 17:292–303. https://doi.org/10.1094/MPMI.2004.17.3.292.
178. Prell J, Poole P. 2006. Metabolic changes of rhizobia in legume nodules. Trends Microbiol 14:161–168. https://doi.org/10.1016/j.tim.2006.02.005.
179. Dunn MF. 2015. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. Crit Rev Microbiol 41:411–451. https://doi.org/10.3109/1040841X.2013.856854.
180. Lodwig EM, Hosie AH, Bourdes A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS. 2003. Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. Nature 422:722–726. https://doi.org/10.1038/nature01527.
181. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009. Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. Proc Natl Acad Sci U S A 106:12477–12482. https://doi.org/10.1073/pnas.0903653106.
182. Prell J, Bourdes A, Kumar S, Lodwig E, Hosie A, Kinghorn S, White J, Poole P. 2010. Role of symbiotic auxotrophy in the *Rhizobium*-legume symbioses. PLoS One 5:e13933. https://doi.org/10.1371/journal.pone.0013933.
183. Ferraioli S, Tate R, Cermola M, Favre R, Iaccarino M, Patriarca EJ. 2002. Auxotrophic mutant strains of *Rhizobium etli* reveal new nodule development phenotypes. Mol Plant Microbe Interact 15:501–510. https://doi.org/10.1094/MPMI.2002.15.5.501.
184. Zheng L, White RH, Cash VL, Dean DR. 1994. Mechanism for the desulfurization of L-cysteine catalyzed by the *nifS* gene product. Biochemistry 33:4714–4720. https://doi.org/10.1021/bi00181a031.
185. Prell J, Bourdes A, Karunakaran R, Lopez-Gomez M, Poole P. 2009. Pathway of $\gamma$-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. J Bacteriol 191:2177–2186. https://doi.org/10.1128/JB.01714-08.
186. Hakoyama T, Niimi K, Watanabe H, Tabata R, Matsubara J, Sato S, Nakamura Y, Tabata S, Jichun L, Matsumoto T, Tatsumi K, Nomura M, Tajima S, Ishizaka M, Yano K, Imaizumi-Anraku H, Kawaguchi M, Kouchi H, Suganuma N. 2009. Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature 462:514–517. https://doi.org/10.1038/nature08594.
187. Ronson CW, Lyttleton P, Robertson JG. 1981. $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. Proc Natl Acad Sci U S A 78:4284–4288. https://doi.org/10.1073/pnas.78.7.4284.
188. Glenn AR, Brewin NJ. 1981. Succinate-resistant mutants of *Rhizobium leguminosarum*. J Gen Microbiol 126:237–241. https://doi.org/10.1099/00221287-126-1-237.
189. Mitsch MJ, diCenzo GC, Cowie A, Finan TM. 2018. Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. Appl Environ Microbiol 84:e01561-17. https://doi.org/10.1128/AEM.01561-17.
190. Driscoll BT, Finan TM. 1993. $NAD^+$-dependent malic enzyme of *Rhizobium meliloti* is required for symbiotic nitrogen fixation. Mol Microbiol 7:865–873. https://doi.org/10.1111/j.1365-2958.1993.tb01177.x.
191. Zhang Y, Aono T, Poole P, Finan TM. 2012. $NAD(P)^+$-malic enzyme mutants of *Sinorhizobium* sp. strain NGR234, but not *Azorhizobium caulinodans* ORS571, maintain symbiotic $N_2$ fixation capabilities. Appl Environ Microbiol 78:2803–2812. https://doi.org/10.1128/AEM.06412-11.
192. Mulley G, Lopez-Gomez M, Zhang Y, Terpolilli J, Prell J, Finan T, Poole P. 2010. Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. J Bacteriol 192:4944–4953. https://doi.org/10.1128/JB.00294-10.
193. Flores-Tinoco CE, Tschan F, Fuhrer T, Margot C, Sauer U, Christen M, Christen B. 2020. Co-catabolism of arginine and succinate drives symbiotic nitrogen fixation. Mol Syst Biol 16:e9419. https://doi.org/10.15252/msb.20199419.
194. Djordjevic MA. 2004. *Sinorhizobium meliloti* metabolism in the root nodule: a proteomic perspective. Proteomics 4:1859–1872. https://doi.org/10.1002/pmic.200300802.
195. Walshaw DL, Wilkinson A, Mundy M, Smith M, Poole PS. 1997. Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in *Rhizobium leguminosarum*. Microbiology (Reading) 143(Part 7):2209–2221. https://doi.org/10.1099/00221287-143-7-2209.
196. Mortimer MW, McDermott TR, York GM, Walker GC, Kahn ML. 1999. Citrate synthase mutants of *Sinorhizobium meliloti* are ineffective and have altered cell surface polysaccharides. J Bacteriol 181:7608–7613. https://doi.org/10.1128/JB.181.24.7608-7613.1999.
197. Krishnan HB, Kim WS, Sun-Hyung J, Kim KY, Jiang G. 2003. Citrate synthase mutants of *Sinorhizobium fredii* USDA257 form ineffective nodules with aberrant ultrastructure. Appl Environ Microbiol 69:3561–3568. https://doi.org/10.1128/aem.69.6.3561-3568.2003.
198. Hernandez-Lucas I, Pardo MA, Segovia L, Miranda J, Martinez-Romero E. 1995. *Rhizobium tropici* chromosomal citrate synthase gene. Appl Environ Microbiol 61:3992–3997. https://doi.org/10.1128/AEM.61.11.3992-3997.1995.
199. Green LS, Emerich DW. 1997. The formation of nitrogen-fixing bacteroids is delayed but not abolished in soybean infected by an $\alpha$-ketoglutarate dehydrogenase-deficient mutant of *Bradyrhizobium japonicum*. Plant Physiol 114:1359–1368. https://doi.org/10.1104/pp.114.4.1359.
200. Shah R, Emerich DW. 2006. Isocitrate dehydrogenase of *Bradyrhizobium japonicum* is not required for symbiotic nitrogen fixation with soybean. J Bacteriol 188:7600–7608. https://doi.org/10.1128/JB.00671-06.
201. Thöny-Meyer L, Künzler P. 1996. The *Bradyrhizobium japonicum* aconitase gene (*acnA*) is important for free-living growth but not for an effective root nodule symbiosis. J Bacteriol 178:6166–6172. https://doi.org/10.1128/jb.178.21.6166-6172.1996.
202. Green LS, Li Y, Emerich DW, Bergersen FJ, Day DA. 2000. Catabolism of $\alpha$-ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. J Bacteriol 182:2838–2844. https://doi.org/10.1128/jb.182.10.2838-2844.2000.
203. Dunn MF. 1998. Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. FEMS Microbiol Rev 22:105–123. https://doi.org/10.1111/j.1574-6976.1998.tb00363.x.
204. Hansen RG, Henning U. 1966. Regulation of pyruvate dehydrogenase activity in *Escherichia coli* K12. Biochim Biophys Acta 122:355–358. https://doi.org/10.1016/0926-6593(66)90076-2.
205. Cevallos MA, Encarnación S, Leija A, Mora Y, Mora J. 1996. Genetic and physiological characterization of a *Rhizobium etli* mutant strain unable to synthesize poly-$\beta$-hydroxybutyrate. J Bacteriol 178:1646–1654. https://doi.org/10.1128/jb.178.6.1646-1654.1996.
206. Salminen SO, Streeter JG. 1990. Factors contributing to the accumulation of glutamate in *Bradyrhizobium japonicum* bacteroids under microaerobic conditions. J Gen Microbiol 136:2119–2126. https://doi.org/10.1099/00221287-136-10-2119.
207. Lodwig EM, Leonard M, Marroqui S, Wheeler TR, Findlay K, Downie JA, Poole PS. 2005. Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. Mol Plant Microbe Interact 18:67–74. https://doi.org/10.1094/MPMI-18-0067.
208. Lodwig E, Poole P. 2003. Metabolism of *Rhizobium* bacteroids. Crit Rev Plant Sci 22:37–78. https://doi.org/10.1080/713610850.
209. Wang C, Saldanha M, Sheng X, Shelswell KJ, Walsh KT, Sobral BWS, Charles TC. 2007. Roles of poly-3-hydroxybutyrate (PHB) and glycogen in symbiosis of *Sinorhizobium meliloti* with *Medicago* sp. Microbiology (Reading) 153:388–398. https://doi.org/10.1099/mic.0.29214-0.
210. Kim SA, Copeland L. 1996. Enzymes of poly-$\beta$-hydroxybutyrate metabolism in soybean and chickpea bacteroids. Appl Environ Microbiol 62:4186–4190. https://doi.org/10.1128/AEM.62.11.4186-4190.1996.
211. Quelas JI, Mesa S, Mongiardini EJ, Jendrossek D, Lodeiro AR. 2016. Regulation of polyhydroxybutyrate synthesis in the soil bacterium *Bradyrhizobium diazoefficiens*. Appl Environ Microbiol 82:4299–4308. https://doi.org/10.1128/AEM.00757-16.
212. Nishihata S, Kondo T, Tanaka K, Ishikawa S, Takenaka S, Kang C-M, Yoshida K-I. 2018. *Bradyrhizobium diazoefficiens* USDA110 PhaR functions for pleiotropic regulation of cellular processes besides PHB accumulation. BMC Microbiol 18:156. https://doi.org/10.1186/s12866-018-1317-2.
213. Allaway D, Lodwig EM, Crompton LA, Wood M, Parsons R, Wheeler TR, Poole PS. 2000. Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. Mol Microbiol 36:508–515. https://doi.org/10.1046/j.1365-2958.2000.01884.x.

Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

214. Rosendahl L, Dilworth MJ, Glenn AR. 1992. Exchange of metabolites across the peribacteroid membrane in pea root nodules. J Plant Physiol 139:635–638. https://doi.org/10.1016/S0176-1617(11)80385-8.
215. Waters JK, Hughes BL, II, Purcell LC, Gerhardt KO, Mawhinney TP, Emerich DW. 1998. Alanine, not ammonia, is excreted from $N_2$-fixing soybean nodule bacteroids. Proc Natl Acad Sci U S A 95:12038–12042. https://doi.org/10.1073/pnas.95.20.12038.
216. Li Y, Parsons R, Day DA, Bergersen FJ. 2002. Reassessment of major products of $N_2$ fixation by bacteroids from soybean root nodules. Microbiology (Reading) 148:1959–1966. https://doi.org/10.1099/00221287-148-6-1959.
217. Patriarca EJ, Tatè R, Iaccarino M. 2002. Key role of bacterial $NH_4^+$ metabolism in *Rhizobium*-plant symbiosis. Microbiol Mol Biol Rev 66:203–222. https://doi.org/10.1128/mmbr.66.2.203-222.2002.
218. Espin G, Moreno S, Guzman J. 1994. Molecular genetics of the glutamine synthetases in *Rhizobium* species. Crit Rev Microbiol 20:117–123. https://doi.org/10.3109/10408419409113551.
219. Lewis TA, Gonzalez R, Botsford JL. 1990. *Rhizobium meliloti* glutamate synthase: cloning and initial characterization of the *glt* locus. J Bacteriol 172:2413–2420. https://doi.org/10.1128/jb.172.5.2413-2420.1990.
220. Kondorosi A, Svab Z, Kiss GB, Dixon RA. 1977. Ammonia assimilation and nitrogen fixation in *Rhizobium meliloti*. Mol Gen Genet 151:221–226. https://doi.org/10.1007/BF00338698.
221. Castillo A, Taboada H, Mendoza A, Valderrama B, Encarnación S, Mora J. 2000. Role of GOGAT in carbon and nitrogen partitioning in *Rhizobium etli*. Microbiology (Reading) 146:1627–1637. https://doi.org/10.1099/00221287-146-7-1627.
222. Mulley G, White JP, Karunakaran R, Prell J, Bourdes A, Bunnewell S, Hill L, Poole PS. 2011. Mutation of GOGAT prevents pea bacteroid formation and $N_2$ fixation by globally downregulating transport of organic nitrogen sources. Mol Microbiol 80:149–167. https://doi.org/10.1111/j.1365-2958.2011.07565.x.
223. Sánchez-Cañizares C, Prell J, Pini F, Rutten P, Kraxner K, Wynands B, Karunakaran R, Poole PS. 2020. Global control of bacterial nitrogen and carbon metabolism by a $PTS^{Ntr}$-regulated switch. Proc Natl Acad Sci U S A 117:10234–10245. https://doi.org/10.1073/pnas.1917471117.
224. Patriarca EJ, Chiurazzi M, Manco G, Riccio A, Lamberti A, De Paolis A, Rossi M, Defez R, Iaccarino M. 1992. Activation of the *Rhizobium leguminosarum glnII* gene by NtrC is dependent on upstream DNA sequences. Mol Gen Genet 234:337–345. https://doi.org/10.1007/BF00538692.
225. diCenzo GC, Zamani M, Checcucci A, Fondi M, Griffitts JS, Finan TM, Mengoni A. 2019. Multidisciplinary approaches for studying rhizobium-legume symbioses. Can J Microbiol 65:1–33. https://doi.org/10.1139/cjm-2018-0377.
226. Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? Nat Biotechnol 28:245–248. https://doi.org/10.1038/nbt.1614.
227. Nielsen J. 2017. Systems biology of metabolism. Annu Rev Biochem 86:245–275. https://doi.org/10.1146/annurev-biochem-061516-044757.
228. Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. 2011. Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. BMC Syst Biol 5:120. https://doi.org/10.1186/1752-0509-5-120.
229. Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BO. 2007. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. PLoS Comput Biol 3:1887–1895. https://doi.org/10.1371/journal.pcbi.0030192.
230. Zhao H, Li M, Fang K, Chen W, Wang J. 2012. *In silico* insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. PLoS One 7:e31287. https://doi.org/10.1371/journal.pone.0031287.
231. Yang Y, Hu XP, Ma BG. 2017. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states. Mol Biosyst 13:607–620. https://doi.org/10.1039/c6mb00553e.
232. Pfau T, Christian N, Masakapalli SK, Sweetlove LJ, Poolman MG, Ebenhoh O. 2018. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. Sci Rep 8:12504. https://doi.org/10.1038/s41598-018-30884-x.
233. diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. Nat Commun 11:765271. https://doi.org/10.1038/s41467-020-16484-2.
234. diCenzo GC, Checcucci A, Bazzicalupo M, Mengoni A, Viti C, Dziewit L, Finan TM, Galardini M, Fondi M. 2016. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nat Commun 7:12219. https://doi.org/10.1038/ncomms12219.
235. Contador CA, Lo SK, Chan SHJ, Lam HM. 2020. Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. mSystems 5:e00516-19. https://doi.org/10.1128/mSystems.00516-19.
236. Marroquí S, Zorreguieta A, Santamaría C, Temprano F, Soberón M, Megías M, Downie JA. 2001. Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. J Bacteriol 183:854–864. https://doi.org/10.1128/JB.183.3.854-864.2001.
237. Kuzma MM, Winter H, Storer P, Oresnik I, Atkins CA, Layzell DB. 1999. The site of oxygen limitation in soybean nodules. Plant Physiol 119:399–408. https://doi.org/10.1104/pp.119.2.399.
238. Miller RW, McRae DG, Al-Jobore A, Berndt WB. 1988. Respiration supported nitrogenase activity of isolated *Rhizobium meliloti* bacteroids. J Cell Biochem 38:35–49. https://doi.org/10.1002/jcb.240380105.
239. Schuster S, Hilgetag C. 1994. On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 02:165–182. https://doi.org/10.1142/S0218339094000131.
240. Kiers ET, Rousseau RA, Denison RF. 2006. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. Evol Ecol Res 8:1077–1086.
241. Kiers ET, Rousseau RA, West SA, Denison RF. 2003. Host sanctions and the legume-rhizobium mutualism. Nature 425:78–81. https://doi.org/10.1038/nature01931.
242. Sachs JL, Quides KW, Wendlandt CE. 2018. Legumes versus rhizobia: a model for ongoing conflict in symbiosis. New Phytol 219:1199–1206. https://doi.org/10.1111/nph.15222.
243. West SA, Kiers ET, Simms EL, Denison RF. 2002. Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proc Biol Sci 269:685–694. https://doi.org/10.1098/rspb.2001.1878.
244. Oono R, Denison RF, Kiers ET. 2009. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? New Phytol 183:967–979. https://doi.org/10.1111/j.1469-8137.2009.02941.x.
245. Berrabah F, Ratet P, Gourion B. 2015. Multiple steps control immunity during the intracellular accommodation of rhizobia. J Exp Bot 66:1977–1985. https://doi.org/10.1093/jxb/eru545.
246. Regus JU, Quides KW, O'Neill MR, Suzuki R, Savory EA, Chang JH, Sachs JL. 2017. Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. Am J Bot 104:1299–1312. https://doi.org/10.3732/ajb.1700165.
247. Daubech B, Remigi P, Doin de Moura G, Marchetti M, Pouzet C, Auriac M-C, Gokhale CS, Masson-Boivin C, Capela D. 2017. Spatio-temporal control of mutualism in legumes helps spread symbiotic nitrogen fixation. Elife 6:e28683. https://doi.org/10.7554/eLife.28683.
248. Heath KD, Tiffin P. 2009. Stabilizing mechanisms in a legume-rhizobium mutualism. Evolution 63:652–662. https://doi.org/10.1111/j.1558-5646.2008.00582.x.
249. Batstone RT, O'Brien AM, Harrison TL, Frederickson ME. 2020. Experimental evolution makes microbes more cooperative with their local host genotype. Science 370:476–478. https://doi.org/10.1126/science.abb7222.
250. Lambert I, Pervent M, Le Quere A, Clement G, Tauzin M, Severac D, Benezech C, Tillard P, Martin-Magniette M-L, Colella S, Lepetit M. 2020. Responses of mature symbiotic nodules to the whole-plant systemic nitrogen signaling. J Exp Bot 71:5039–5052. https://doi.org/10.1093/jxb/eraa221.
251. Benezech C, Doudement M, Gourion B. 2020. Legumes tolerance to rhizobia is not always observed and not always deserved. Cell Microbiol 22:e13124. https://doi.org/10.1111/cmi.13124.
252. Gourion B, Berrabah F, Ratet P, Stacey G. 2015. *Rhizobium*-legume symbioses: the crucial role of plant immunity. Trends Plant Sci 20:186–194. https://doi.org/10.1016/j.tplants.2014.11.008.
253. Berrabah F, Ratet P, Gourion B. 2019. Legume nodules: massive infection in the absence of defense induction. Mol Plant Microbe Interact 32:35–44. https://doi.org/10.1094/MPMI-07-18-0205-FI.

Downloaded from https://journals.asm.org/journal/jb on 08 August 2021 by 129.67.119.210.

**Supplementary information**

**Supplementary methods**

To find common adaptations of rhizobia to the nitrogen-fixing bacteroid state, we performed a meta-analysis with available transcriptomics and proteomics datasets which compare bacteroids with a defined non-bacteroid state, usually a free-living liquid culture. We only included datasets that compare wild-type bacteria in different lifestyles. Hence, we excluded datasets comparing bacteroids on different hosts, mutant analyses, and datasets acquired for bacteroids without a free-living reference state. COG classifications were obtained from the website of the Joint Genome Institute (https://genome.jgi.doe.gov/). For *Bradyrhizobium* ORS285 and *Mesorhizobium huakuii* 7653R, FASTA files containing the protein products annotated on the genome were obtained from the NCBI website and the COG classification was performed using WebMGA (1). The microarray data from (2) (GEO Series GSE48090) were analysed with the samr package (3) in R (4) using default parameters. Differentially expressed genes were determined according to the methods stated in the respective study. A COG category was considered as up- or downregulated in a given dataset if all differentially expressed genes associated with this category were consistently up- or downregulated, i.e. a category was not considered if it included both up- and downregulated genes in this dataset.
To determine significant up- or downregulation of a subsystem, a hypergeometric test was performed using the `phyper` function in R. *P* values were subjected to Bonferroni multiple test correction.

**Table S1. COG categories significantly enriched within differentially expressed genes.**

| Dataset | COG categories enriched within upregulated genes | COG categories enriched within downregulated genes |
|---|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 – *Pisum sativum* | Energy production and conversion | Translation, ribosomal structure and biogenesis |
| *Rhizobium leguminosarum* bv. *viciae* 3841 – *Vicia cracca* | Defense mechanisms<br>Energy production and conversion<br>Posttranslational modification, protein turnover, chaperones | Cell cycle control, cell division, chromosome partitioning<br>Cell wall/membrane/envelope biogenesis<br>Translation, ribosomal structure and biogenesis |
| *Rhizobium leguminosarum* bv. *viciae* A34 – *Pisum sativum* | Energy production and conversion | Intracellular trafficking, secretion, and vesicular transport |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 – *Phaseolus vulgaris* | Energy production and conversion<br>Posttranslational modification, protein turnover, chaperones<br>Translation, ribosomal structure and biogenesis | Amino acid transport and metabolism<br>Carbohydrate transport and metabolism<br>Cell motility<br>Inorganic ion transport and metabolism |
| *Rhizobium etli* CFN42 – *Phaseolus vulgaris* (exponential phase) | Energy production and conversion<br>Inorganic ion transport and metabolism<br>Mobilome: prophages, transposons | Cell motility<br>Translation, ribosomal structure and biogenesis |
| *Rhizobium etli* CFN42 – *Phaseolus vulgaris* (stationary phase) | Energy production and conversion<br>Inorganic ion transport and metabolism<br>Translation, ribosomal structure and biogenesis | Function unknown<br>Posttranslational modification, protein turnover, chaperones |
| *Sinorhizobium meliloti* 2011 – *Medicago truncatula* | Cell motility<br>General function prediction only | Cell wall/membrane/envelope biogenesis<br>Intracellular trafficking, secretion, and vesicular transport<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Sinorhizobium meliloti* 1021 – *Medicago truncatula* | Transcription | Cell motility<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Sinorhizobium meliloti* 1021 – *Medicago sativa* cv. Europe | Energy production and conversion | Cell motility<br>Inorganic ion transport and metabolism<br>Translation, ribosomal structure and biogenesis |

| | | |
|---|---|---|
| *Sinorhizobium medicae* – *Medicago truncatula* | Translation, ribosomal structure and biogenesis | Translation, ribosomal structure and biogenesis |
| *Sinorhizobium* sp. NGR234 - *Vigna unguiculata* | Energy production and conversion<br>Inorganic ion transport and metabolism<br>Posttranslational modification, protein turnover, chaperones | Cell motility<br>Cell wall/membrane/envelope biogenesis<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Sinorhizobium* sp. NGR234 – *Leucaena leucocephala* | Energy production and conversion<br>General function prediction only<br>Inorganic ion transport and metabolism | Cell motility<br>Cell wall/membrane/envelope biogenesis<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Bradyrhizobium* sp. ORS285 – *Aeschynomene afraspera* | Cell wall/membrane/envelope biogenesis<br>Inorganic ion transport and metabolism<br>Posttranslational modification, protein turnover, chaperones | Cell motility<br>Extracellular structures<br>Signal transduction mechanisms |
| *Bradyrhizobium* sp. ORS285 – *Aeschynomene indica* | Inorganic ion transport and metabolism<br>Posttranslational modification, protein turnover, chaperones | Extracellular structures<br>Signal transduction mechanisms |
| *Bradyrhizobium japonicum* USDA110 – *Glycine max* | Energy production and conversion<br>Inorganic ion transport and metabolism<br>Posttranslational modification, protein turnover, chaperones | Cell cycle control, cell division, chromosome partitioning<br>Cell motility<br>Cell wall/membrane/envelope biogenesis<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Bradyrhizobium japonicum* USDA110 – *Vigna unguiculata* | Inorganic ion transport and metabolism | Cell motility<br>Nucleotide transport and metabolism<br>Translation, ribosomal structure and biogenesis |
| *Bradyrhizobium japonicum* USDA110 – *Macroptilium atropurpureum* | Inorganic ion transport and metabolism<br>Energy production and conversion | Cell cycle control, cell division, chromosome partitioning<br>Cell motility<br>Cell wall/membrane/envelope biogenesis<br>Nucleotide transport and metabolism<br>Replication, recombination and repair<br>Translation, ribosomal structure and biogenesis |
| *Azorhizobium caulinodans* ORS571 – *Sesbania rostrata* | Energy production and conversion<br>Posttranslational modification, protein turnover, chaperones | Cell motility<br>Cell wall/membrane/envelope biogenesis |

| | | |
|---|---|---|
| *Mesorhizobium haukuii* 7653R – *Astragalus sinicus* | General function prediction only<br>Inorganic ion transport and metabolism | Cell motility<br>Cell wall/membrane/envelope biogenesis<br>Extracellular structures<br>Intracellular trafficking, secretion, and vesicular transport<br>Translation, ribosomal structure and biogenesis |

$P < 0.05$ was considered to be significant.

**Table S2. COGs upregulated in at least 10 of the 19 analysed datasets.**

| COG | Description | Number of datasets |
|---|---|---|
| | **Electron transfer and respiratory chain** | |
| COG0348 | Polyferredoxin | 17 |
| COG2993 | Cbb3-type cytochrome oxidase, cytochrome c subunit | 15 |
| COG3278 | Cbb3-type cytochrome oxidase, subunit 1 | 15 |
| COG2440 | Ferredoxin-like protein FixX | 13 |
| COG2025 | Electron transfer flavoprotein, alpha subunit | 12 |
| COG2086 | Electron transfer flavoprotein, alpha and beta subunits | 12 |
| COG4736 | Cbb3-type cytochrome oxidase, subunit 3 | 12 |
| COG2010 | Cytochrome c, mono- and diheme variants | 11 |
| | **Nitrogenase and Fe-S cluster** | |
| COG3197 | Uncharacterized protein, possibly involved in nitrogen fixation | 15 |
| COG1348 | Nitrogenase subunit NifH | 14 |
| COG0535 | Radical SAM superfamily enzyme, MoaA/NifB/PqqE/SkfB family | 14 |
| COG5456 | Nitrogen fixation protein FixH | 13 |
| COG2710 | Nitrogenase molybdenum-iron protein, alpha and beta chains | 13 |
| COG0316 | Fe-S cluster assembly iron-binding protein IscA | 12 |
| COG2151 | Metal-sulfur cluster biosynthetic enzyme | 11 |
| COG3604 | Transcriptional regulator containing GAF, AAA-type ATPase, and DNA-binding Fis domains | 11 |
| COG1104 | Cysteine sulfinate desulfinase/cysteine desulfurase or related enzyme | 11 |
| | **Other** | |
| COG0071 | Molecular chaperone IbpA, HSP20 family | 14 |
| COG2146 | Ferredoxin subunit of nitrite reductase or a ring-hydroxylating dioxygenase | 13 |
| COG0799 | Ribosomal silencing factor RsfS, regulates association of 30S and 50S subunits | 12 |
| COG0644 | Dehydrogenase (flavoprotein) | 12 |
| COG3221 | ABC-type phosphate/phosphonate transport system, periplasmic component | 12 |
| COG2908 | UDP-2,3-diacylglucosamine pyrophosphatase LpxH | 11 |
| COG2766 | Predicted Ser/Thr protein kinase | 11 |
| COG3467 | Nitroimidazol reductase NimA or a related FMN-containing flavoprotein, pyridoxamine 5'-phosphate oxidase superfamily | 11 |
| COG0450 | Alkyl hydroperoxide reductase subunit AhpC (peroxiredoxin) | 10 |
| COG3247 | Uncharacterized membrane protein HdeD, DUF308 family | 10 |
| COG0229 | Peptide methionine sulfoxide reductase MsrB | 10 |
| COG0265 | Periplasmic serine protease, S1-C subfamily, contain C-terminal PDZ domain | 10 |
| COG1117 | ABC-type phosphate transport system, ATPase component | 10 |
| COG0568 | DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32) | 10 |

| | | |
|---|---|---|
| COG3396 | 1,2-phenylacetyl-CoA epoxidase, catalytic subunit | 10 |
| COG5457 | Uncharacterized conserved protein YjiS, DUF1127 family | 10 |

The number of datasets refers to those datasets in which the COG was unambiguously upregulated, i.e. if a COG was associated with both up- and downregulated genes in a dataset, it was not included in the numbers given in this table.

**Table S3. COGs downregulated in at least 10 out of 19 analysed datasets.**

| COG | Description | Number of datasets |
|---|---|---|
| | **Translation** | |
| COG0048 | Ribosomal protein S12 | 15 |
| COG0088 | Ribosomal protein L4 | 14 |
| COG0522 | Ribosomal protein S4 or related protein | 14 |
| COG0186 | Ribosomal protein S17 | 14 |
| COG0049 | Ribosomal protein S7 | 14 |
| COG0096 | Ribosomal protein S8 | 14 |
| COG0097 | Ribosomal protein L6P/L9E | 13 |
| COG0098 | Ribosomal protein S5 | 13 |
| COG0087 | Ribosomal protein L3 | 13 |
| COG0198 | Ribosomal protein L24 | 13 |
| COG0197 | Ribosomal protein L16/L10AE | 13 |
| COG0200 | Ribosomal protein L15 | 13 |
| COG0228 | Ribosomal protein S16 | 13 |
| COG0093 | Ribosomal protein L14 | 13 |
| COG0094 | Ribosomal protein L5 | 13 |
| COG0238 | Ribosomal protein S18 | 13 |
| COG0089 | Ribosomal protein L23 | 12 |
| COG0051 | Ribosomal protein S10 | 12 |
| COG0090 | Ribosomal protein L2 | 12 |
| COG0092 | Ribosomal protein S3 | 12 |
| COG0103 | Ribosomal protein S9 | 12 |
| COG0050 | Translation elongation factor EF-Tu, a GTPase | 12 |
| COG0185 | Ribosomal protein S19 | 12 |
| COG1825 | Ribosomal protein L25 (general stress protein Ctc) | 12 |
| COG0199 | Ribosomal protein S14 | 12 |
| COG0360 | Ribosomal protein S6 | 12 |
| COG0244 | Ribosomal protein L10 | 11 |
| COG0080 | Ribosomal protein L11 | 11 |
| COG0091 | Ribosomal protein L22 | 11 |
| COG0255 | Ribosomal protein L29 | 11 |
| COG0102 | Ribosomal protein L13 | 11 |
| COG0052 | Ribosomal protein S2 | 11 |
| COG1841 | Ribosomal protein L30/L7E | 11 |
| COG0806 | Ribosomal 30S subunit maturation factor RimM, required for 16S rRNA processing | 11 |
| COG0256 | Ribosomal protein L18 | 11 |
| COG0359 | Ribosomal protein L9 | 10 |
| COG0292 | Ribosomal protein L20 | 10 |
| COG0081 | Ribosomal protein L1 | 10 |
| COG0100 | Ribosomal protein S11 | 10 |

| | | |
|---|---|---|
| COG0203 | Ribosomal protein L17 | 10 |
| COG0539 | Ribosomal protein S1 | 10 |
| | **Lipid and cell envelope synthesis** | |
| COG0236 | Acyl carrier protein | 15 |
| COG3206 | Uncharacterized protein involved in exopolysaccharide biosynthesis | 11 |
| COG1686 | D-alanyl-D-alanine carboxypeptidase | 11 |
| COG0764 | 3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratase | 11 |
| COG1044 | UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase | 10 |
| COG1560 | Lipid A biosynthesis lauroyl acyltransferase | 10 |
| COG0416 | Phosphate:acyl-[acyl carrier protein] acyltransferase | 10 |
| COG0774 | UDP-3-O-acyl-N-acetylglucosamine deacetylase | 10 |
| | **Chemotaxis and motility** | |
| COG1344 | Flagellin and related hook-associated protein FlgL | 16 |
| COG0784 | CheY chemotaxis protein or a CheY-like REC (receiver) domain | 13 |
| COG0835 | Chemotaxis signal transduction protein | 12 |
| COG1558 | Flagellar basal body rod protein FlgC | 12 |
| COG5442 | Flagellar biosynthesis regulator FlaF | 12 |
| COG5443 | Flagellar biosynthesis regulator FlbT | 12 |
| COG0840 | Methyl-accepting chemotaxis protein | 11 |
| COG1291 | Flagellar motor component MotA | 11 |
| COG1749 | Flagellar hook protein FlgE | 11 |
| COG4786 | Flagellar basal body rod protein | 11 |
| COG3847 | Flp pilus assembly protein, pilin Flp | 11 |
| COG0643 | Chemotaxis protein histidine kinase CheA | 10 |
| COG4960 | Flp pilus assembly protein, protease CpaA | 10 |
| COG1815 | Flagellar basal body rod protein FlgB | 10 |
| COG4962 | Pilus assembly protein | 10 |
| | **ATP synthase** | |
| COG0356 | FoF1-type ATP synthase, membrane subunit a | 12 |
| COG0056 | FoF1-type ATP synthase, alpha subunit | 11 |
| COG0712 | FoF1-type ATP synthase, delta subunit | 11 |
| COG0224 | FoF1-type ATP synthase, gamma subunit | 11 |
| COG0355 | FoF1-type ATP synthase, epsilon subunit | 11 |
| COG0055 | FoF1-type ATP synthase, beta subunit | 10 |
| | **DNA/RNA- associated proteins** | |
| COG0776 | Bacterial nucleoid DNA-binding protein | 13 |
| COG0188 | DNA gyrase/topoisomerase IV, subunit A | 12 |
| COG1197 | Transcription-repair coupling factor (superfamily II helicase) | 10 |
| COG0863 | DNA modification methylase | 10 |
| | **Cell division** | |
| COG1674 | DNA segregation ATPase FtsK/SpoIIIE and related proteins | 11 |
| COG0849 | Cell division ATPase FtsA | 10 |

| | | |
|---|---|---|
| COG0206 | Cell division GTPase FtsZ | 10 |
| COG1192 | Putative partition protein | 10 |
| | **Other** | |
| COG0499 | S-adenosylhomocysteine hydrolase | 13 |
| COG0209 | Ribonucleotide reductase alpha subunit | 12 |
| COG2945 | Alpha/beta superfamily hydrolase | 12 |
| COG5330 | Uncharacterized conserved protein, DUF2336 family | 12 |
| COG0690 | Preprotein translocase subunit SecE | 11 |
| COG0439 | Biotin carboxylase | 11 |
| COG0544 | FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor) | 11 |
| COG4214 | ABC-type xylose transport system, permease component | 11 |
| COG0652 | Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family | 11 |
| COG0609 | Hemin ABC transporter permease protein | 11 |
| COG0109 | Protoheme IX farnesyltransferase | 11 |
| COG0136 | Aspartate-semialdehyde dehydrogenase | 10 |
| COG3750 | Uncharacterized conserved protein, UPF0335 family | 10 |
| COG3672 | Predicted transglutaminase-like cysteine proteinase | 10 |
| COG0250 | Transcription antitermination factor NusG | 10 |
| COG4530 | Uncharacterized protein | 10 |
| COG1034 | NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) | 10 |
| COG1862 | Preprotein translocase subunit YajC | 10 |
| COG3195 | 2-oxo-4-hydroxy-4-carboxy--5-ureidoimidazoline (OHCU) decarboxylase | 10 |
| COG0054 | 6,7-dimethyl-8-ribityllumazine synthase (Riboflavin synthase beta chain) | 10 |
| COG0757 | 3-dehydroquinate dehydratase | 10 |
| COG3175 | Cytochrome *c* oxidase assembly protein | 10 |
| COG4775 | Beta-barrel assembly machine subunit BamA | 10 |
| COG0201 | Protein translocase subunit SecY/Sec61 alpha | 10 |
| COG0797 | Probable rare lipoprotein A | 10 |
| COG4454 | Uncharacterized copper-binding protein, cupredoxin-like subfamily | 10 |
| COG0406 | Phosphoglycerate mutase | 10 |
| COG0105 | Nucleoside diphosphate kinase | 10 |

Number of datasets refers to those datasets in which the COG was unambiguously downregulated, i.e. if a COG was associated with both up- and downregulated genes in a dataset, it was not included in the numbers given in this table.

**Table S4. Metabolic modelling studies of bacteroids.**

| Species | Model size | Predictions and key messages | Reference |
|---|---|---|---|
| *Rhizobium etli* | Bacteroid-specific;<br>363 genes<br>387 reactions<br>371 metabolites | Usage of an incomplete TCA cycle and active oxidative phosphorylation required for SNF; no activity of GS-GOGAT; importance of *myo*-inositol catabolism for maximum nitrogen fixation; increase in SNF when PHB and glycogen synthesis is abolished; importance of arginine deiminase for SNF when arginine uptake and an arginine-ornithine antiporter are included in the model; oxygen limitation of bacteroid metabolism | (5) |
| *Rhizobium etli* | Bacteroid-specific expansion of (*1*)<br>450 genes<br>405 reactions<br>377 metabolites | Extensive comparison with genome-scale data (transcriptome, proteome); expansion of objective function to maximise agreement with *omics* data; activity of the full TCA cycle; pyruvate dehydrogenase not essential, but deletion reduces SNF; PHB and glycogen produced during SNF; activity of gluconeogenesis as well as purine and pyrimidine metabolism | (6) |
| *Sinorhizobium meliloti* | Bacteroid-specific;<br>565 genes<br>503 reactions<br>522 metabolites | Activity of the full TCA cycle; flux through gluconeogenesis; dme essential for SNF; amino acid cycling between plant and bacteroid: plant provides glutamate, bacteroid secretes alanine and aspartate | (7) |
| *Sinorhizobium meliloti* | Genome-scale;<br>1575 genes<br>1825 reactions<br>1579 metabolites | Importance of genes on the different replicons for adaptation to different environments; strong decrease of active reactions during SNF compared to free-living bacteria; important metabolic functions during SNF: energy production and conversion, coenzyme transport and metabolism; decreased importance of chromosomal genes during SNF; contribution of megaplasmid pSymA to metabolic functions relevant for establishing the symbiosis | (8) |

| | | | |
|---|---|---|---|
| *Bradyrhizobium japonicum* | Genome-scale;<br>1101 genes<br>1031 reactions<br>661 metabolites | Condition-specific model derived from transcriptome and proteome data; highest total nitrogen output when ammonia, alanine and aspartate are secreted by the bacteroid; full TCA cycle and active oxidative phosphorylation; PHB accumulation; strongly reduced number of active reactions during SNF | (9) |
| *Sinorhizobium meliloti* – *Medicago truncatula* | *S. meliloti*: genome-scale<br><br>*M. truncatula*:<br>3403 genes<br>2909 reactions<br>2780 metabolites | Oxygen limitation of bacteroid metabolism; increasing use of succinate instead of malate at higher oxygen concentrations; amino acid cycling: plant provides glutamate, bacteroid secretes alanine, especially at low oxygen concentrations | (10) |
| *Sinorhizobium meliloti* – *Medicago truncatula* | *S. meliloti*: genome-scale;<br>1348 genes<br>1407 reactions<br>1160 metabolites<br><br>*M. truncatula* (updated from(10)):<br>2522 genes<br>2920 reactions<br>2742 metabolites | Activity of TCA cycle, pentose-phosphate pathway, oxidative phosphorylation and purine metabolism during SNF; sugars provided as additional carbon source for differentiating bacteroids; ammonium as main export product; increased plant growth when bacteroids catabolise sucrose instead of dicarboxylates; protons required for dicarboxylate metabolism; oxygen limitation of nodule metabolism | (11) |
| *Sinorhizobium fredii* | Bacteroid-specific;<br>541 genes<br>538 reactions<br>508 metabolites | Activity of full TCA cycle and oxidative phosphorylation; PHB and glycogen synthesis during SNF; *myo*-inositol as a rate-limiting substrate; integration with transcriptomic data predicts improved efficiency of SNF in cultivated soybean compared to wild variant | (12) |

SNF: symbiotic nitrogen fixation

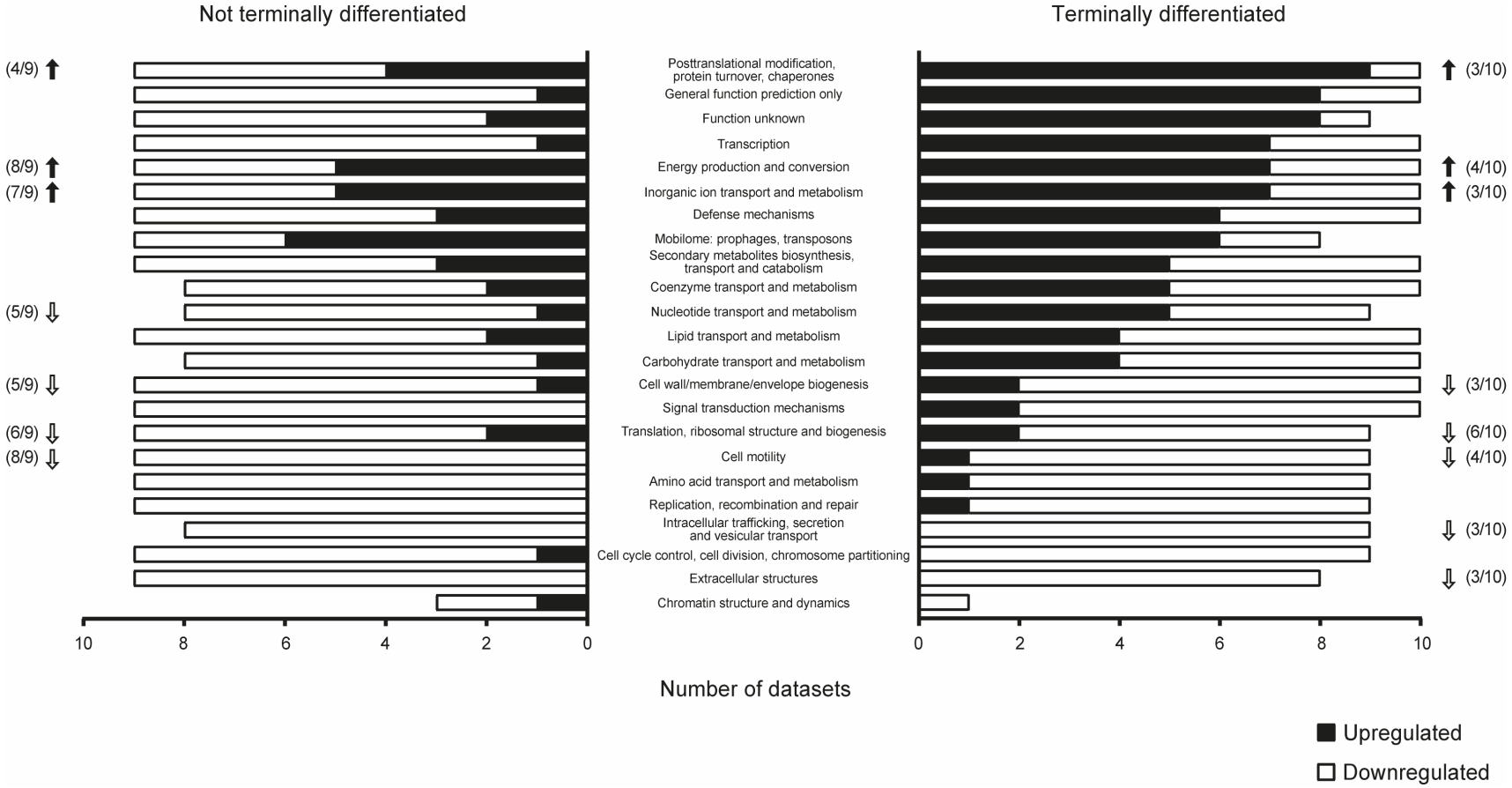



**Figure S1. Central features of terminally differentiated and not terminally differentiated bacteroids.** The data shown in Fig. 1 were divided into those for not terminally differentiated (left) and terminally differentiated (right) bacteroids. The bar graphs show the number of datasets in which a COG category contained more upregulated than downregulated (black) or more downregulated than upregulated (white) genes/proteins. Where the total number is smaller than 9 (left) or 10 (right), some datasets contained either no differentially expressed genes/proteins or an equal number of up- and downregulated genes/proteins in this category. Arrows indicate the 3 categories that were most often significantly enriched within upregulated (black) or downregulated (white) genes/proteins, with numbers in brackets indicating the number of datasets for which significant up-/downregulation was observed. *P* values were determined with a hypergeometric test followed by Bonferroni multiple test correction and $P < 0.05$ was considered significant.

**References**


1. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: A customizable web server for fast metagenomic sequence analysis. BMC Genomics 12:444.
2. Peng J, Hao B, Liu L, Wang S, Ma B, Yang Y, Xie F, Li Y. 2014. RNA-Seq and microarrays analyses reveal global differential transcriptomes of *Mesorhizobium huakuii* 7653R between bacteroids and free-living cells. PLoS One 9:e93626.
3. Tusher VG, Tibshirani R, Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98:5116-21.
4. R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
5. Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BO. 2007. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. PLoS Comput Biol 3:1887-95.
6. Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. 2011. Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling. BMC Syst Biol 5:120.
7. Zhao H, Li M, Fang K, Chen W, Wang J. 2012. *In silico* insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. PLoS One 7:e31287.
8. diCenzo GC, Checcucci A, Bazzicalupo M, Mengoni A, Viti C, Dziewit L, Finan TM, Galardini M, Fondi M. 2016. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nat Commun 7:12219.
9. Yang Y, Hu XP, Ma BG. 2017. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: A comparison between free-living and symbiotic states. Mol BioSyst 13:607-620.
10. Pfau T, Christian N, Masakapalli SK, Sweetlove LJ, Poolman MG, Ebenhoh O. 2018. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. Sci Rep 8:12504.
11. diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. Nat Commun 11:765271.
12. Contador CA, Lo SK, Chan SHJ, Lam HM. 2020. Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. mSystems 5:00516-19.

## Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

| | |
|---|---|
| Title of Paper | How rhizobia adapt to the nodule environment |
| Publication Status | Published |
| Publication Details | R. Ledermann*, C. C. M. Schulte*, P. S. Poole, How rhizobia adapt to the nodule environment. *Journal of Bacteriology* **203** (12), e00539 – 20 (2021). *equal contribution |

### Student Confirmation

| | | | |
|---|---|---|---|
| Student Name: | Carolin C. M. Schulte | | |
| Contribution to the Paper | performed meta-analysis, wrote manuscript together with R. Ledermann | | |
| Signature | Carolin Schulte | Date | 9th August 2021 |

### Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

| | | | |
|---|---|---|---|
| Supervisor name and title: Professor Philip Poole | | | |
| Supervisor comments<br>Excellent review written jointly by Raphael Ledermann and Carolin Schulte with all the modelling incorporated into the review performed by Carolin | | | |
| Signature | Philip Poole | Date | 10th August 2021 |

This completed form should be included in the thesis, at the end of the relevant chapter.

# 3. Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum*

## Summary

In this chapter, I reconstruct a genome-scale metabolic model for Rlv3841. Rlv3841 is a rhizobial model strain for which an extensive body of literature exists, making it an ideal candidate for the application of metabolic modelling approaches. After curation and validation of the model, I use various experimental datasets to create condition-specific models describing the metabolism of Rlv3841 as free-living bacteria in the rhizosphere, as undifferentiated nodule bacteria, and as nitrogen-fixing bacteroids. Through the integration of experimental data, nutrient uptake is defined based on the expression of genes associated with metabolic and transport reactions to avoid constraining nutrient uptake rates for which direct experimental measurements are not available. The condition-specific models are then analysed to elucidate nutrient uptake and metabolic pathway usage for the different lifestyles. Model predictions are validated through comparison with the literature where possible, and predictions warranting further experimental studies are discussed. This manuscript has been submitted for publication[47].

# Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum*

Carolin C. M. Schulte[1,2], Vinoy K. Ramachandran[1], Antonis Papachristodoulou[2]#, Philip S. Poole[1]#

[1]Department of Plant Sciences, University of Oxford, Oxford, UK

[2]Department of Engineering Science, University of Oxford, Oxford, UK

Running title: Metabolic modelling of rhizobial lifestyles

#correspondence to antonis@eng.ox.ac.uk or philip.poole@plants.ox.ac.uk

**Abstract**

Biological nitrogen fixation in rhizobium-legume symbioses is of major importance for sustainable agricultural practices. To establish a mutualistic relationship with their plant host, rhizobia transition from free-living bacteria in soil to growth down infection threads inside plant roots and finally differentiate into nitrogen-fixing bacteroids. We reconstructed a genome-scale metabolic model for *Rhizobium leguminosarum* and integrated the model with transcriptome, proteome, metabolome and gene essentiality data to investigate nutrient uptake and metabolic fluxes characteristic of these different lifestyles. Synthesis of leucine, polyphosphate and AICAR is predicted to be important in the rhizosphere, while *myo*-inositol catabolism is active in undifferentiated nodule bacteria in agreement with experimental evidence. The model indicates that bacteroids utilize xylose and glycolate in addition to dicarboxylates, which could explain previously described gene expression patterns. Histidine is predicted to be actively synthesized in bacteroids, consistent with transcriptome and proteome data for several rhizobial species. These results provide the basis for targeted experimental investigation of metabolic processes specific to the different stages of the rhizobium-legume symbioses.

## Importance

Rhizobia are soil bacteria that induce nodule formation on plant roots and differentiate into nitrogen-fixing bacteroids. A detailed understanding of this complex symbiosis is essential for advancing ongoing efforts to engineer novel symbioses with cereal crops for sustainable agriculture. Here, we reconstruct and validate a genome-scale metabolic model for *Rhizobium leguminosarum* bv. *viciae* 3841. By integrating the model with various experimental datasets specific to different stages of symbiosis formation, we elucidate the metabolic characteristics of rhizosphere bacteria, undifferentiated bacteria inside root nodules, and nitrogen-fixing bacteroids. Our model predicts metabolic flux patterns for these three distinct lifestyles, thus providing a framework for the interpretation of genome-scale experimental datasets and identifying targets for future experimental studies.

## Introduction

Nitrogen is commonly the main limiting nutrient in agriculture because plants are unable to assimilate atmospheric $N_2$ (1). Some legumes, such as peas, beans and lentils, circumvent this problem by entering into complex symbiotic relationships with soil bacteria called rhizobia. Legumes secrete signaling molecules (flavonoids) that are recognized by compatible rhizobia, which produce their own signaling molecules (Nod factors) in response. As a result of this signal exchange, rhizobia are typically entrapped by root hairs and grow down so-called infection threads until they are endocytosed by plant cells in the developing nodule. The bacteria then undergo further cell division and eventually differentiate into bacteroids converting atmospheric $N_2$ into ammonia, which is secreted to the plant host in exchange for carbon sources, mainly dicarboxylates (2–4).

Symbiosis formation is a multi-stage process, requiring distinct metabolic capabilities at each stage. The ability of rhizobia to adapt to various environmental conditions is reflected in their large genomes, which often comprise several replicons (5–7), and in the importance of different genomic regions for each lifestyle (8, 9). While significant research efforts have focused on understanding bacteroid metabolism in rhizobium-legume symbioses, several recent studies have begun to unravel the plant-bacteria interactions preceding the formation of differentiated nitrogen-fixing bacteroids. For example, transcriptomic changes in response to root exudates of different plants have been investigated (10, 11) and biosensors have been developed to elucidate nutrient availability in the rhizosphere (12). Importantly, a study using transposon-based insertion sequencing (INSeq) assessed gene essentiality in *Rhizobium leguminosarum* for rhizosphere bacteria, root-attached

bacteria, undifferentiated nodule bacteria and nitrogen-fixing bacteroids (13). It was found that 603 genetic regions were essential for a successful transition from free-living bacteria to bacteroids, highlighting the complexity of development during formation of a successful symbiosis. Understanding the metabolic features at different stages of symbiosis is required for developing effective rhizobial inocula for agricultural applications. Rhizobia that efficiently fix nitrogen are not necessarily adapted to persistence in the rhizosphere as wells as nodulating a plant host in the presence of genetically different bacterial strains, a characteristic described as competitiveness (13–15). Knowledge of the nutrient exchanges between plants and rhizosphere bacteria is thus required for the design of microbial inocula that are competitive and stably persist when applied in the field (14, 16). Once rhizobia have successfully entered the plant root, elucidating the metabolism of undifferentiated rhizobia inside the nodule is also important to avoid delays in the onset of nitrogen fixation.

Due to the complexity of nutrient exchanges in symbioses, metabolic modelling has become a popular tool for investigating rhizobium-legume interactions (3, 17). Metabolic models describe the reactions that are catalyzed by the enzymes annotated in an organism's genome (18, 19). By defining nutrient availability as well as an objective function reflecting the metabolic strategy of the organism, flux distributions at steady state can be calculated using flux balance analysis (20). Due to the gene-protein-reaction associations contained in metabolic models, they also provide a convenient framework for contextualizing genome-scale data obtained by omics technologies, such as transcriptomics or proteomics (21). Most metabolic models of rhizobial species so far have focused on fully differentiated bacteroids

(22–25). One *in silico* study of *Sinorhizobium meliloti* has addressed the differences in metabolism for free-living growth in the bulk soil, growth the rhizosphere, and symbiotic nitrogen fixation during the bacteroid stage (9). However, this study focused on the contributions of the different replicons to fitness in the different environments rather than specifics of changes in metabolic flux distributions and did not integrate experimental data. Another study compared the metabolism of free-living *Bradyrhizobium japonicum* with bacteroids (26). While transcriptome and proteome datasets were used to generate models for rhizosphere bacteria and bacteroids, the data for the rhizosphere model were obtained for bacteria grown in a laboratory culture. Only one recent study has addressed metabolic differences in the different nodule zones for the symbiosis between *S. meliloti* and *Medicago truncatula* (27).

In this study, we reconstruct and extensively curate a genome-scale metabolic model (GSM) for *R. leguminosarum* bv. *viciae* 3841 (Rlv3841). Various experimental datasets exist for this strain at different stages of symbiosis with its native host pea. By integrating transcriptome, proteome and gene essentiality data with the GSM, we perform a detailed investigation of nutrient uptake and metabolic pathway usage of Rlv3841 in the rhizosphere, as nodule bacteria and as nitrogen-fixing bacteroids. This genome-scale approach for data integration reproduced experimentally observed phenotypes and particularly highlighted the role of different carbon sources and amino acids throughout the different stages of symbiosis. The metabolic model developed herein provides a valuable resource for targeted investigation of metabolic requirements of the different rhizobial lifestyles and may enable the

identification of strategies for engineering strains that are metabolically advantaged at all stages of symbiosis formation.

## Results

### Reconstruction of a genome-scale metabolic model for *Rhizobium leguminosarum*

Most published metabolic models for rhizobia focus on bacteroids and are therefore limited to metabolic pathways active during nitrogen fixation. Curated genome-scale reconstructions are so far only available for *B. japonicum* (26) and *S. meliloti* (27). With the aim of investigating metabolism in the rhizosphere and during different stages of bacteroid development, we developed a GSM for Rlv3841 using multiple sources of information. As shown in Fig. 1A, automated reconstructions based on the KEGG (28) and MetaCyc (29) databases were combined with a homology-based reconstruction using a GSM for *S. meliloti* as a template and reactions from our previously reconstructed bacteroid model of Rlv3841 (23) (Fig. 1B). Extensive curation was then performed based on literature evidence, gene essentiality data (13, 30) and enzymatic functions predicted by DeepEC (31). Comparison with iML1515, a high-quality model for *Escherichia coli* (32) as well as the CarveMe template for Gram-negative bacteria (33) was further used to correct reaction stoichiometry and reversibility if required. We next defined a biomass function based on evidence from the literature (Table S1). Since our previous work showed the dependence of carbon polymer synthesis on environmental conditions (23), demand reactions for polymers such as glycogen, polyhydroxybutyrate (PHB) and exopolysaccharides were included in the model to allow for their flexible

accumulation. The final model (Supplementary Data 1 and 2) contained 1224 genes, 1257 reactions and 984 metabolites (Table 1), and was named iCS1224 according to standard naming conventions. The largest groups of metabolic reactions were associated with amino acid and lipid metabolism (14.1% and 13.5% of model reactions, respectively), followed by cofactor metabolism (10.7%) and purine/pyrimidine metabolism (9.0%) (Fig. 1C). Cluster of orthologous genes (COG) (34) analysis of the model genes showed that all COG categories associated with metabolic reactions were represented in iCS1224 (Fig. S1). The quality of the reconstruction was evaluated using MEMOTE (35), where iCS1224 achieved an overall score of 89% (Supplementary Data 3).

**Model validation**

We validated our model for free-living Rlv3841 growing in minimal media using various experimental datasets. First, we experimentally assessed growth on 190 different carbon sources using phenotype microarrays (36) (Supplementary Data 4). For the 109 carbon sources that were present as metabolites in iCS1224, an overall predictive accuracy of 89.9% with 90.9% precision and 96.4% recall was achieved (Fig. 2A), which is similar to the performance of curated GSMs for well-investigated bacteria, such as *Pseudomonas aeruginosa* (37) or *E. coli* (38). In addition, we evaluated the quality of gene essentiality predictions by comparing *in silico* gene essentiality with the results of an INSeq gene essentiality screen of Rlv3841 performed in minimal media supplemented with succinate and ammonia (30). Since the classification of genes based on transposon mutagenesis screen is subject to some variability (39), the list of essential genes was further curated by comparison

with INSeq data for growth on complex media (13). Predictions by iCS1224 achieved an accuracy, precision and recall of 91.0%, 89.6% and 87.8%, respectively (Fig. 2B), thus showing good agreement with the INSeq data and indicating high quality of the gene-protein-reaction associations as well as suitability of the biomass objective function.

Finally, we performed quantitative validation of our model by comparing the predicted flux values for 17 reactions involved in central carbon metabolism with published values measured by $^{13}C$ metabolic flux analysis of Rlv3841 grown in minimal media with succinate and ammonia (40). As shown in Fig. 2C, we observed excellent agreement between predicted and experimentally measured flux values. In all cases, the measured flux was within the range determined by flux variability analysis. iCS1224 thus appears to be an accurate representation of the metabolism of Rlv3841, both qualitatively and quantitatively.

**Metabolism of rhizosphere bacteria**

Having validated the predictive capabilities of iCS1224, we sought to extract condition-specific models for metabolism of Rlv3841 (i) in the rhizosphere, (ii) as undifferentiated nodule bacteria and (iii) as nitrogen-fixing bacteroids (Fig. 3). We chose the recently developed RIPTiDe algorithm (41) to obtain condition-specific metabolic models. Based on gene expression data, RIPTiDe assigns weights to all gene-associated reactions, assuming that higher transcript abundance makes it more likely that the corresponding reaction is used in a certain environmental condition. The overall flux through the network is then minimized and inactive reactions are removed. Finally, flux sampling of the solution space is performed,

where flux through reactions associated with highly expressed genes is favored. In contrast to other methods for transcriptome data integration, RIPTiDe does not impose arbitrary thresholds on the gene expression data, it produces functional models with flux through the objective reaction, and takes flux parsimony into account, i.e., the overall flux is minimized to find cost-efficient solutions (41). Generation of a rhizosphere-specific model thus required information about available nutrients as well as gene expression data. Nutrient availability in the rhizosphere is mainly determined by plant root exudates, and plants modulate the composition of their root exudates to select for specific soil microbes (42, 43). However, only a subset of metabolites is used by the soil microbiota (44, 45), and elucidation of nutrient uptake by rhizosphere bacteria usually requires extensive metabolomics profiling (46, 47). Taking a top-down approach for defining the rhizosphere environment, we first compiled a list of compounds present in pea root exudates based on published experimental data (10, 12, 48, 49) (Table S2 and S3, Supplementary Text). For those compounds that could be matched to model metabolites, exchange reactions were added to the model with reaction bounds set to allow for unlimited uptake. RNA-Seq data for Rlv3841 in the rhizosphere of pea plants 7 days post inoculation (Supplementary Data 5) was used as an input dataset for model contextualization. In addition, a list of genes that were classified as essential or defective in the rhizosphere in an INSeq screen (13) was provided to prevent removal of reactions associated with these genes from the rhizosphere-specific model. Biomass production was set as the objective and additional positive lower bounds were placed on exopolysaccharide, lipopolysaccharide and Nod factor synthesis, all of which are known to be important in the rhizosphere (4). During data

integration, constraining flux through the objective function to values between 50% and 95% of its maximum was tested to identify the scenario that gave the best match with the transcriptome data. Within the range of objective values tested, the highest correlation (Spearman's Rho=0.237, $P<0.001$) between metabolic fluxes and transcript abundances was obtained with the biomass reaction constrained to carry at least 77.5% of its maximum flux, and the rhizosphere-specific model contained 606 reactions and 576 metabolites (Supplementary Data 6). Remarkably, out of the 134 nutrients available for uptake before data integration, only 51 were present in the rhizosphere-specific model.

For the analysis of the contextualized model, we focused on those metabolic pathways that are either not universally essential or that are retained in the model despite their end product being available for uptake from the environment. Pathways such as membrane lipid or PHB synthesis, for instance, will always be retained in the model, since they are required to maintain flux through the biomass objective function and uptake of lipids and PHB is not possible. In addition, we limit our discussion to reactions that had a non-zero median flux value based on the flux sampling results, since those reactions are most likely to be active in the rhizosphere. The TCA cycle was predicted to be a central catabolic pathway (Fig. 4), which is consistent with previous reports of organic acids being the predominant carbon sources for rhizobia in the rhizosphere (9, 10). In particular, the model predicted high uptake of glycolate in agreement with the induction of C2 metabolism observed in previous gene expression studies (10). Glycolate was converted into pyruvate via glycolate oxidase and an aminotransferase. The model also showed high uptake rates for aspartate, which could explain the induction of a *dctA* biosensor

234 in Rlv3841 in the pea rhizosphere (12). Aspartate and 2-oxoglutarate were
235 transaminated to produce glutamate and oxaloacetate, which is a TCA cycle
236 intermediate.
237 In addition to organic acids, amylotriose, which is hydrolyzed into glucose, was partly
238 metabolized via the Entner-Doudoroff pathway and glycolysis in the model and
239 entered the pentose phosphate pathway to enable production of nucleotides required
240 for the synthesis of various polysaccharides and Nod factors. The solute-binding
241 protein of a carbohydrate uptake transporter-1 (CUT1) family transporter (RL3840)
242 was 2.6-fold upregulated in the pea rhizosphere compared to free-living cells (10),
243 which supports the predicted uptake of a di- or oligosaccharide. Ribulose, a
244 monosaccharide metabolized via the pentose phosphate pathway was also predicted
245 to be taken up. Catabolism of a monosaccharide in the rhizosphere is highly
246 probable considering the strong signals of a fructose and a xylose biosensor in the
247 pea rhizosphere (12). The fructose biosensor is based on the solute-binding protein
248 of the CUT2 family *frcABC* transporter, which has been shown to transport ribose in
249 addition to fructose in *S. meliloti* (50) and may therefore also contribute to pentose
250 uptake in the rhizosphere. The model further contained reactions for glycerol uptake
251 and catabolism, which could explain the decreased competitiveness observed for a
252 glycerol catabolism mutant of *R. leguminosarum* bv. *viciae* VF39 (51).
253 With regard to amino acids, all of which are present in root exudates, biosynthetic
254 pathways were generally retained in the rhizosphere model due to the essentiality of
255 the associated genes. Low levels of uptake were however predicted for most amino
256 acids, mainly to support protein synthesis. Notably, the biosynthetic pathway for
257 leucine was predicted to be active, which was partly supported by uptake of 2-

isopropylmalate, an intermediate of branched-chain amino acid synthesis. The need for leucine synthesis in the rhizosphere agrees with a *leuD* mutant of Rlv3841 requiring the addition of 1 mM leucine to nodulate pea (52). Mutation of the isopropylmalate synthase gene in *S. meliloti* impaired nodulation even in the presence of leucine, and it was shown that either isopropylmalate synthase or intermediates of the leucine biosynthetic pathway are required for the activation of *nod* gene expression (53). It is therefore possible that the predicted leucine synthesis is at least partly related to the synthesis of Nod factors in the rhizosphere. High uptake rates were further predicted for glutamine, which is consistent with its high concentration in pea root exudates (49). Glutamine was converted into glutamate, which was mostly used to sustain leucine synthesis. The model also contained active uptake reactions for several nucleotides. This agrees with the reported uptake of nucleosides and nucleotides by rhizosphere bacteria (44, 47, 54) and agrees with the gene essentiality predictions for rhizosphere bacteria obtained by INSeq, where purine auxotrophs appear to be rescued by plant root exudates (13).

Among biomass components that are present in root exudate but not predicted to be taken up, biosynthesis of the polyamine putrescine was retained in the model, attesting to the ability of the RIPTiDe algorithm to choose metabolic reactions that agree with gene expression and/or essentiality rather than choosing the least resource-intensive solution. Putrescine and related polyamines are important for survival under stress conditions and their synthesis has been suggested to play an important role during root colonization (55). As part of the model reconstruction, several demand reactions were included for compounds such as carbon polymers whose accumulation can vary with environmental conditions. The only non-essential

demand reactions that were not removed during the pruning process were those for glutathione and polyphosphate, where polyphosphate synthesis in particular had a non-zero median flux. Glutathione is important to deal with stress conditions, such as osmotic and oxidative stress, encountered in the rhizosphere, and mutants in glutathione biosynthesis are severely affected in rhizosphere colonization (56). The predicted catabolism of glycolate via glyoxylate produces the reactive oxygen species hydrogen peroxide, which could contribute to the need for glutathione synthesis. Polyphosphate has recently been suggested to play a role in the global carbon regulatory system (57), but its function remains to be investigated in detail. It is interesting to note that an exopolyphosphatase gene (RL1600) was classified as essential for persistence in the rhizosphere (13), indicating an important role for phosphate homeostasis in rhizosphere colonization and/or competition.

Catabolism of several other compounds, such as erythritol, *myo*-inositol and homoserine, has been described to be important for competitiveness (10, 58, 59), however, these compounds were not included in the rhizosphere model. This could be due to the catabolism of these compounds being important at later stages of the symbiosis, e.g. for growth in infection threads rather than in the rhizosphere. Alternatively, uptake of these compounds could be masked in the model due to catabolic routes that are shared with other metabolites. For example, erythritol is metabolized via the pentose phosphate pathway (60), hence the predicted uptake and metabolism of ribulose could partly be due to erythritol catabolism.

**Reporter metabolites highlight plant-specific rhizosphere metabolism**

As an independent validation and extension of our analysis of metabolic changes, we identified reporter metabolites using previously published microarray data comparing Rlv3841 in the rhizosphere of pea plants with free-living cells grown on minimal media with glucose and ammonium chloride (10). Based on the network topology defined by a metabolic model, the reporter metabolite algorithm identifies those compounds around which significant changes in gene expression occur (61). This method is therefore independent of specifying nutrient uptake from the environment. Reporter metabolites associated with upregulated genes matched several observations from the RNA-Seq data integration described in the previous section. In particular, several intermediates of branched-chain amino acid synthesis, such as acetolactate, 2-hydroxyethyl-thiamin diphosphate and 2-aceto-2-hydroxybutanoate, were identified as reporter metabolites (Fig. 5A). Significant transcriptional changes were also observed around various nucleobase derivatives. This may be related to their predicted uptake from plant root exudates but could also indicate an increased need for nucleotide synthesis for the production of polysaccharides and Nod factor. Phosphoribosyl-AMP and phosphoribosylformiminoaicar-phosphate are intermediates of histidine biosynthesis and direct precursors of AICAR, which is involved in purine metabolism. Seeing as no additional metabolites of the histidine biosynthetic pathway were identified as reporter metabolites, this analysis indicates an increase in AICAR synthesis, which seems to be required for successful legume infection by various *Rhizobium* species (62).

Comparison of the rhizosphere reporter metabolites for pea (host legume for Rlv3841) with those for alfalfa (non-host legume) (Fig. 5B) and sugar beet (non-legume) (Fig. 5C) highlighted several plant-specific features. For alfalfa, phosphoribosyl-AMP was identified as a reporter metabolite similar to pea. In addition, phenylalanine and tyrosine support the role of aromatic amino acid metabolism in colonization competitiveness (10). Significant transcriptional changes also occurred around the carbon polymer beta-glucan and the diamine putrescine. While beta-glucan generally appears to be important for persistence in the rhizosphere (10, 13), its identification as a reporter metabolite together with putrescine indicates increased osmotic stress in the alfalfa rhizosphere compared to pea. For sugar beet, the identification of several compounds involved in nitrogen metabolism (ammonia, urea, urate) agrees with the suggested nitrogen limitation in the sugar beet rhizosphere, but nitrogen sufficiency in legume rhizospheres (10). This could also explain why the carbon polymer glycogen was a reporter metabolite specifically in the sugar beet rhizosphere, since glycogen synthesis is probably linked to nitrogen limitation (63). Notably, multiple mono- and disaccharides and their derivatives indicate an increased importance of sugar metabolism compared to legume rhizospheres. However, many genes involved in saccharide metabolism are associated with multiple reactions (e.g. unspecific glucoside hydrolases), and therefore the identity of the metabolized sugar cannot be derived from this analysis. Finally, the reporter metabolites 3-dehydrocarnitine and betainyl-CoA indicate either accumulation of amines for osmoprotection or catabolism of carnitine or related amines. These findings present interesting targets for future investigations using gene essentiality screens on different plant hosts.

Overall, both the context-specific model obtained by transcriptome data integration and the reporter metabolite analysis were in good agreement with experimental data for rhizobial metabolism in the rhizosphere without forcing the uptake of any compound through arbitrary constraints. Instead, insights into nutrient uptake were facilitated by the integration of gene expression and gene essentiality data with iCS1224. If biomass production were simply maximized with unlimited availability of all root exudate compounds, this would result in uptake of all available compounds that are required for biomass formation, which would not reflect a biologically meaningful scenario.

**Metabolism of undifferentiated nodule bacteria**

We next sought to develop models for Rlv3841 inside the nodule environment. For this purpose, it is important to differentiate between nodule bacteria at the tip of the nodule, which are dividing and undergoing differentiation, and bacteroids in the central nitrogen fixation zone of the nodule (64). While nodule bacteria are still dividing, bacteroids are growth-arrested and mainly catabolize plant-provided dicarboxylates to fix atmospheric $N_2$ into ammonia. The distinction between these developmental stages is required in the context of gene essentiality analyses since genes required for the differentiation process may not be essential for nitrogen fixation and vice versa. Similar to the approach for the rhizosphere model, we used RIPTiDe to obtain models for nodule bacteria and bacteroids and performed flux sampling to identify those reactions that are most likely to be active in each contextualized model.

The model for nodule bacteria was obtained using published dRNA-Seq data for RNA extracted from nodule tips (65), as well as a list of genes that were predicted to be specifically essential for nodule bacteria (13). Nutrient availability was defined based on a study using biosensors to detect metabolites inside nodules (12) and our direct measurement of metabolites in pea root exudate, in pea bacteroids and in the nodule cytosol as described previously (40) (Table S4 and S5, Supplementary Data 7). The biomass objective function was used to account for the cell division occurring as rhizobia grow down infection threads and differentiate into bacteroids and positive lower bounds were placed on demand reactions for exopolysaccharides and lipopolysaccharides. The nodule bacteria model contained 510 reactions and 502 metabolites and achieved highest correlation with the transcriptome data (Spearman's Rho=0.335, $P$<0.001) when the objective value was constrained to 65% of its maximum (Supplementary Data 8). The observation that higher correlation was obtained for lower flux through the objective reaction (compared to the rhizosphere) indicates that the metabolism of nodule bacteria is not oriented towards maximum growth. This agrees with experimental data showing that growth of infection threads proceeds at highly variable rates controlled by the plant host (66). The improved correlation of flux predictions and gene expression data compared to the rhizosphere model can be explained by the lower number of essential genes, which places fewer constraints on the reactions included in the contextualized model.

Malate, fructose, xylose, *myo*-inositol and γ-aminobutyrate (GABA) were all predicted to be taken up by nodule bacteria (Figs. 6A and S2). Biosensors for these carbon sources were strongly induced in young nodules, whereas biosensors for the carbon

sources that were removed during the data integration process (erythritol, mannitol, formate, malonate, tartrate) only showed weak induction (12). Malate and GABA are both catabolized in the TCA cycle, indicating that it is an important catabolic route in differentiating nodule bacteria despite transport of dicarboxylates being non-essential for differentiation into bacteroids (67). Enzymes involved in GABA metabolism are highly induced in bacteroids, although GABA catabolism is not essential for effective nitrogen fixation (68). The predicted catabolism of fructose is consistent with the strong induction of a fructose-specific biosensor in nodules (12) as well as a previous modelling study of *S. meliloti* suggesting the use of sucrose-derived sugars as a carbon source by differentiating nodule bacteria (27). Sucrose uptake was removed from the nodule bacteria model, which may be due to the inability of the model to accurately distinguish between sucrose and fructose uptake based on the gene expression data. It is interesting to note that Rlv3841 bacteroids mutated in a subunit of succinyl-CoA synthetase, which had severely reduced nitrogen fixation capacity, had 168-fold higher levels of fructose than wild-type bacteroids and 151-fold elevated levels of sucrose (Supplementary Data 7), which may be the result of carbon source build-up in the developmentally impaired nodule bacteria and bacteroids. *Myo*-inositol is present in the rhizosphere (12) and abundant in pea nodules (69), and mutants in *myo*-inositol catabolism have strongly reduced competitiveness compared to wild-type Rlv3841 (58). However, the activity of enzymes involved in *myo*-inositol catabolism is very low in mature bacteroids (69), and mutants in *myo*-inositol catabolism were not disadvantaged during growth in the rhizosphere compared to wild-type Rlv3841 (58). In addition, it has been proposed that catabolism of rhizopines, which are inositol derivatives, by undifferentiated

nodule bacteria may be important as a kin selection strategy (70). Catabolism of *myo*-inositol is therefore most likely to play a role during infection and in undifferentiated nodule bacteria, which is correctly predicted by the model. Xylose enters the pentose phosphate pathway, and its predicted uptake could be related to the importance of nucleotide synthesis, both for DNA endoreduplication and synthesis of exopolysaccharides and lipopolysaccharides. Similarly, uptake reactions for the nucleoside guanosine and uridine as well as the nucleobase adenine were present in the model.

Our nodule bacteria model predicted uptake of most amino acids, which agrees with the severe symbiotic defect of a *gltB* mutant unable to transport amino acids (71) but may also be a result of a beginning general downregulation of biosynthetic functions as rhizobia transition into growth-arrested bacteroids. Similar to the rhizosphere bacteria, leucine was predicted to be synthesized from 2-isopropylmalate. Expression of *nod* genes is elevated in nodule bacteria at 7 days post inoculation (72), which could explain the predicted leucine synthesis as discussed for the rhizosphere model.

**Metabolism of bacteroids**

To extract a model specific for nitrogen-fixing bacteroids, we used dRNA-Seq data derived from the middle of nodules (65), which contains fully differentiated bacteroids performing nitrogen fixation (64). In addition, a list of 38 genes that were present in the model and encoded proteins significantly upregulated in bacteroids compared to free-living bacteria (23) and the *dct* genes (73) were specified to ensure inclusion of those genes in the bacteroid model. Nitrogenase activity was set as the objective

445 function while low levels of protein and fatty acid production were enforced through
446 demand reactions. Nutrient availability was specified similar to the considerations for
447 nodule bacteria (Table S6). Gene essentiality data from the INSeq screen were not
448 included for model contextualization due to the aforementioned difficulty of
449 determining the developmental stage where a gene is essential inside the nodule
450 environment. The bacteroid model contained 307 reactions and 308 metabolites
451 (Supplementary Data 9) and achieved significant correlation with the transcriptome
452 data (Spearman's Rho=0.348, $P$<0.001) when nitrogenase activity was constrained
453 to 65% of its maximum. The reduced model size compared to both the rhizosphere
454 and the nodule bacteria model is in agreement with the reduced physiological
455 complexity of the non-dividing bacteroids (3, 72).
456 The bacteroid model contained the C4 dicarboxylates malate, succinate and
457 fumarate as the main carbon sources in agreement with experimental evidence (67,
458 74) (Figs. 6B and S3), and only low levels of GABA uptake were predicted. Ammonia
459 was the only nitrogenous export product. Consistent with our previous modelling
460 study of Rlv3841 bacteroids (23), constraining the oxygen uptake prior to data
461 integration resulted in nitrogen partly being secreted as alanine. With the metabolites
462 provided in initial simulations, the glyoxylate cycle comprising isocitrate lyase and
463 malate synthase was contained in the model, which is consistent with the high
464 induction of malate synthase (72) but disagrees with the lack of isocitrate lyase
465 activity in pea bacteroids (75). The source of glyoxylate for the malate synthase
466 reaction has so far not been elucidated. Because the metabolomics data showed
467 that glycolate is present in the nodule cytosol and glycolate concentrations in
468 bacteroids are 2-fold elevated compared to free-living cells (Supplementary Data 7),

469 we allowed for glycolate uptake by the bacteroid model and inactivated the isocitrate
470 lyase reaction. This resulted in substantial uptake of glycolate, which was converted
471 into glyoxylate that was used in the malate synthase reaction. Glycolate provision by
472 the plant may therefore explain the increase in malate synthase expression in the
473 absence of isocitrate lyase activity.
474 The model also contained uptake of xylose, which was metabolized in the pentose
475 phosphate pathway and supported synthesis of nucleotides. Dicarboxylate
476 catabolism generally requires gluconeogenesis to provide precursors for the
477 synthesis of nucleotides and some amino acids. Due to the predicted xylose uptake,
478 only minor flux through the reactions involved in gluconeogenesis was predicted,
479 highlighting the importance of this pathway in bacteroids as an interesting question
480 to explore using targeted mutant studies. Proton uptake by bacteroids was further
481 required as previously predicted for *S. meliloti* bacteroids (27) and a demand
482 reaction for PHB was retained in the model. PHB synthesis was highly variable
483 across flux samples, which is in agreement with its previously suggested role for
484 carbon and redox balancing (23, 40).
485 Low levels of uptake were predicted for most amino acids to support the required
486 synthesis of protein, but no significant catabolism of any amino acid was observed.
487 Mutant studies have shown a requirement for branched-chain amino acid supply to
488 bacteroids (52), and isoleucine was predicted to be supplied by the plant.
489 Interestingly, histidine was predicted to be synthesized rather than taken up by
490 bacteroids. Several proteins involved in histidine synthesis were upregulated or
491 unchanged in abundance in the bacteroid proteome compared to free-living Rlv3841
492 (23), in contrast to the general downregulation of amino acid biosynthesis (3). Similar

493 results were obtained in a proteome study of *Rhizobium etli* (24) and RNA-Seq data
494 for bacteroids of *R. leguminosarum* bv. *viciae* A34 and *R. leguminosarum* bv.
495 *phaseoli* 4292 (76). In addition, mutants of *R. leguminosarum* bv. *trifolii* lacking
496 histidinol dehydrogenase activity formed ineffective nodules on clover (77). To
497 investigate the requirement for histidine biosynthesis, we compared the amino acid
498 composition of the Nif and Fix proteins, which are highly expressed in bacteroids,
499 with the overall amino acid composition of the Rlv3841 proteome (Table S7). We
500 found a significant ($P$=0.042) enrichment of histidine in the Nif and Fix proteins,
501 which could at least partly explain why histidine biosynthesis is required in
502 bacteroids.
503
504 **Discussion**
505 In this study, we present the first curated GSM for Rlv3841, a model strain for
506 investigating rhizobium-legume interactions and a natural symbiont of the
507 agriculturally important crop pea. GSMs have emerged as promising tools for
508 informing experimental design, addressing fundamental research questions, and
509 contextualizing experimental data (78). In order to obtain a high-quality model,
510 integration of experimental data during model curation and validation is essential
511 (79). We therefore evaluated our model using carbon source utilization, gene
512 essentiality data and flux data obtained by $^{13}C$ labelling and observed high
513 agreement between model predictions and experimental data.
514 We further used the GSM to elucidate metabolic changes in Rlv3841 as it transitions
515 from a free-living soil bacterium in the rhizosphere to an undifferentiated nodule
516 bacterium and finally to a nitrogen-fixing bacteroid. While significant advances in

517 determining metabolic requirements for successful symbiosis formation have been
518 made using transcriptome data (10) and gene essentiality screens (13), genome-
519 scale datasets are often difficult to interpret without the framework of a
520 comprehensive model, especially when information about nutrient uptake is missing.
521 To this end, we employed approaches integrating gene expression and metabolome
522 data as well as gene essentiality predicted by INSeq to obtain condition-specific
523 models. This allowed us to contextualize our model based on experimental data
524 without assuming uptake rates for any nutrient. In addition, during the process of
525 data integration, different fractions of the optimum objective value were tested as
526 constraints to find a solution with the highest correlation between gene expression
527 and associated reaction fluxes. Especially for nodule bacteria and bacteroids, using
528 sub-optimal fluxes through the objective function as constraints during model
529 contextualization was found to produce better agreement with experimental data.
530 Objective functions can be difficult to define outside of defined growth in a laboratory
531 culture, and our results highlight the need to adopt strategies beyond maximization
532 of a biomass objective function to accurately capture metabolic behavior in complex
533 environmental settings. A clear limitation of our approach is the imperfect correlation
534 of gene expression and protein abundance, as well as protein abundance and
535 enzyme activity (80, 81). Catabolic pathways common to multiple different
536 compounds can further make it difficult to specifically determine which nutrient is
537 taken up. Nevertheless, our model predictions are in good agreement with known
538 metabolic characteristics of the different lifestyles of Rlv3841, attesting to the
539 biological relevance of our findings.

The rhizosphere model showed substantial uptake of glycolate, aspartate and glutamine as well as mono- and oligosaccharides. These predictions are consistent with previous studies about gene expression (10) as well as nutrient uptake of a *Rhizobium* sp. from root exudates of *Arabidopsis* (44). We further identified a requirement for leucine synthesis in the rhizosphere, as well as a potentially important role for polyphosphate synthesis. However, the predicted nutrient uptake was not supported by gene essentiality predictions in all cases. While both INSeq gene essentiality assignments and our metabolic model generate predictions that warrant detailed investigation using isolated mutant strains, there are other possible explanations for this observation. First, root exudates might not contain sufficient quantities of a compound to complement an auxotrophy. In addition, the composition of plant root exudates has been shown to change over time (47), and compounds present at the time of RNA extraction may not be present from the time of inoculation, causing the loss of some mutants. Finally, for genes that are essential on complex media, the corresponding mutants may already be lost from or underrepresented in the bacterial population that is inoculated onto plants.

The model for nodule bacteria confirmed previous results suggesting supply of nutrients other than dicarboxylates, in particular sucrose-derived sugars, during the differentiation process (27). Interestingly, we found that *myo*-inositol catabolism was only predicted for nodule bacteria, but not in the rhizosphere or in bacteroids. While the importance of *myo*-inositol catabolism for competitiveness has been established (58), our results suggest that it may be particularly important for differentiating bacteria rather than those in the rhizosphere. For bacteroids, biosynthesis of histidine was found to be important in contrast to the general uptake predicted for

most other amino acids. In addition, low levels of xylose uptake were predicted to support nucleotide synthesis in bacteroids. This result indicates that a carbon source metabolized in the pentose phosphate pathway may be provided to bacteroids, which presents an interesting area to explore experimentally using mutants affected in gluconeogenesis. Initial predictions of isocitrate lyase activity, which disagree with measured enzyme activities in bacteroids, led us to hypothesize that glycolate is provided to bacteroids. This is supported by metabolomics data and could explain the induction of malate synthase in bacteroids without concomitant expression of isocitrate lyase.

In summary, our results provide insights into rhizobial metabolism in the rhizosphere, which can inform the design of more competitive rhizobial inocula as well as plants that secrete metabolites to specifically enrich beneficial bacterial strains. Our understanding of the nutrient exchanges between plants and rhizobia at different developmental stages inside nodules remains incomplete (3, 82), and the predictions presented herein provide a foundation for targeted investigation of amino acid and central carbon metabolism in particular. We anticipate that the highly curated metabolic model for Rlv3841 presented in this paper will provide a valuable resource for the reconstruction of GSMs for related species.

## Materials and Methods

### Model reconstruction

To reconstruct a GSM for Rlv3841, we combined information from multiple databases, which has been shown to significantly improve the scope of metabolic network reconstructions (83). All reconstructions were performed based on RefSeq assembly GCF_000009265.1. We used the RAVEN Toolbox 2.0 (84) to create draft models from KEGG (28) and MetaCyc (29) using the functions `getKEGGModelForOrganism` and `getMetaCycModelForOrganism`, respectively. In addition, template-based reconstruction based on BLAST bidirectional hits was performed using a curated GSM for *S. meliloti* 1021 (iGD1348 (27)) as a template for the function `getModelFromHomology`. All models were merged into one reaction list and reaction and compound identifiers were unified based on the reaction database provided by the ModelSEED (85), followed by removal of duplicate reactions. Starting from this database of reactions compiled from different sources, the reconstruction was curated. First, reactions without gene association were removed. Reactions involving non-specific compounds such as “acceptor” or “protein” were also deleted, as well as reactions involved in the biosynthesis and catabolism of secondary metabolites and non-metabolic processes, such as DNA and RNA modification because those were outside of the scope of our model. Extensive curation was then performed by evaluating metabolic pathways guided by the literature and the KEGG database. Pathways for catabolism of small carbon sources in particular were reconstructed based on predictions obtained from GapMind (86). Gene-protein-reaction associations were curated based on published

gene essentiality data for growth in minimal (30) and complete (13) media as well as enzyme commission (EC) number predictions obtained from DeepEC (31). Transport reactions were annotated based on literature evidence, in particular homology to experimentally characterized transporters in *S. meliloti* (87) and the annotation obtained from TransportDB 2.0 (88). We manually reconstructed pathways for organism-specific biomass components, such as lipopolysaccharides and exopolysaccharides, as well as pathways which were not present in any of the databases used for reconstruction, such as carnitine metabolism. To improve information on reaction directionality, upper and lower bounds were adjusted according to the information in a highly curated model for *E. coli* (iML1515 (32)) and the CarveMe template model for Gram-negative bacteria (33).

The biomass objective function was defined as follows: The composition of DNA was determined from the RefSeq genome sequence. Similarly, RNA and protein composition were determined by counting nucleotides or amino acids in the annotated RNAs and protein coding sequences, respectively. Since the lipid composition of Rlv3841 has not been investigated so far, we adopted the values reported *R. leguminosarum* bv. *trifolii* ANU843 (89). *R. leguminosarum* produces predominantly C18 fatty acids, as well as smaller quantities of C16 fatty acids (40, 90), and representative phospholipids in our model included fatty acids with these chain lengths. Lipopolysaccharides and exopolysaccharides were included with the fractions previously reported for *S. meliloti* (91). Cyclic beta-glucans have so far not been considered in metabolic models for rhizobia, however they can make up a significant fraction of the cellular dry weight (92) and were therefore also included as a biomass component. Apart from the main cell components, trace amounts of

cofactors identified as universally essential in prokaryotes (93) were included in the biomass objective function. Phytoene was also added to the biomass reaction due to the essentiality of the genes associated with its biosynthetic pathway. Carbon polymers such as glycogen, PHB and fatty acids, as well as polyamines such as homospermidine and putrescine are known to be produced by Rlv3841, however, the quantities in which they are produced vary depending on nutrient availability. Similar to our previous work (23), we therefore added demand reactions for these compounds to allow for variable accumulation. Glycogen and PHB were also included in the biomass objective function as they are commonly synthesized by free-living *R. leguminosarum* (94). A complete description of the biomass composition used in this study is given in Table S1.

**General modelling procedures**

Standard metabolic modelling computations were performed in MATLAB R2020b (Mathworks) using scripts from the COBRA Toolbox v3.0 (95) and the Gurobi 9.1.1 solver (www.gurobi.com). When using the `optimizeCbModel` function, the Taxicab norm was minimized to avoid loops in the calculated flux distributions. All scripts are available on Github (https://github.com/CarolinSchulte/Rlv3481-lifestyles).

**Model validation**

To evaluate the agreement between model predictions and experimentally determined carbon source utilization, we limited our analysis to those compounds that were either present in the model or showed a positive growth phenotype in the phenotype microarray experiment. The lower bounds for the exchange reactions

were then adjusted according to the composition of universal minimal salts (UMS) media (30) with ammonium as a nitrogen source, and flux through the biomass reaction was evaluated for each carbon source individually added to the model. Accuracy, precision and recall for carbon source utilization and gene essentiality analysis were calculated according to the following equations:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{FP + FN}$$

TP: true positives FP: false positives

TN: true negatives FN: false negatives

Gene essentiality analysis was performed using the function `singleGeneDeletion` with the MOMA (minimization of metabolic adjustment) option in the COBRA Toolbox, while all components of UMS media with succinate and ammonia were available without constraints on their uptake rate. The predictions were compared with gene essentiality data for Rlv3841 determined by INSeq (13, 30). Genes that were experimentally classified as essential or defective were considered essential in our analysis. The threshold for a gene to be considered essential *in silico* was set to 50% of the wild-type growth rate since all mutant strains are grown in a single culture for an INSeq experiment, and a slower growth rate will

675 therefore decrease the abundance of a mutant even if the gene carrying the insertion
676 is not absolutely essential.
677 For comparison with $^{13}C$ labelling data, boundary conditions were set to allow for
678 unlimited uptake of UMS media components. The succinate uptake rate was
679 constrained to 1 flux unit and flux balance analysis was performed maximizing the
680 biomass objective function. In addition, loopless flux variability analysis was
681 performed where the objective fraction was set to 95% of the optimum value.
682
683 **Data integration for model contextualization**
684 The Python implementation of RIPTiDe (https://github.com/mjenior/riptide) (41) was
685 used to generate condition-specific models of iCS1224. Max fit RIPTiDe was run for
686 objective flux fractions between 0.5 and 0.95 with 0.05 increments, and the context-
687 specific models with the highest correlation between flux values and transcriptome
688 data were used in further analyses.
689 In addition to the nutrient availability determined based on experimental data, trace
690 elements and vitamins required for flux through the objective function were added to
691 the *in silico* representation of each environment. If an exchange and transport
692 reaction for a compound already existed in the model, the lower bound of the
693 exchange reaction was set to -1000. If a compound was only present as an
694 intracellular metabolite, a sink reaction for this metabolite with lower bound set to -
695 1000 and upper bound set to 0 was added. This was done to avoid erroneous
696 exclusion of metabolites which are present in the environment, but for which
697 transporters have not been identified. Some cofactors and central intermediates of
698 carbon metabolism, such as glyceraldehyde 3-phsophate, were excluded from

699 environmental representations since their uptake would result in unspecific
700 predictions for metabolic pathway activity (Supplementary Text).
701
702 **Data integration for rhizosphere model**
703 For the rhizosphere model, compounds that have been detected in pea root
704 exudates (10, 12, 48, 49) and that could be matched to model metabolites were
705 specified with unlimited availability (Table S2 and S3). Flux through the biomass
706 reaction as described in the previous section was set as the objective function and in
707 addition, a lower bound of one flux unit was set for demand reactions for Nod factor,
708 lipopolysaccharides and exopolysaccharides, since these compounds are known to
709 be produced as part of the root colonization process (4). RPKM values for RNA-Seq
710 data obtained from Rlv3841 in the pea rhizosphere 7 days post inoculation were
711 provided as an input, and all genes that are present in the model and were classified
712 as essential or defective in the rhizosphere (13) were specified as model tasks to
713 prevent removal of the associated reactions during the pruning process.
714
715 **Data integration for nodule bacteria model**
716 For nodule bacteria, all metabolites that were detected by rhizobial biosensors in pea
717 nodules (12) were allowed to be taken up without limitation, as well as all amino
718 acids and metabolites whose abundance was at least ten-fold higher in the nodule
719 cytosol compared to root exudates (Table S4 and S5). A lower bound of one flux unit
720 was set for lipopolysaccharide and exopolysaccharide demand reactions and
721 biomass production was used as the objective function. RPKM values for dRNA-Seq
722 data obtained from the tip of pea nodules were provided as an input, and all genes

723 that were classified as essential or defective for nodule bacteria (13) were specified
724 as model tasks to prevent removal of the associated reactions during the pruning
725 process.
726
727 **Data integration for bacteroids**
728 Similar to the nodule bacteria model, metabolites detected in nodules by rhizobial
729 biosensors, and all amino acids were made available to the bacteroid model (Table
730 S6). However, fructose and sucrose were not included since they are known to be
731 poorly oxidized by bacteroids (96). Inclusion of the metabolites increased in the
732 nodule cytosol compared to root exudates led to a decrease in correlation of flux
733 predictions and gene expression data, and those metabolites were therefore omitted
734 from the nutrients available to bacteroids. A lower bound of one flux unit was set for
735 the synthesis of fatty acids and proteins and flux through the nitrogenase reaction
736 was used as the objective function. RPKM values for dRNA-Seq data obtained from
737 the middle of pea nodules were provided as an input, and all genes associated with
738 proteins significantly upregulated in bacteroids compared to free-living cells (23)
739 were specified as model tasks to prevent removal of the associated reactions during
740 the pruning process.
741
742 **Phenotype MicroArray™ analysis**
743 Carbon source utilization of Rlv3841 was assessed using the phenotype microarray
744 technology (Biolog, Hayward, USA). A liquid culture of Rlv3841 was grown at 28 ºC
745 in UMS media supplemented with 10 mM glucose, 10 mM ammonium chloride and a
746 vitamin solution as previously described (30). Cells were spun down and washed

three times in UMS without addition of a carbon or nitrogen source. The optical density at 600 nM was then adjusted to 0.1 with UMS supplemented with 10 mM ammonium chloride and vitamins, and 100 ml of cell suspension were added to each well of the phenotype microarray plate. After overnight incubation without shaking at 28 ºC, 10 ml of a 0.1 % (w/v) stock solution of 2,3,5-triphenyltetrazolium chloride in water were added to each well. Plates were then incubated in an Omega FluoStar plate reader with double orbital shaking at 500 rpm and the absorbance at 505 nm was measured every 15 min. Absorbance values were analyzed using the DuctApe software (97), and all carbon sources with an activity value higher than the water control were considered to support growth. For activity values close to the growth threshold, curves were manually inspected, and literature searches were performed to determine if the carbon source supports growth of *R. leguminosarum*. The full DuctApe output for the phenotype microarray analysis is available in Supplementary Data 4.

**Metabolomics data**

Metabolomics data were obtained in a previous study (40), where only values for metabolites relevant to the investigated metabolic pathways were published . The full metabolomics dataset is included as Supplementary Data 7.

**Sample preparation for RNA-Seq of rhizosphere bacteria**

For total RNA extraction from Rlv3841 in the pea rhizosphere, *Pisum sativum* cv. Avola seeds were surface sterilized and sown in sterilized boiling tubes with fine vermiculite and nitrogen-free rooting solution. Pea seeds were grown in the dark for

771 3 days and then transferred to a controlled environment room, where they were
772 grown at 25 °C with a 16:8 h photoperiod for another 4 days. On day 7, 1 ml (108
773 CFU/ml) of washed Rlv3841 cells was added near the root. At 7 days post
774 inoculation, rhizobial cells were harvested from the rhizosphere as previously
775 described (10). RNA was extracted for three biological replicates where the total
776 RNA extracted from the pea rhizosphere of 16 boiling tubes was pooled for each
777 replicate. Quality and quantity of the total RNA was assessed using Experion
778 StdSens (Standard Sensitivity) and HighSens (High Sensitivity) analysis kits. Total
779 RNA (3 mg per sample) was treated with the TURBO DNA-free kit (Invitrogen
780 AM1907) as previously described (10). Depletion of genomic DNA was confirmed by
781 performing a Qubit fluorometer double-stranded DNA broad range assay. Finally, the
782 ribosomal RNA was depleted from the RNA sample using the Illumina Ribo-Zero
783 rRNA removal kit - Gram-negative (MRZGN126) according to the manufacturer's
784 instructions. The rRNA-depleted mRNA was purified using the ZymoResearch RNA
785 Clean & Concentrator 50. mRNA samples were used to generate barcoded cDNA
786 libraries for multiplexing during sequencing using the Ion Total RNA-Seq kit v2
787 (Thermo Fisher Scientific). Each barcoded cDNA library was quantified using the
788 Agilent Bioanalyzer High Sensitivity DNA kit and diluted to a final concentration of 70
789 pM. Equal volumes of the diluted cDNA libraries were pooled before loading on the
790 IonChef for template preparation and chip loading. Finally, the chips were sequenced
791 in an Ion Proton Semiconductor based sequencing platform (Thermo Fisher
792 Scientific). The full dataset is available on the NCBI SRA database, BioProject
793 number PRJNA748006.
794

**Data analysis for RNA-Seq of rhizosphere bacteria**

RNA-Seq data was de-multiplexed based on valid barcodes and data for each library was downloaded in fastq format. The overall quality of the sequencing and the data was assessed based on the Torrent Browser suite sequencing run report summary. Data from each library was assessed using FastQC (Babraham Institute; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and any remaining adapters and low-quality reads were filtered using cutadapt (98). The data for each library was mapped against the Rlv3841 genome using EDGE-pro (99) developed specifically for bacterial RNA-Seq data. EDGE-pro uses Bowtie2 to map the reads to the genome and calculates the frequencies per nucleotide. EDGE-pro calculates the number of reads and RPKM value for each gene feature in the genome including mRNA, rRNA, tRNA. The mapped reads from each library were visualized with the Integrative Genomics Viewer (100) for further analysis.

**dRNA-Seq data for nodule bacteria and bacteroids**

The dRNA-Seq data used for creation of the nodule bacteria and the bacteroid model were described previously (65) and are available on the NCBI SRA database, BioProject number PRJNA667846.

## References


1. Ågren GI, Wetterstedt JÅM, Billberger MFK. 2012. Nutrient limitation on terrestrial plant growth – modeling the interaction between nitrogen and phosphorus. New Phytol 194:953–960.
2. Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol 64:781–805.
3. Ledermann R, Schulte CCM, Poole PS. 2021. How rhizobia adapt to the nodule environment. J Bacteriol 203:e00539-20.
4. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. Nat Rev Microbiol 16:291–303.
5. Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol 7:R34.
6. Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R,

838 Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh
839 KC, Batut J. 2001. The composite genome of the legume symbiont
840 *Sinorhizobium meliloti*. Science 293:668–672.
841 7. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto
842 A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-
843 Vides J, Dávila G. 2006. The partitioned *Rhizobium etli* genome: Genetic and
844 metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci
845 103:3834–3839.
846 8. Jiao J, Ni M, Zhang B, Zhang Z, Young JPW, Chan T-F, Chen WX, Lam H-M,
847 Tian CF. 2018. Coordinated regulation of core and accessory genes in the
848 multipartite genome of *Sinorhizobium fredii*. PLOS Genet 14:e1007428.
849 9. DiCenzo GC, Checcucci A, Bazzicalupo M, Mengoni A, Viti C, Dziewit L, Finan
850 TM, Galardini M, Fondi M. 2016. Metabolic modelling reveals the specialization
851 of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nat
852 Commun 7:1–10.
853 10. Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS. 2011.
854 Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet
855 rhizospheres investigated by comparative transcriptomics. Genome Biol
856 12:R106.
857 11. Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali
858 C, diCenzo GC, Mengoni A. 2021. Nonadditive transcriptomic signatures of
859 genotype-by-genotype interactions during the initiation of plant-rhizobium
860 symbiosis. mSystems 6:e00974-20.
861 12. Pini F, East AK, Appia-Ayme C, Tomek J, Karunakaran R, Mendoza-Suárez

M, Edwards A, Terpolilli JJ, Roworth J, Downie JA, Poole PS. 2017. Bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion. Plant Physiol 174:1289–1306.

13. Wheatley RM, Ford BL, Li L, Aroney STN, Knights HE, Ledermann R, East AK, Ramachandran VK, Poole PS. 2020. Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. Proc Natl Acad Sci 117:23823–23834.

14. Checcucci A, DiCenzo GC, Bazzicalupo M, Mengoni A. 2017. Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. Front Microbiol 8:2207.

15. Mendoza-Suárez MA, Geddes BA, Sánchez-Cañizares C, Ramírez-González RH, Kirchhelle C, Jorrin B, Poole PS. 2020. Optimizing *Rhizobium*-legume symbioses by simultaneous measurement of rhizobial competitiveness and $N_2$ fixation in nodules. Proc Natl Acad Sci 117:9822–9831.

16. Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA, Leach JE, Dangl JL. 2017. Research priorities for harnessing plant microbiomes in sustainable agriculture. PLOS Biol 15:e2001793.

17. DiCenzo GC, Zamani M, Checcucci A, Fondi M, Griffitts JS, Finan TM, Mengoni A. 2019. Multidisciplinary approaches for studying rhizobium–legume symbioses. Can J Microbiol 65:1-33.

18. Nielsen J. 2017. Systems biology of metabolism. Annu Rev Biochem 86:245–275.

19. Thiele I, Palsson B. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5:93–121.

20. Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? Nat

Biotechnol 28:245-248.

21. Blazier AS, Papin JA. 2012. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 3:299.

22. Contador CA, Lo S-K, Chan SHJ, Lam H-M. 2020. Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. mSystems 5:e00516-19.

23. Schulte CCM, Borah K, Wheatley RM, Terpolilli JJ, Saalbach G, Crang N, de Groot DH, Ratcliffe RG, Kruger NJ, Papachristodoulou A, Poole PS. 2021. Metabolic constraints on nitrogen fixation by rhizobia in legume nodules. bioRxiv https://doi.org/10.1101/2021.02.16.431433.

24. Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. 2011. Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. BMC Syst Biol 5:120.

25. Zhao H, Li M, Fang K, Chen W, Wang J. 2012. *In silico* insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. PLoS One 7:e31287.

26. Yang Y, Hu X-P, Ma B-G. 2017. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states. Mol BioSyst 13:607.

27. diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. Nat Commun 11:1–11.

28. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2016. KEGG: new

perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:353–361.

29. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res 48:D445–D453.

30. Wheatley RM, Ramachandran VK, Geddes BA, Perry BJ, Yost CK, Poole PS. 2017. Role of O2 in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate. J Bacteriol 199:e00572-16.

31. Ryu JY, Kim HU, Lee SY. 2019. Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers. Proc Natl Acad Sci 116:13996–14001.

32. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, Feist AM, Palsson BO. 2017. iML1515, a knowledgebase that computes *Escherichia coli* traits. Nat Biotechnol 35:904–908.

33. Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res 46:7542-7553.

34. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin E V. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49:D274–D281.

35. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, Diener C, Dräger A, Ebert BE, Edirisinghe JN, Faria JP, Feist AM, Fengos G, Fleming RMT, García-Jiménez

B, Hatzimanikatis V, van Helvoirt W, Henry CS, Hermjakob H, Herrgård MJ, Kaafarani A, Kim HU, King Z, Klamt S, Klipp E, Koehorst JJ, König M, Lakshmanan M, Lee DY, Lee SY, Lee S, Lewis NE, Liu F, Ma H, Machado D, Mahadevan R, Maia P, Mardinoglu A, Medlock GL, Monk JM, Nielsen J, Nielsen LK, Nogales J, Nookaew I, Palsson BO, Papin JA, Patil KR, Poolman M, Price ND, Resendis-Antonio O, Richelle A, Rocha I, Sánchez BJ, Schaap PJ, Malik Sheriff RS, Shoaie S, Sonnenschein N, Teusink B, Vilaça P, Vik JO, Wodke JAH, Xavier JC, Yuan Q, Zakhartsev M, Zhang C. 2020. MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol 38:272–276.

36. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11:1246–1255.

37. Bartell JA, Blazier AS, Yen P, Thøgersen JC, Jelsbak L, Goldberg JB, Papin JA. 2017. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. Nat Commun 8:14631.

38. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ. 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci 110:20338–20343.

39. Blazier AS, Papin JA. 2019. Reconciling high-throughput gene essentiality data with metabolic network reconstructions. PLOS Comput Biol 15:e1006507.

40. Terpolilli JJ, Masakapalli SK, Karunakaran R, Webb IUC, Green R, Watmough NJ, Kruger NJ, Ratcliffe RG, Poole PS. 2016. Lipogenesis and redox balance

in nitrogen-fixing pea bacteroids. J Bacteriol 198:2864–2875.

41. Jenior ML, Moutinho Jr. TJ, Dougherty B V, Papin JA. 2020. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. PLOS Comput Biol 16:e1007099.

42. Rudrappa T, Czymmek KJ, Paré PW, Bais HP. 2008. Root-secreted malic acid recruits beneficial soil bacteria. Plant Physiol 148:1547–1556.

43. Biedrzycki ML, Bais HP. 2009. Root secretions: from genes and molecules to microbial associations. J Exp Bot 60:1533–1534.

44. Jacoby RP, Martyn A, Kopriva S. 2018. Exometabolomic profiling of bacterial strains as cultivated using *Arabidopsis* root extract as the sole carbon source. Mol Plant Microbe Interact 31:803–813.

45. Baran R, Brodie EL, Mayberry-Lewis J, Hummel E, Da Rocha UN, Chakraborty R, Bowen BP, Karaoz U, Cadillo-Quiroz H, Garcia-Pichel F, Northen TR. 2015. Exometabolite niche partitioning among sympatric soil bacteria. Nat Commun 6:8289.

46. Jacoby RP, Kopriva S. 2019. Metabolic niches in the rhizosphere microbiome: new tools and approaches to analyse metabolic mechanisms of plant-microbe nutrient exchange. J Exp Bot 70:1087–1094.

47. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP, Firestone MK, Northen TR, Brodie EL. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat Microbiol 3:470–480.

48. Gaworzewska ET, Carlile MJ. 1982. Positive chemotaxis of *Rhizobium*

*leguminosarum* and other bacteria towards root exudates from legumes and other plants. Microbiology 128:1179–1188.

49. Knee EM, Gong F-C, Gao M, Teplitski M, Jones AR, Foxworthy A, Mort AJ, Bauer WD. 2001. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. Mol Plant-Microbe Interact 14:775–784.

50. Annie L, Magne Ø, Karine M, Marie-Christine P, Daniel LR. 2001. Fructose uptake in *Sinorhizobium meliloti* is mediated by a high-affinity ATP-binding sassette transport system. J Bacteriol 183:4709–4717.

51. Ding H, Yip CB, Geddes BA, Oresnik IJ, Hynes MF. 2012. Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. Microbiology 158:1369–1378.

52. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009. Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. Proc Natl Acad Sci 106:12477–12482.

53. Sanjuán-Pinilla JM, Muñoz S, Nogales J, Olivares J, Sanjuán J. 2002. Involvement of the *Sinorhizobium meliloti leuA* gene in activation of nodulation genes by NodD1 and luteolin. Arch Microbiol 178:36–44.

54. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, Wang G, Ul-Hasan S, McDonald M, Yoshikuni Y, Malmstrom RR, Deutschbauer AM, Dangl JL, Visel A. 2017. Genome-wide identification of bacterial plant colonization genes. PLOS Biol 15:e2002860.

55. Becerra-Rivera VA, Dunn MF. 2019. Polyamine biosynthesis and biological roles in rhizobia. FEMS Microbiol Lett 366:fnz084.

56. Cheng G, Karunakaran R, East AK, Munoz-Azcarate O, Poole PS. 2017.

Glutathione affects the transport activity of *Rhizobium leguminosarum* 3841 and is essential for efficient nodulation. FEMS Microbiol Lett 364:fnx045.

57. Webb I, Xu J, Sanchez-Cañizares C, Karunakaran R, Ramachandran V, Rutten P, East A, Huang W, Watmough N, Poole P. 2021. Regulation and characterization of mutants of *fixABCX* in *Rhizobium leguminosarum*. Mol Plant-Microbe Interact https://doi.org/10.1094/MPMI-02-21-0037-R.

58. Fry J, Wood M, Poole PS. 2001. Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. Mol Plant-Microbe Interact 14:1016–1025.

59. Yost CK, Rath AM, Noel TC, Hynes MF. 2006. Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. Microbiology 152:2061–2074.

60. Barbier T, Collard F, Zúñiga-Ripa A, Moriyón I, Godard T, Becker J, Wittmann C, Van Schaftingen E, Letesson JJ. 2014. Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in *Brucella*. Proc Natl Acad Sci 111:17815–17820.

61. Patil KR, Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci 102:2685–2689.

62. Newman JD, Diebold RJ, Schultz BW, Noel KD. 1994. Infection of soybean and pea nodules by *Rhizobium* spp. purine auxotrophs in the presence of 5-aminoimidazole-4-carboxamide riboside. J Bacteriol 176:3286–3294.

63. Ofaim S, Sulheim S, Almaas E, Sher D, Segrè D. 2021. Dynamic allocation of carbon storage and nutrient-dependent exudation in a revised genome-scale model of *Prochlorococcus*. Front Genet 12:586293.

64. Vasse J, F de B, Camut S, Truchet G. 1990. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol 172:4295–4306.

65. Rutten PJ, Steel H, Hood GA, Ramachandran VK, McMurtry L, Geddes B, Papachristodoulou A, Poole PS. 2021. Multiple sensors provide spatiotemporal oxygen regulation of gene expression in a *Rhizobium*-legume symbiosis. PLOS Genet 17:e1009099.

66. Fournier J, Timmers ACJ, Sieberer BJ, Jauneau A, Chabaud M, Barker DG. 2008. Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. Plant Physiol 148:1985–1995.

67. Ronson CW, Lyttleton P, Robertson JG. 1981. C(4)-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. Proc Natl Acad Sci 78:4284–8.

68. Prell J, Bourdés A, Karunakaran R, Lopez-Gomez M, Poole P. 2009. Pathway of γ-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. J Bacteriol 191:2177–2186.

69. Poole PS, Blyth A, Reid CJ, Walters K. 1994. *Myo*-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. Microbiology 140:2787–2795.

70. Oono R, Denison RF, Kiers ET. 2009. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? New Phytol 183:967–979.

71. Mulley G, White JP, Karunakaran R, Prell J, Bourdes A, Bunnewell S, Hill L, Poole PS. 2011. Mutation of GOGAT prevents pea bacteroid formation and N2

fixation by globally downregulating transport of organic nitrogen sources. Mol Microbiol 80:149–167.

72. Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS. 2009. Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J Bacteriol 191:4002–14.

73. Finan TM, Wood JM, Jordan DC. 1983. Symbiotic properties of C4-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. J Bacteriol 154:1403–1413.

74. Mitsch MJ, DiCenzo GC, Cowie A, Finan TM. 2018. Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. Appl Environ Microbiol 84:e01561-17.

75. Johnson G V, Evans HJ, Ching T. 1966. Enzymes of the glyoxylate cycle in rhizobia and nodules of legumes. Plant Physiol 41:1330–1336.

76. Green RT, East AK, Karunakaran R, Downie JA, Poole PS. 2019. Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. Microb Genomics 5:e000254.

77. Yadav, Vashishat, Kuykendall, Hashem. 1998. Biochemical and symbiotic properties of histidine-requiring mutants of *Rhizobium leguminosarum* biovar *trifolii*. Lett Appl Microbiol 26:22–26.

78. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. 2019. Current status and applications of genome-scale metabolic models. Genome Biol 20:121.

79. Kim WJ, Kim HU, Lee SY. 2017. Current state and applications of microbial genome-scale metabolic models. Curr Opin Syst Biol 2:10-18.

80. Machado D, Herrgård M. 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol 10:e1003580.
81. Liu Y, Beyer A, Aebersold R. 2016. On the dependency of cellular protein levels on mRNA Abundance. Cell 165:535–550.
82. Dunn MF. 2015. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. Crit Rev Microbiol 41:411–451.
83. Griesemer M, Kimbrel JA, Zhou CE, Navid A. 2018. Combining multiple functional annotation tools increases coverage of metabolic annotation. BMC Genomics 19:948.
84. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ. 2018. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. PLOS Comput Biol 14:e1006541.
85. Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:977–982.
86. Price MN, Deutschbauer AM, Arkin AP. 2020. GapMind: Automated annotation of amino acid biosynthesis. mSystems 5:e00291-20.
87. Mauchline TH, Fowler JE, East AK, Sartor AL, Zaheer R, Hosie AHF, Poole PS, Finan TM. 2006. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc Natl Acad Sci 103:17933-17938.
88. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all

domains of life. Nucleic Acids Res 45:D320–D324.

89. Orgambide GG, Huang Z-H, Gage DA, Dazzo FB. 1993. Phospholipid and fatty acid compositions ofRhizobium leguminosarum biovartrifolii ANU843 in relation to flavone-activated pSymnod gene expression. Lipids 28:975–979.

90. Théberge M-C, Prévost D, Chalifour F-P. 1996. The effect of different temperatures on the fatty acid composition of *Rhizobium leguminosarum* bv. *viciae* in the faba bean symbiosis. New Phytol 134:657–664.

91. diCenzo GC, Benedict AB, Fondi M, Walker GC, Finan TM, Mengoni A, Griffitts JS. 2018. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. PLoS Genet 14:e1007357.

92. Breedveld MW, Miller KJ. 1994. Cyclic beta-glucans of members of the family Rhizobiaceae. Microbiol Rev 58:145-161.

93. Xavier JC, Patil KR, Rocha I. 2017. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. Metab Eng 39:200-208.

94. Lodwig EM, Leonard M, Marroqui S, Wheeler TR, Findlay K, Downie JA, Poole PS. 2005. Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. Mol Plant-Microbe Interact 18:67–74.

95. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel L V., Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A,

Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson BØ, Thiele I, Fleming RMT. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc 14:639–702.

96. Glenn AR, Dilworth MJ. 1981. Oxidation of substrates by isolated bacteroids and free-living cells of *Rhizobium leguminosarum* 3841. Microbiology 126:243–247.

97. Galardini M, Mengoni A, Biondi EG, Semeraro R, Florio A, Bazzicalupo M, Benedetti A, Mocali S. 2014. DuctApe: A suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data. Genomics 103:1–10.

98. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10-12.

99. Magoc T, Wood D, Salzberg SL. 2013. EDGE-pro: Estimated degree of gene expression in prokaryotic genomes. Evol Bioinform Online 9:127–136.

100. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

## Data availability

All data need to evaluate the conclusions in this paper are present in the paper and/or the Supplementary Materials. RNA-Seq data for Rlv3841 in the pea rhizosphere are available on the NCBI SRA database, BioProject number PRJNA748006. All code is available on Github (https://github.com/CarolinSchulte/Rlv3481-lifestyles).


## Acknowledgements

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant numbers BB/M011224/1, BB/R017859/1, BB/T001801/1 and BB/T006722/1]. C.C.M.S. is supported by the Clarendon Fund (Oxford University) and the Keble College De Breyne Scholarship. A.P. was funded in part by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/M002454/1].


## Author contributions

Conceptualization: CCMS, AP, PSP

Methodology: CCMS, AP, PSP

Formal analysis: CCMS, VKR

Investigation: CCMS, VKR, AP, PSP

Visualization: CCMS

Supervision: AP, PSP

Writing—original draft: CCMS

Writing—review & editing: CCMS, VKR, AP, PSP

**Supplementary Data**

Supplementary Data 1: sbml file of iCS1224

Supplementary Data 2: Excel file of iCS1224

Supplementary Data 3: MEMOTE report for iCS1224

Supplementary Data 4: Phenotype microarray data for Rlv3841

Supplementary Data 5: RNA-Seq data for Rlv3841 in the pea rhizosphere

Supplementary Data 6: Rhizosphere model of Rlv3841

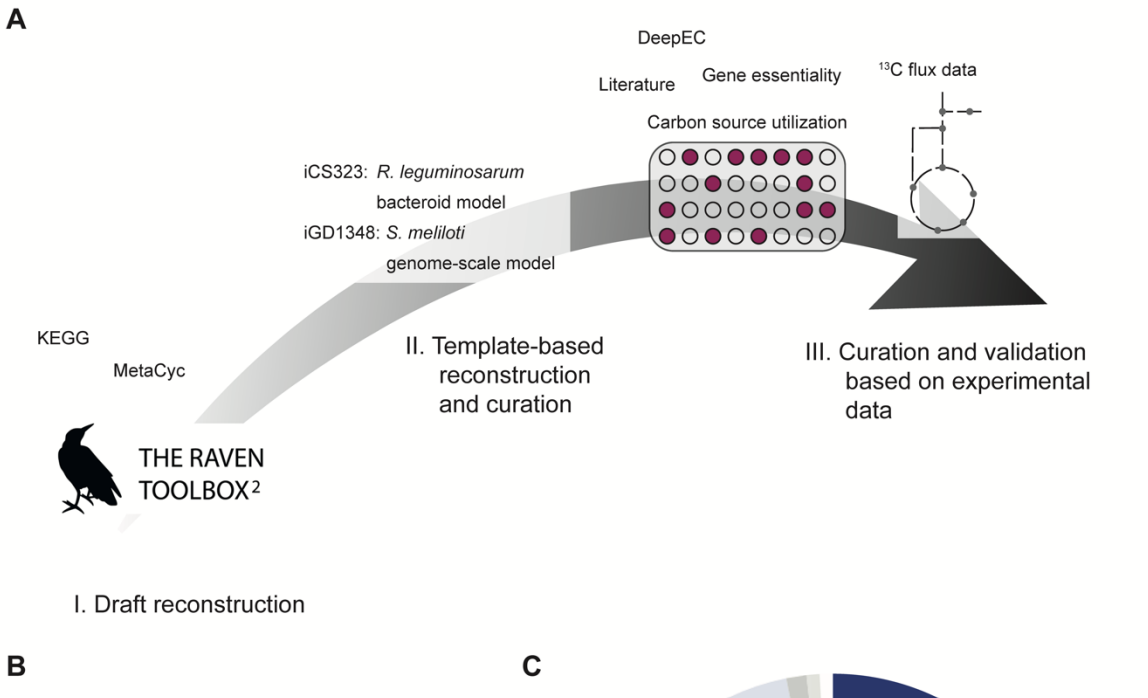
Supplementary Data 7: Metabolomics data

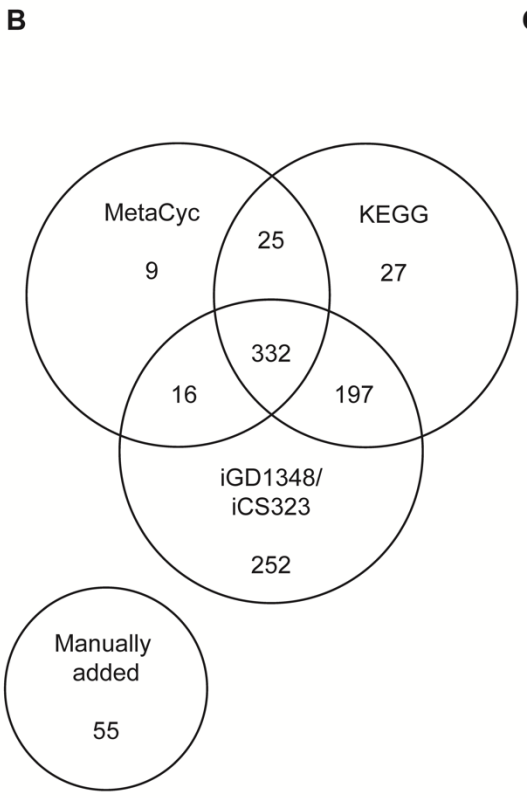
Supplementary Data 8: Nodule bacteria model of Rlv3841

Supplementary Data 9: Bacteroid model of Rlv3841

**Table 1: Properties of iCS1224.**

| Feature | Value |
|---|---|
| Genes | 1224 |
| Metabolites | 984 |
| Unique EC identifiers | 603 |
| Reactions | 1257 |
| Metabolic reactions | 913 |
| Gene-associated metabolic reactions | 897 |
| Transport reactions | 162 |
| Gene-associated transport reactions | 142 |
| Sink reactions | 155 |
| Demand reactions | 15 |
| Other reactions (e.g. DNA synthesis, protein synthesis, biomass objective function) | 12 |


A
DeepEC
Literature
Gene essentiality
13C flux data
Carbon source utilization
iCS323: R. leguminosarum
bacteroid model
iGD1348: S. meliloti
genome-scale model
KEGG
MetaCyc
THE RAVEN
TOOLBOX 2
I. Draft reconstruction
II. Template-based
reconstruction
and curation
III. Curation and validation
based on experimental
data


B
MetaCyc
9
25
KEGG
27
332
16
197
iGD1348/
iCS323
252
Manually
added
55

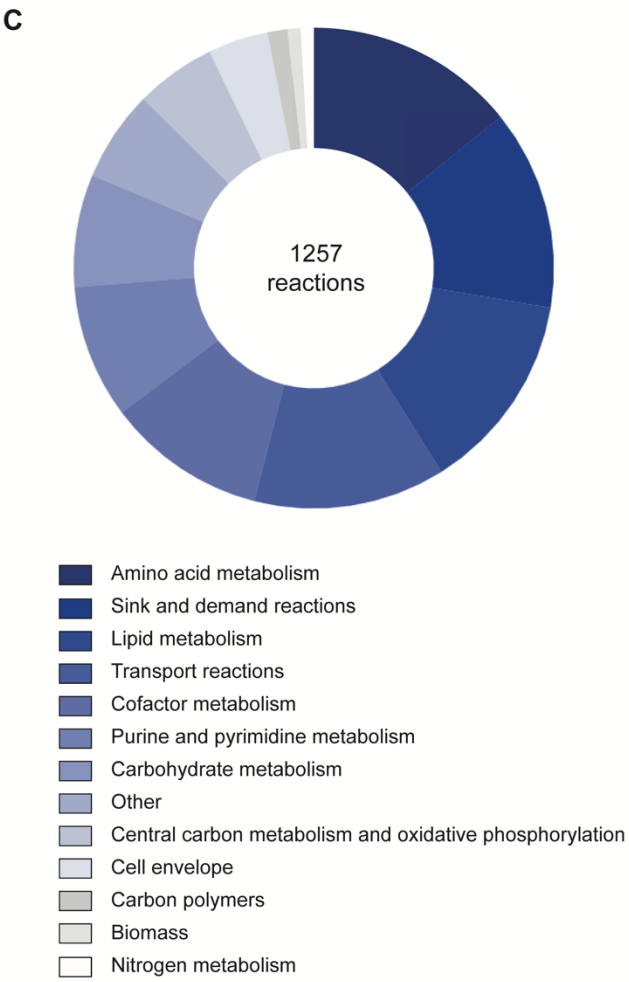

C
1257
reactions
Amino acid metabolism
Sink and demand reactions
Lipid metabolism
Transport reactions
Cofactor metabolism
Purine and pyrimidine metabolism
Carbohydrate metabolism
Other
Central carbon metabolism and oxidative phosphorylation
Cell envelope
Carbon polymers
Biomass
Nitrogen metabolism

**Fig. 1: Reconstruction of a genome-scale model for *Rhizobium leguminosarum* bv. *viciae* 3841.** A: Reconstruction process for iCS1224 using automated reconstruction, template-based reconstruction, and data-based curation. B: Sources for the 913 metabolic reactions in iCS1224. Numbers indicate how many reactions from KEGG, MetaCyc or the template-based reconstruction were included in the final model, with numbers in the overlapping areas indicating reactions that were present in multiple draft reconstructions. C: Classification of the reactions in iCS1224.

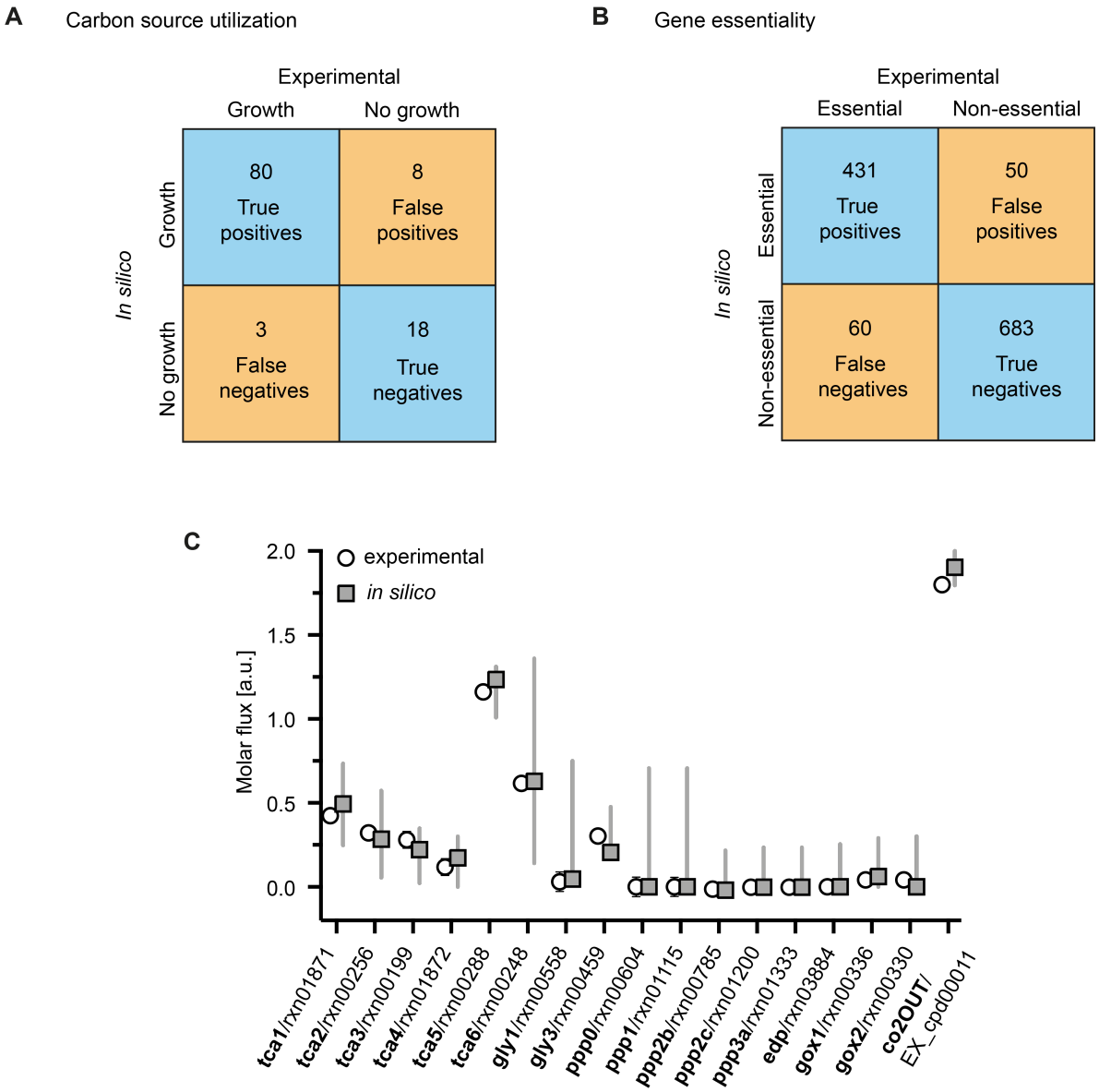



**Fig 2. Validation of iCS1224.** A: Table showing the agreement between carbon source utilization experimentally measured with phenotype microarrays and predicted by iCS1224. B: Table showing the agreement between gene essentiality determined by insertion sequencing (13, 30) and predicted by iCS1224. C: Comparison of metabolic fluxes determined by $^{13}C$ metabolic flux analysis for Rlv3841 grown on succinate (40) with flux rates predicted by iCS1224. For experimental data, symbols and bars indicate mean ± SD. Note the error bars are

too small to be visible for most data points. For *in silico* data, symbols represent the flux rate predicted by flux balance analysis, with lines indicating upper and lower bounds for each flux determined by flux variability analysis with at least 95% of the optimum flux through the biomass objective function. Labels on the x axis indicate the name of the reaction as reported in (40) (in bold), as well as the reaction identifier in the model.

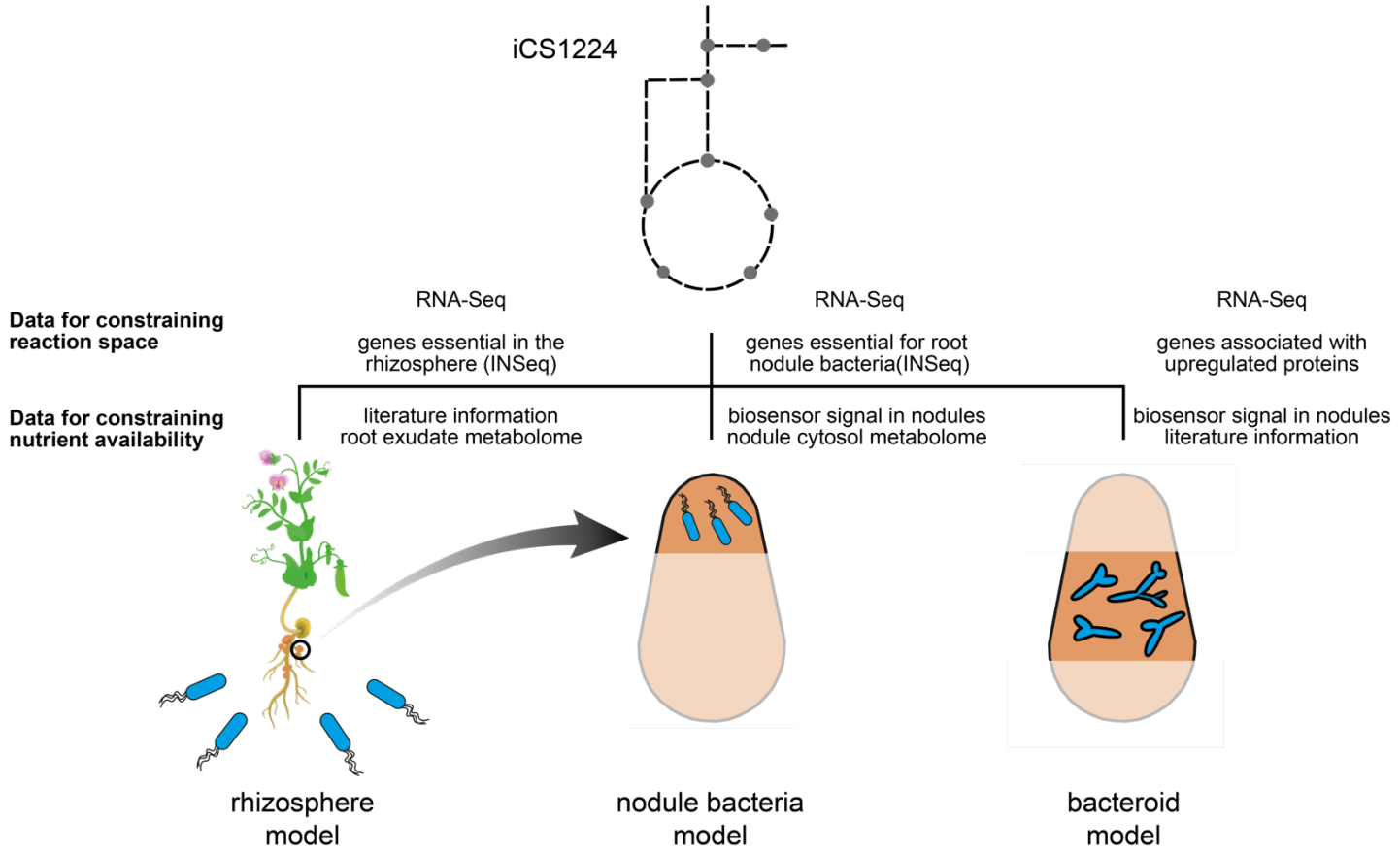



**Fig. 3: Approach for generating lifestyle-specific models for *Rhizobium leguminosarum* bv. *viciae*.** Based on iCS1224, transcriptome, gene essentiality and proteome data specific to a certain lifestyle were used to inform the extraction of context-specific models for the rhizosphere, nodule bacteria and nitrogen-fixing bacteroids. Boundary conditions were defined based on metabolome data and/or literature information.

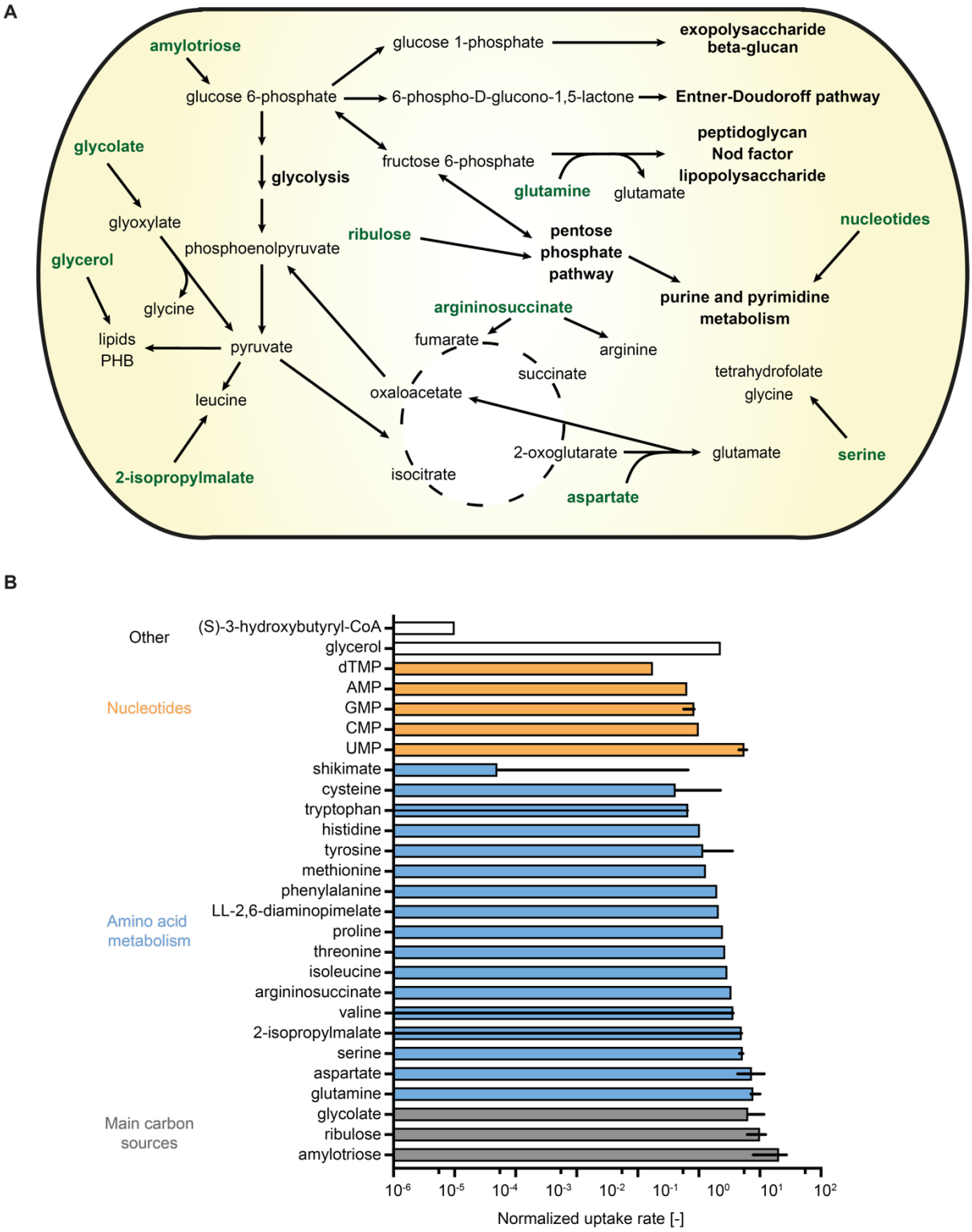



**Fig. 4: Metabolism of *Rhizobium leguminosarum* in the pea rhizosphere.** A rhizosphere-specific model was extracted from iCS1224 using the RIPTiDe algorithm with RNA-Seq and gene essentiality data for *R. leguminosarum* in the rhizosphere of pea plants. A: Schematic representation of the main pathways predicted to be active

in the rhizosphere-specific model. Compounds predicted to be taken up are indicated in bold green. Note that the magnitude of flux is not indicated in this summary map. B: Bar graph showing the uptake rates of metabolites predicted to be taken up from pea root exudate. Absolute flux values for the exchange reactions were normalized by flux through the biomass reaction in each sample. Only metabolites with non-zero median uptake for the 500 samples of the contextualized model are shown. Uptake of ions and cofactors has been omitted for clarity. Bars and lines indicate median and interquartile range, respectively.

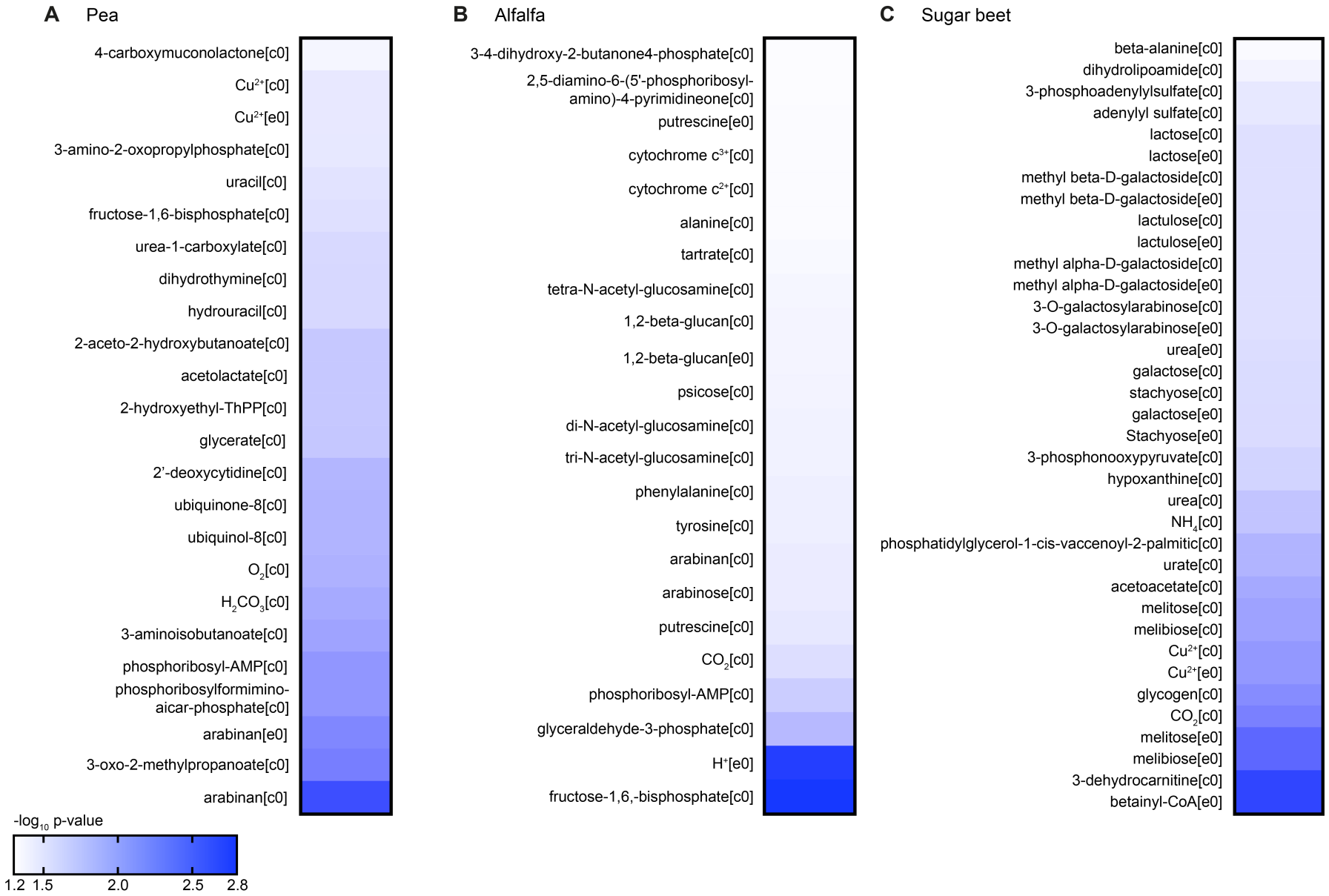

A Pea
4-carboxymuconolactone[c0]
Cu2+[c0]
Cu2+[e0]
3-amino-2-oxopropylphosphate[c0]
uracil[c0]
fructose-1,6-bisphosphate[c0]
urea-1-carboxylate[c0]
dihydrothymine[c0]
hydrouracil[c0]
2-aceto-2-hydroxybutanoate[c0]
acetolactate[c0]
2-hydroxyethyl-ThPP[c0]
glycerate[c0]
2'-deoxycytidine[c0]
ubiquinone-8[c0]
ubiquinol-8[c0]
O2[c0]
H2CO3[c0]
3-aminoisobutanoate[c0]
phosphoribosyl-AMP[c0]
phosphoribosylformimino-aicar-phosphate[c0]
arabinan[e0]
3-oxo-2-methylpropanoate[c0]
arabinan[c0]
-log10 p-value
1.2 1.5 2.0 2.5 2.8
B Alfalfa
3-4-dihydroxy-2-butanone4-phosphate[c0]
2,5-diamino-6-(5'-phosphoribosyl-amino)-4-pyrimidineone[c0]
putrescine[e0]
cytochrome c3+[c0]
cytochrome c2+[c0]
alanine[c0]
tartrate[c0]
tetra-N-acetyl-glucosamine[c0]
1,2-beta-glucan[c0]
1,2-beta-glucan[e0]
psicose[c0]
di-N-acetyl-glucosamine[c0]
tri-N-acetyl-glucosamine[c0]
phenylalanine[c0]
tyrosine[c0]
arabinan[c0]
arabinose[c0]
putrescine[c0]
CO2[c0]
phosphoribosyl-AMP[c0]
glyceraldehyde-3-phosphate[c0]
H+[e0]
fructose-1,6,-bisphosphate[c0]
C Sugar beet
beta-alanine[c0]
dihydrolipoamide[c0]
3-phosphoadenylylsulfate[c0]
adenylyl sulfate[c0]
lactose[c0]
lactose[e0]
methyl beta-D-galactoside[c0]
methyl beta-D-galactoside[e0]
lactulose[c0]
lactulose[e0]
methyl alpha-D-galactoside[c0]
methyl alpha-D-galactoside[e0]
3-O-galactosylarabinose[c0]
3-O-galactosylarabinose[e0]
urea[e0]
galactose[c0]
stachyose[c0]
galactose[e0]
Stachyose[e0]
3-phosphonooxypyruvate[c0]
hypoxanthine[c0]
urea[c0]
NH4[c0]
phosphatidylglycerol-1-cis-vaccenoyl-2-palmitic[c0]
urate[c0]
acetoacetate[c0]
melitose[c0]
melibiose[c0]
Cu2+[c0]
Cu2+[e0]
glycogen[c0]
CO2[c0]
melitose[e0]
melibiose[e0]
3-dehydrocarnitine[c0]
betainyl-CoA[e0]

**Fig. 5: Reporter metabolites in different rhizospheres.** Reporter metabolites were calculated using microarray data for *Rhizobium leguminosarum* bv. *viciae* 3841 in the rhizosphere of pea (A), alfalfa (B), and sugar beet (C) compared to free living cells grown in minimal media with glucose and ammonia (10). The heatmaps show the negative decimal logarithm of the *P* value for those metabolites that were associated with significant ($P<0.05$) transcriptional changes among genes upregulated in the rhizosphere. [c0] and [e0] indicate cytosolic and extracellular metabolites, respectively.

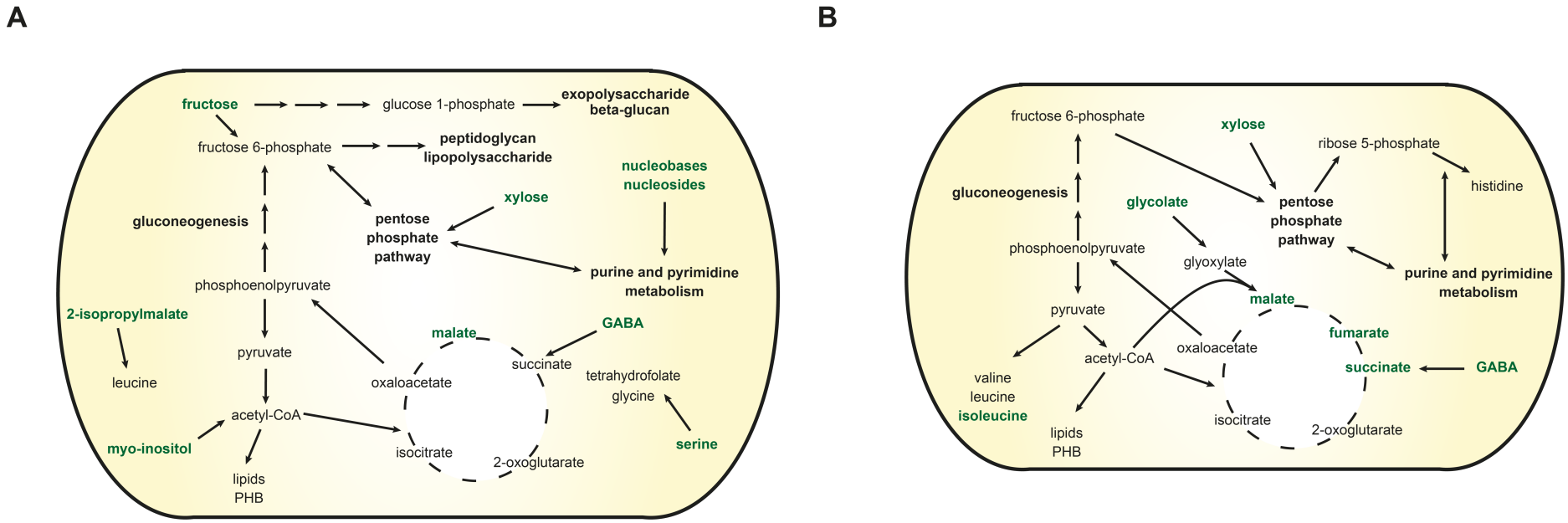



**Fig. 6: Metabolism of undifferentiated nodule bacteria and nitrogen-fixing bacteroids.** Maps showing schematic representations of the main pathways predicted to be active in undifferentiated nodule bacteria (A) and nitrogen-fixing bacteroids (B). Compounds predicted to be taken up are indicated in bold green. Note that the magnitude of flux is not indicated in these summary maps.

# Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum*

## Supplementary Information

Carolin C. M. Schulte, Vinoy K. Ramachandran, Antonis Papachristodoulou, Philip S. Poole

## Supplementary Text

### Note on the selection of boundary conditions

A key objective of integrating gene expression data with metabolic models in complex environments, such as the rhizosphere or nodule, is the elucidation of metabolites which are taken up, since experimental determination of uptake rates for all nutrients would require extensive metabolomics studies. We therefore sought to define nutrient availability prior to data integration as broadly as possible to enable unbiased investigation of nutrients that are predicted to be taken up. However, it was observed that allowing uptake of central intermediates of carbon metabolism, such as 2-phosphoglycerate or pyruvate, generally resulted in high uptake rates of these metabolites. This can be explained by high metabolic activity around these intermediates, which is expected irrespective of the specific carbon source, however high uptake of these intermediates is unlikely to be biologically relevant. We therefore chose to omit these compounds from the environmental representations. Metabolites that were converted into amino acids in a single reaction were excluded to distinguish

between amino acid synthesis and uptake. Uptake of cofactors, such as $NAD^+$ or FAD, was not permitted in the model to ensure a balanced redox metabolism.

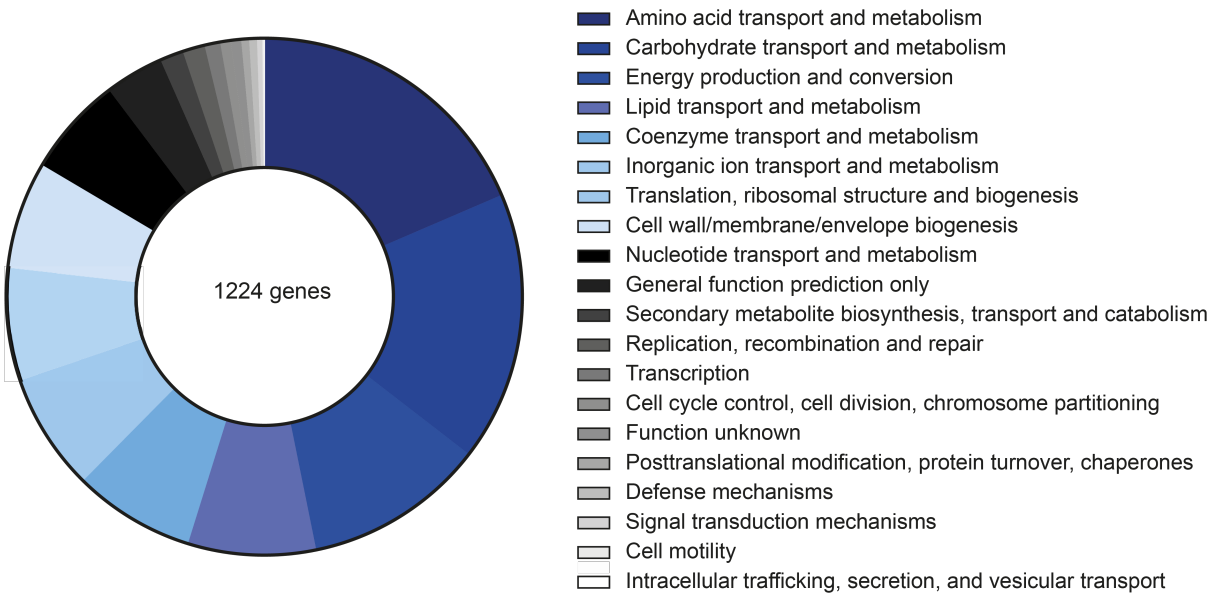



**Fig. S1: COG classification of model genes.** Pie chart showing the distribution of COG classifications for the genes in iCS1224. COG classification was performed with WebMGA (1).

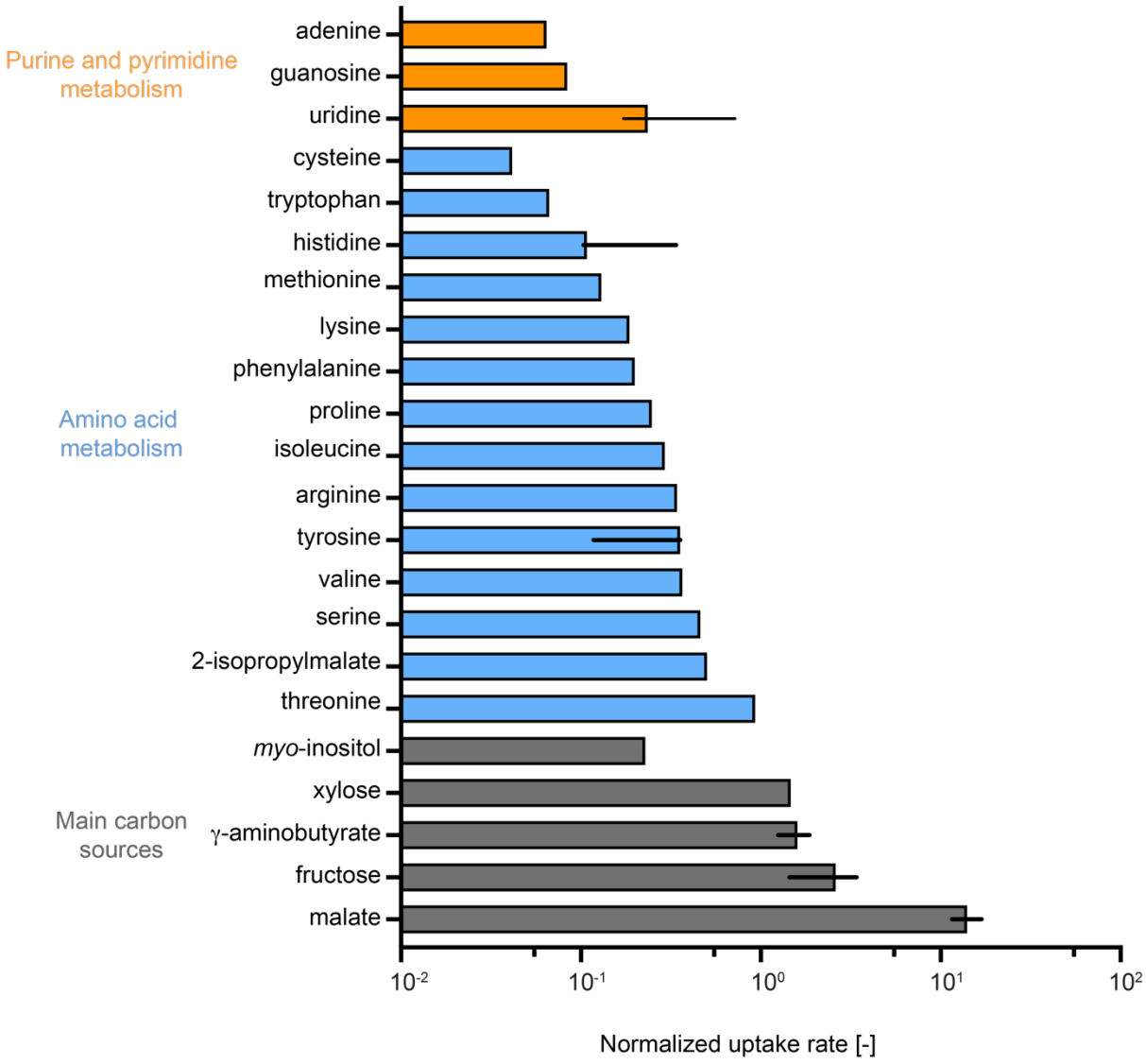



**Fig. S2: Nutrient uptake of nodule bacteria.** Bar graph showing the uptake rates of metabolites predicted to be taken up by undifferentiated nodule bacteria. Absolute flux values for the exchange reactions were normalized by flux through the biomass reaction in each sample. Only metabolites with non-zero median uptake for the 500 flux samples of the contextualized model are shown. Uptake of ions and cofactors has been omitted for clarity. Bars and lines indicate median and interquartile range, respectively.

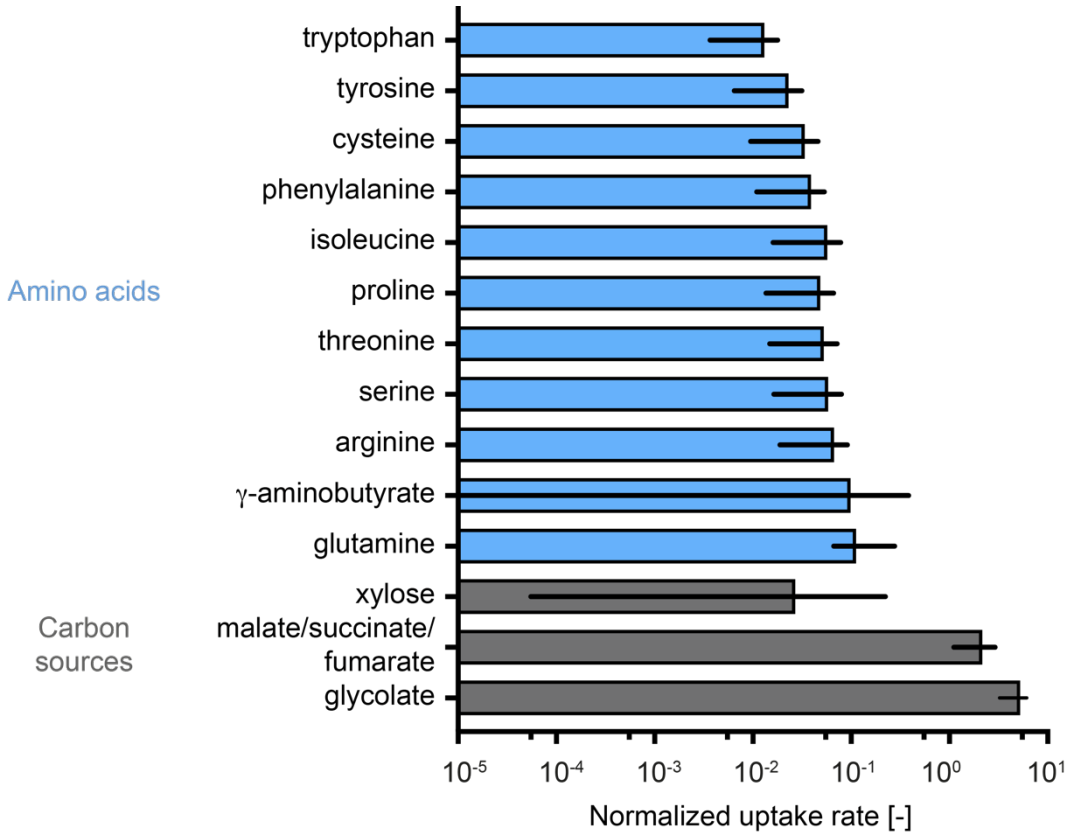



**Fig. S3: Nutrient uptake of bacteroids.** Bar graph showing the uptake rates of metabolites predicted to be taken up by nitrogen-fixing bacteroids. Absolute flux values for the exchange reactions were normalized by flux through the nitrogenase reaction in each sample. Only metabolites with non-zero median uptake for the 500 flux samples of the contextualized model are shown. Uptake of ions and cofactors has been omitted for clarity. Bars and lines indicate median and interquartile range, respectively.

**Table S1. Biomass composition for free-living *Rhizobium leguminosarum* bv. *viciae* 3841.**

| Biomass component | Cellular content (g/g) | Comment | Reference |
|---|---|---|---|
| DNA | 0.028 | Value for *S. meliloti* | (2) |
| RNA | 0.074 | Interpolated from values for *E. coli* assuming a growth rate of 0.16 $h^{-1}$ | (3) |
| protein | 0.55 | Value for *E. coli* | BNID 106154 (4) |
| phospholipids | 0.091 | Value for *E. coli*; composition for *R. leguminosarum* bv. *trifolii*<br>2% phosphatidylcholine<br>5% cardiolipin<br>15% phosphatidyl-ethanolamine<br>19% phosphatidylglycerol<br>47% monomethyl-phosphatidylethanolamine<br>9% dimethylphosphatidyl-ethanolamine | BNID 101436 (4, 5) |
| peptidoglycan | 0.025 | Value for *E. coli* | BNID 106154 (4) |
| LPS | 0.015 | | (6) |
| EPS | 0.05 | Value for *S. meliloti* | (2) |
| beta-glucan | 0.0927 | | (7) |
| glycogen | 0.0386 | | (8) |

| | | | |
|---|---|---|---|
| polyhydroxybutyrate | 0.0358 | | (8) |
| Cofactors | Trace | | (9) |
| NAD | ($1e^{-5}$) | | |
| NADP | | | |
| FAD | | | |
| S-adenosyl-methionine | | | |
| CoA | | | |
| FMN | | | |
| riboflavin | | | |
| tetrahydrofolate | | | |
| 10-formyltetrahydrofolate | | | |
| 5-methyltetrahydrofolate | | | |
| thiamine diphosphate | | | |
| pyridoxal-5-phosphate | | | |
| hemeO | | | |
| ubiquinone-8 | | | |
| biotinyl-5-AMP | | | |
| lipoyl-acyl carrier protein | | | |
| cobalamin | | | |
| UDP-N-acetylglucosamine | | | |
| putrescine | | | |
| phytoene | | | |
| folate | | | |
| homospermidine | | | |
| putrescine | | | |
| trace metals | | | |

**Table S2. Compounds included in the *in silico* rhizosphere representation.**

| Compound | ModelSEED ID | Reference |
|---|---|---|
| shikimate | cpd00383 | (10) |
| protocatechuate | cpd00197 | (10) |
| mannitol | cpd00314 | (10) |
| arabinan | cpd12115 | (10) |
| tartrate | cpd00666 | (10, 11) |
| homoserine | cpd00227 | (10, 12) |
| tryptophan | cpd00065 | (10, 12) |
| phenylalanine | cpd00066 | (10–13) |
| *myo*-inositol | cpd00121 | (11) |
| malonate | cpd00308 | (11) |
| fructose | cpd00082 | (11, 12) |
| succinate | cpd00036 | (11, 12) |
| malate | cpd00130 | (11, 12) |
| proline | cpd00129 | (13, 14) |
| alanine | cpd00035 | (12, 13) |
| arginine | cpd00051 | (12, 13) |
| aspartate | cpd00041 | (12) |
| asparagine | cpd00132 | (12) |
| cysteine | cpd00084 | (12) |
| glutamate | cpd00023 | (12, 13) |
| glutamine | cpd00053 | (12, 13) |
| glycine | cpd00033 | (12, 13) |
| histidine | cpd00119 | (12) |
| isoleucine | cpd00322 | (12, 13) |
| leucine | cpd00107 | (12, 13) |
| lysine | cpd00039 | (12, 13) |
| serine | cpd00054 | (12, 13) |
| threonine | cpd00161 | (12, 13) |
| valine | cpd00156 | (12, 13) |

| | | |
|---|---|---|
| glucose | cpd00027 | (12, 13) |
| maltose | cpd00179 | (12) |
| raffinose | cpd00382 | (12) |
| sucrose | cpd00076 | (12) |
| xylose | cpd00154 | (12, 13) |
| citrate | cpd00137 | (12) |
| fumarate | cpd00106 | (11, 12) |
| glycolate | cpd00139 | (12) |
| arabinose | cpd00185 | (13) |
| fucose | cpd00751 | (13) |
| galactose | cpd00108 | (13) |
| mannose | cpd00138 | (13) |
| rhamnose | cpd00396 | (13) |
| *trans*-4-hydroxyproline | cpd00851 | (13) |
| tyrosine | cpd00069 | (13) |
| 5-oxoproline | cpd01293 | (11) |
| erythritol | cpd00392 | (11) |
| arabitol | cpd01307 | (11) |
| arabonate | cpd00651 | (11) |
| ribitol | cpd00366 | (11) |
| 2'-deoxyguanosine | cpd00277 | (11) |
| 2'-deoxyadenosine | cpd00438 | (11) |
| adenosine | cpd00182 | (11) |
| allantoin | cpd01092 | (11) |
| sorbitol | cpd00588 | (11) |
| xylitol | cpd00306 | (11) |
| lactate | cpd00159 | (11) |
| beta-hydroxypyruvate | cpd00145 | (11) |
| glycerate | cpd00223 | (11) |
| pantothenate | cpd00644 | (11) |
| betaine | cpd00540 | (11) |
| cytidine | cpd00367 | (11) |

| | | |
|---|---|---|
| thymidine | cpd00184 | (11) |
| guanine | cpd00207 | (11) |
| urocanate | cpd00581 | (11) |
| allantoic acid | cpd00388 | (11) |
| 2'-deoxycytidine | cpd00654 | (11) |
| maltotriose | cpd01262 | (11) |
| uridine | cpd00249 | (11) |
| xanthosine 5'-monophosphate | cpd00497 | (11) |
| cytosine | cpd00307 | (11) |
| ribulose | cpd00258 | (11) |
| ribose | cpd00105 | (11) |
| orotate | cpd00247 | (11) |
| galactitol | cpd01171 | (11) |
| heme | cpd00028 | (11) |
| thymine | cpd00151 | (11) |
| glycerol | cpd00100 | (11) |
| 2'-deoxycytidine 5'-monophosphate | cpd00206 | (11) |
| gamma-aminobutyrate | cpd00281 | (11) |
| acetylphosphate | cpd00196 | (11) |
| pyridoxine | cpd00263 | (11) |
| cytidine 5'-monophosphate | cpd00046 | (11) |
| thymidine 5'-monophosphate | cpd00298 | (11) |
| 3-hydroxybutyryl CoA | cpd00842 | (11) |
| 3-hydroxybutyrate | cpd00797 | (11) |
| *cis*-vaccenate | cpd25615 | (11) |
| uracil | cpd00092 | (11) |
| inosine | cpd00246 | (11) |
| guanosine 5'- monophosphate | cpd00126 | (11) |
| ornithine | cpd00064 | (11) |
| thiamin diphosphate | cpd00056 | (11) |
| 2'-deoxyguanosine 5'-monophosphate | cpd00296 | (11) |
| methionine | cpd00060 | (11) |

| | | |
|---|---|---|
| nicotinate | cpd00218 | (11) |
| xanthine | cpd00309 | (11) |
| O-acetylserine | cpd00722 | (11) |
| pyrophosphate | cpd00012 | (11) |
| 4-hydroxyphenylpyruvate | cpd00868 | (11) |
| anthranilate | cpd00093 | (11) |
| xanthosine | cpd01217 | (11) |
| 3-ureidopropionate | cpd01720 | (11) |
| diaminopimelate | cpd00504 | (11) |
| adenine | cpd00128 | (11) |
| 2-isopropylmalate | cpd01646 | (11) |
| argininosuccinate | cpd02152 | (11) |
| trehalose | cpd00794 | (11) |
| phosphopantetheine | cpd00834 | (11) |
| ethanolamine | cpd00162 | (11) |
| 5'-uridine monophosphate | cpd00091 | (11) |
| guanosine | cpd00311 | (11) |
| urate | cpd00300 | (11) |
| 2'-deoxyuridine | cpd00412 | (11) |
| 2'-deoxyinosine | cpd03279 | (11) |
| riboflavin | cpd00220 | (11) |
| putrescine | cpd00118 | (11) |
| agmatine | cpd00152 | (11) |
| adenosine 5'-monophosphate | cpd00018 | (11) |
| hypoxanthine | cpd00226 | (11) |
| $Ca^{2+}$ | cpd00063 | required for biomass |
| $Cl^{-}$ | cpd00099 | required for biomass |
| $K^{+}$ | cpd00205 | required for biomass |

| | | |
|---|---|---|
| $Mg^{2+}$ | cpd00254 | required for biomass |
| $Fe^{2+}$ | cpd10515 | required for biomass |
| $SO_4^{2-}$ | cpd00048 | required for biomass |
| $Mn^{2+}$ | cpd00030 | required for biomass |
| $Zn^{2+}$ | cpd00034 | required for biomass |
| $Cu^{2+}$ | cpd00058 | required for biomass |
| $MoO_4$ | cpd11574 | required for biomass |
| $Co^{2+}$ | cpd00149 | required for biomass |
| $Na^{+}$ | cpd00971 | required for biomass |
| biotin | cpd00104 | required for biomass |
| $H^{+}$ | cpd00067 | |
| $PO_4^{3-}$ | cpd00009 | |
| $O_2$ | cpd00007 | |

**Table S3. Compounds excluded from the rhizosphere composition.**

| Compound | SEED ID | Reason for exclusion |
|---|---|---|
| 3-phosphoglycerate | cpd00169 | Central intermediate of carbon metabolism |
| 1,3-dihydroxyacetone | cpd00157 | Central intermediate of carbon metabolism |
| glycerol 3-phosphate | cpd00080 | Central intermediate of carbon metabolism |
| pyruvate | cpd00020 | Central intermediate of carbon metabolism |
| acetyl-CoA | cpd00022 | Central intermediate of carbon metabolism |
| S-adenosylmethionine | cpd00017 | Cofactor |
| S-adenosyl homocysteine | cpd00019 | Cofactor |
| flavin mononucleotide | cpd00050 | Electron carrier |
| $NADP^+$ | cpd00006 | Electron carrier |
| $NAD^+$ | cpd00003 | Electron carrier |
| FAD | cpd00015 | Electron carrier |
| glutathione, reduced | cpd00042 | Electron carrier |
| glutathione, oxidized | cpd00111 | Electron carrier |
| coenzyme A | cpd00010 | Cofactor |
| beta-alanine | cpd00085 | Inclusion resulted in unrealistically high uptake rate |
| 2-keto-isoleucine | cpd00508 | Direct precursor of isoleucine |
| 2-keto-isocaproate | cpd00200 | Direct precursor of leucine |
| 2-oxoisovalerate | cpd00123 | Direct precursor of valine |
| phenylpyruvate | cpd00143 | Direct precursor of phenylalanine |
| 2-oxoglutarate | cpd00024 | Not utilized by free-living Rlv3841 according to phenotype microarray |

| | | |
|---|---|---|
| mannose 6-phosphate | cpd00235 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| trehalose 6-phosphate | cpd00523 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| ribulose 5-phosphate | cpd00171 | Central intermediate of pentose phosphate pathway; sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| fructose 6-phosphate | cpd00072 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| glucose 6-phosphate | cpd00079 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| N-acetylglutamate | cpd00477 | Not utilized by free-living Rlv3841 according to phenotype microarray |
| KDO | cpd00875 | Component of lipopolysaccharides; unlikely to be taken up |

**Table S4. Compounds included in the *in silico* environment for nodule bacteria.**

The list of compounds is based on the induction of rhizobial biosensors in pea nodules (11) and metabolomics data for the nodule cytosol. In addition, all amino acids and vitamins present in the nodule cytosol were included.

| Compound | SEED ID |
|---|---|
| xylose | cpd00154 |
| fructose | cpd00082 |
| sucrose | cpd00076 |
| erythritol | cpd00392 |
| mannitol | cpd00314 |
| *myo*-inositol | cpd00121 |
| formate | cpd00034 |
| malonate | cpd00308 |
| malate | cpd00130 |
| succinate | cpd00036 |
| fumarate | cpd00106 |
| tartrate | cpd00666 |
| GABA | cpd00281 |
| phenylalanine | cpd00066 |
| tyrosine | cpd00069 |
| tryptophan | cpd00065 |
| leucine | cpd00107 |
| valine | cpd00156 |
| isoleucine | cpd00322 |
| alanine | cpd00035 |
| aspartate | cpd00041 |
| glycine | cpd00033 |
| glutamate | cpd00023 |
| glutamine | cpd00053 |

| | |
|---|---|
| asparagine | cpd00132 |
| histidine | cpd00119 |
| proline | cpd00129 |
| serine | cpd00054 |
| threonine | cpd00161 |
| methionine | cpd00060 |
| cysteine | cpd00084 |
| lysine | cpd00039 |
| arginine | cpd00051 |
| xanthosine | cpd01217 |
| 2-isopropylmalate | cpd01646 |
| uridine | cpd00249 |
| uracil | cpd00092 |
| agmatine | cpd00152 |
| hypoxanthine | cpd00226 |
| adenine | cpd00128 |
| homoserine | cpd00227 |
| $H^+$ | cpd00067 |
| heme | cpd00028 |
| inosine | cpd00246 |
| argininosuccinate | cpd02152 |
| guanosine | cpd00311 |
| O-acetylserine | cpd00722 |
| ethanolamine | cpd00162 |
| thiamin diphosphate | cpd00056 |
| pyridoxine | cpd00263 |
| homocitrate | cpd00919 |
| biotin | cpd00104 |
| pantothenate | cpd00644 |
| phosphopantetheine | cpd00834 |
| niacin | cpd00218 |
| riboflavin | cpd00220 |

| | |
|---|---|
| $O_2$ | cpd00007 |
| $PO_4^{3-}$ | cpd00009 |
| $Cu^{2+}$ | cpd00058 |
| $Ca^{2+}$ | cpd00063 |
| $Cl^-$ | cpd00099 |
| $Co^{2+}$ | cpd00149 |
| $K^+$ | cpd00205 |
| $Mg^{2+}$ | cpd00254 |
| $N_2$ | cpd00528 |
| $Na^+$ | cpd00971 |
| $Fe^{2+}$ | cpd10515 |
| $MoO_4$ | cpd11574 |
| $Zn^{2+}$ | cpd00034 |
| $SO_4^{2-}$ | cpd00048 |

**Table S5. Compounds excluded from the environment for nodule bacteria.**

| Compound | SEED ID | Reason for exclusion |
|---|---|---|
| glycerol 3-phosphate | cpd00080 | Central intermediate of carbon metabolism |
| 3-phosphoglycerate | cpd00169 | Central intermediate of carbon metabolism |
| mannose-6-phosphate | cpd00235 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| fructose 6-phosphate | cpd00072 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| glucose-6-phosphate | cpd00079 | Sugar phosphates not utilized by free-living Rlv3841 according to phenotype microarray |
| pyruvate | cpd00020 | Central intermediate of carbon metabolism |
| 2-oxoglutarate | cpd00024 | Not utilized by free-living Rlv3841 according to phenotype microarray |
| N-acetylglutamate | cpd00477 | Not utilized by free-living Rlv3841 according to phenotype microarray |
| 2-keto-isoleucine | cpd00508 | Direct precursor of isoleucine |
| 2-keto-isocaproate | cpd00200 | Direct precursor of leucine |

| | | |
|---|---|---|
| KDO | cpd00875 | Component of lipopolysaccharides; unlikely to be taken up |
| S-adenosylhomocysteine | cpd00019 | Cofactor |
| $NAD^+$ | cpd00003 | Electron carrier |
| glutathione, reduced | cpd00042 | Electron carrier |
| glutathione, oxidized | cpd00111 | Electron carrier |

**Table S6. Nutrients available to bacteroids.** The list of compounds is based on the induction of rhizobial biosensors in nodules (11). In addition, all amino acids and vitamins present in the nodule cytosol were included.

| Compound | SEED ID |
|---|---|
| xylose | cpd00154 |
| erythritol | cpd00392 |
| mannitol | cpd00314 |
| *myo*-inositol | cpd00121 |
| formate | cpd00034 |
| malonate | cpd00308 |
| malate | cpd00130 |
| succinate | cpd00036 |
| fumarate | cpd00106 |
| tartrate | cpd00666 |
| GABA | cpd00281 |
| phenylalanine | cpd00066 |
| tyrosine | cpd00069 |
| tryptophan | cpd00065 |
| leucine | cpd00107 |
| valine | cpd00156 |
| isoleucine | cpd00322 |
| alanine | cpd00035 |
| aspartate | cpd00041 |
| glycine | cpd00033 |
| glutamate | cpd00023 |
| glutamine | cpd00053 |
| asparagine | cpd00132 |
| histidine | cpd00119 |
| proline | cpd00129 |

| | |
|---|---|
| serine | cpd00054 |
| threonine | cpd00161 |
| methionine | cpd00060 |
| cysteine | cpd00084 |
| lysine | cpd00039 |
| arginine | cpd00051 |
| glycolate | cpd00139 |
| pantothenate | cpd00644 |
| phosphopantetheine | cpd00834 |
| $H^+$ | cpd00067 |
| heme | cpd00028 |
| $Mn^{2+}$ | cpd00030 |
| $O_2$ | cpd00007 |
| $PO_4^{3-}$ | cpd00009 |
| thiamin diphosphate | cpd00056 |
| $Cu^{2+}$ | cpd00058 |
| $Ca^{2+}$ | cpd00063 |
| $Cl^-$ | cpd00099 |
| biotin | cpd00104 |
| $Co^{2+}$ | cpd00149 |
| $K^+$ | cpd00205 |
| niacin | cpd00218 |
| riboflavin | cpd00220 |
| $Mg^{2+}$ | cpd00254 |
| pyridoxine | cpd00263 |
| $N_2$ | cpd00528 |
| homocitrate | cpd00919 |
| $Na^+$ | cpd00971 |
| $Fe^{2+}$ | cpd10515 |
| $MoO_4$ | cpd11574 |
| $Zn^{2+}$ | cpd00034 |
| $SO_4^{2-}$ | cpd00048 |

**Table S7. Enrichment of amino acids in Nif and Fix proteins.** The amino acid compositions of all coding sequences in the Rlv3841 genome and the Nif and Fix proteins were compared using a hypergeometric test followed by Benjamini-Hochberg false discovery rate correction.

| **Amino acid** | **Overall percentage in coding sequences [%]** | **Percentage in Nif and Fix proteins [%]** | **Adjusted *P* value** |
|---|---|---|---|
| alanine | 11.81 | 10.22 | 0.9998 |
| cysteine | 0.82 | 1.53 | $3.0026*10^{-6}$ |
| aspartate | 5.70 | 6.08 | 0.3860 |
| glutamate | 5.72 | 5.94 | 0.5768 |
| phenylalanine | 3.94 | 3.31 | 0.9998 |
| glycine | 8.24 | 8.75 | 0.3460 |
| histidine | 2.04 | 2.57 | 0.0420 |
| isoleucine | 5.79 | 5.96 | 0.6195 |
| lysine | 3.68 | 5.14 | $2.4365*10^{-6}$ |
| leucine | 9.95 | 9.16 | 0.9998 |
| methionine | 2.57 | 2.77 | 0.4936 |
| asparagine | 2.85 | 3.39 | 0.0681 |
| proline | 4.90 | 4.44 | 0.9998 |
| glutamine | 3.09 | 3.15 | 0.7740 |
| arginine | 6.77 | 6.06 | 0.9998 |
| serine | 5.89 | 5.62 | 0.9998 |
| threonine | 5.33 | 5.20 | 0.9998 |
| valine | 7.24 | 6.89 | 0.9998 |
| tryptophan | 1.32 | 1.22 | 0.9998 |
| tyrosine | 2.33 | 2.63 | 0.3460 |

1. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12:444.
2. diCenzo GC, Benedict AB, Fondi M, Walker GC, Finan TM, Mengoni A, Griffitts JS. 2018. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. PLoS Genet 14:e1007357.
3. Valgepea K, Adamberg K, Seiman A, Vilu R. 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. Mol Biosyst 9:2344–2358.
4. Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010. BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res 38:D750-D753.
5. Orgambide GG, Huang Z-H, Gage DA, Dazzo FB. 1993. Phospholipid and fatty acid compositions of *Rhizobium leguminosarum* biovar *trifolii* ANU843 in relation to flavone-activated pSymnod gene expression. Lipids 28:975–979.
6. Forsberg LS, Carlson RW. 2008. Structural characterization of the primary O-antigenic polysaccharide of the *Rhizobium leguminosarum* 3841 lipopolysaccharide and identification of a new 3-acetimidoylamino-3-deoxyhexuronic acid glycosyl component: a unique O-methylated glycan of uniform. J Biol Chem 283:16037–16050.
7. Breedveld MW, Miller KJ. 1994. Cyclic beta-glucans of members of the family Rhizobiaceae. Microbiol Rev 58:145–161.
8. Lodwig EM. 2001. PhD thesis. University of Reading. Regulation of carbon metabolism in *Rhizobium leguminosarum*.

9. Xavier JC, Patil KR, Rocha I. 2017. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. Metab Eng 39:200-208.

10. Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS. 2011. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. Genome Biol 12:R106.

11. Pini F, East AK, Appia-Ayme C, Tomek J, Karunakaran R, Mendoza-Suárez M, Edwards A, Terpolilli JJ, Roworth J, Downie JA, Poole PS. 2017. Bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion. Plant Physiol 174:1289–1306.

12. Gaworzewska ET, Carlile MJ. 1982. Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. Microbiology 128:1179–1188.

13. Knee EM, Gong F-C, Gao M, Teplitski M, Jones AR, Foxworthy A, Mort AJ, Bauer WD. 2001. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. Mol Plant-Microbe Interact 14:775–784.

14. Rubia MI, Ramachandran VK, Arrese-Igor C, Larrainzar E, Poole PS. 2020. A novel biosensor to monitor proline in pea root exudates and nodules under osmotic stress and recovery. Plant Soil 452:413–422.

## Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

| | |
|---|---|
| Title of Paper | Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum* |
| Publication Status | Submitted for publication |
| Publication Details | C. C. M. Schulte, V. K. Ramachandran, A. Papachristodoulou, P. S. Poole, Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum*. *bioRxiv*, https://doi.org/10.1101/2021.07.28.454262 (2021). |

### Student Confirmation

| | | | |
|---|---|---|---|
| Student Name: | Carolin C. M. Schulte | | |
| Contribution to the Paper | performed model reconstruction and analysis, performed phenotype microarray experiments, wrote manuscript | | |
| Signature | Carolin Schulte | Date | 9th August 2021 |

### Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

| | | |
|---|---|---|
| Supervisor name and title: Professor Philip Poole | | |
| Supervisor comments<br>A highly original contribution to the scientific literature that is currently under review. It is the first such study to integrate multiple datasets and use modelling to understand the metabolic transition of rhizobia. | | |
| Signature | Date | 10th August 2021 |

This completed form should be included in the thesis, at the end of the relevant chapter.

# 4. Metabolic control of nitrogen fixation in rhizobium-legume symbioses

## Summary

After elucidating general metabolic features of Rlv3841 throughout the process of symbiosis formation in Chapter 3, the focus in this chapter is on central features of carbon and nitrogen metabolism in bacteroids. I reconstruct a reduced metabolic model for bacteroids and apply various modelling strategies, notably a recently developed tool for elementary conversion mode enumeration, to investigate the role of experimentally observed storage polymer synthesis and amino acid secretion by bacteroids. I further describe in detail the consequences of dicarboxylate catabolism at the low oxygen levels required for nitrogen fixation. Based on model predictions and experimental data for oxygen uptake rates, $NADH/NAD^+$ ratios and metabolic flux analysis (performed by collaborators), I propose a novel hypothesis explaining why legumes supply bacteroids with dicarboxylates, rather than the abundant sucrose. This work has been published in *Science Advances*[39].

PLANT SCIENCES

# Metabolic control of nitrogen fixation in rhizobium-legume symbioses

Carolin C. M. Schulte[1,2†], Khushboo Borah[1†‡], Rachel M. Wheatley[1§], Jason J. Terpolilli[3‖], Gerhard Saalbach[3], Nick Crang[1¶], Daan H. de Groot[4#], R. George Ratcliffe[1]*, Nicholas J. Kruger[1]*, Antonis Papachristodoulou[2]*, Philip S. Poole[1,3]*

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

**Rhizobia induce nodule formation on legume roots and differentiate into bacteroids, which catabolize plant-derived dicarboxylates to reduce atmospheric $N_2$ into ammonia. Despite the agricultural importance of this symbiosis, the mechanisms that govern carbon and nitrogen allocation in bacteroids and promote ammonia secretion to the plant are largely unknown. Using a metabolic model derived from genome-scale datasets, we show that carbon polymer synthesis and alanine secretion by bacteroids facilitate redox balance in microaerobic nodules. Catabolism of dicarboxylates induces not only a higher oxygen demand but also a higher NADH/$NAD^+$ ratio than sugars. Modeling and $^{13}C$ metabolic flux analysis indicate that oxygen limitation restricts the decarboxylating arm of the tricarboxylic acid cycle, which limits ammonia assimilation into glutamate. By tightly controlling oxygen supply and providing dicarboxylates as the energy and electron source donors for $N_2$ fixation, legumes promote ammonia secretion by bacteroids. This is a defining feature of rhizobium-legume symbioses.**

## INTRODUCTION

Biological nitrogen fixation provides 50 to 70 Tg of bioavailable nitrogen in agricultural systems per year (*1*) and sustains global food security. The most efficient contribution to biologically fixed nitrogen is from symbioses between legumes and rhizobia (*2*), which are soil bacteria that induce formation of nodules on plant roots. Inside nodules, rhizobia differentiate into bacteroids that reduce atmospheric $N_2$ into ammonia for secretion to the plant host in exchange for dicarboxylates, primarily succinate and malate (*3*, *4*). Succinate and malate are typically metabolized via malic enzyme and pyruvate dehydrogenase, yielding acetyl–coenzyme A that can be oxidized in the tricarboxylic acid (TCA) cycle (*5*). Whether bacteroids need a complete TCA cycle remains unclear, as it is essential for *Rhizobium* and *Sinorhizobium* species, but mutants of *Bradyrhizobium japonicum* lacking 2-oxoglutarate dehydrogenase activity achieve wild-type levels of nitrogen fixation on a per-bacteroid basis (*6*, *7*). Flux through the TCA cycle produces reduced nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide ($FADH_2$), which provide electrons for nitrogen fixation and adenosine triphosphate (ATP) generation (*8*). Despite rhizobia requiring oxygen for ATP synthesis, oxygen levels in nodules are only 10 to 40 nM (*9*), which is a requirement for highly oxygen-sensitive nitrogenase enzyme activity (*10*). The challenge of balancing carbon allocation under these conditions is evidenced by the synthesis of lipids and carbon polymers, such as polyhydroxybutyrate (PHB), which indicate imbalances in nutrient supply (*8*, *11*). PHB and lipids have been suggested to play a role as carbon and redox sinks for bacteroids, although their accumulation is variable between rhizobial strains and their role remains to be elucidated (*8*, *12*, *13*).

The defining distinction between nitrogen fixation by rhizobial bacteroids compared to free-living bacteria is the secretion of fixed ammonia to the plant. However, there is no known metabolic mechanism forcing secretion of fixed nitrogen to the plant instead of assimilation by the bacteroid. In addition to the main secretion product ammonia, a substantial portion of fixed nitrogen is apparently secreted in the form of alanine and aspartate (*14*, *15*). At the same time, ammonia assimilation by the glutamine synthetase–glutamine oxoglutarate aminotransferase (GS-GOGAT) system, which is active in free-living rhizobia, is down-regulated during the symbiosis (*16*–*18*).

Because of the complexity of bacteroid metabolism, several studies have used computational approaches to investigate the symbiosis. Notably, metabolic models for various rhizobial species have been reconstructed (*19*–*22*), and some reconstructions have been integrated with models for the host plant (*23*, *24*). Metabolic models describe all enzymatic and transport reactions encoded in an organism's genome and enable the simulation of metabolic flux distributions under defined environmental conditions (*25*, *26*). The most common approach for analyzing metabolic models is flux balance analysis (FBA), where an objective function reflecting the evolutionary goal of the organism under investigation is optimized to calculate a flux distribution (*27*). Maximum cellular growth is the most commonly used objective function and has been found to reproduce experimental results for strains adapted to growth under laboratory conditions (*28*), but this is clearly not applicable in the case of growth-arrested bacteroids. In silico studies of bacteroid metabolism, which have so far mostly applied standard FBA methods, defined an objective reaction comprising nitrogen export to the plant and synthesis of storage compounds (*19*–*22*).

[1]Department of Plant Sciences, University of Oxford, Oxford, UK. [2]Department of Engineering Science, University of Oxford, Oxford, UK. [3]John Innes Centre, Norwich Research Park, Norwich, UK. [4]Systems Biology Lab, AIMMS, Vrije Universiteit Amsterdam, Amsterdam, Netherlands.
*Corresponding author. Email: george.ratcliffe@plants.ox.ac.uk (R.G.R.); nick.kruger@plants.ox.ac.uk (N.J.K.); antonis@eng.ox.ac.uk (A.P.); philip.poole@plants.ox.ac.uk (P.S.P.)
†These authors contributed equally to this work.
‡Present address: Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK.
§Present address: Department of Zoology, University of Oxford, Oxford, UK.
‖Present address: Centre for Rhizobium Studies, Murdoch University, Murdoch, Australia.
¶Present address: SciLifeLab, Kungliga Tekniska Högskolan, Solna, Sweden.
#Present address: Biozentrum, University of Basel, Basel, Switzerland.

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

The result of FBA calculations is a single flux distribution, which is often not a unique solution for the optimization problem (*27*). In contrast, methods such as elementary flux mode enumeration describe all minimal sets of steady-state fluxes through the metabolic network (*29*). While this approach is more computationally expensive than FBA and currently not feasible for genome-scale metabolic networks (*30*), it provides a more comprehensive and unbiased description of metabolism that does not rely on an artificially defined objective. Addressing the problem of combinatorial explosion during elementary flux mode enumeration, elementary conversion modes (ECMs) have been proposed as an alternative approach to compute all possible stoichiometries between input and output metabolites (*31*). While ECMs do not provide information on the underlying metabolic pathways for a specific conversion, they can still capture all metabolic capabilities of an organism. The feasibility of ECM enumeration for genome-scale metabolic networks when focusing on subsets of metabolites has recently been demonstrated (*32*).

Despite decades of research efforts, a comprehensive view of the links between central carbon and nitrogen metabolism in bacteroids is missing. Most experimental studies have focused on individual metabolic pathways, and computational models have used artificial objective functions that may not be relevant in the natural system.

In this study, we combine experimental and metabolic modeling approaches to explain fundamental features of bacteroid metabolism. Using proteome, transcriptome, and gene essentiality data, we reconstruct a model of metabolic pathways active during nitrogen fixation in *Rhizobium leguminosarum* bv. *viciae*. We implement modeling strategies that circumvent the limitations of traditional FBA to explain the importance of the experimentally observed storage polymer synthesis and amino acid secretion in bacteroids. We further investigate the role of the TCA cycle during nitrogen fixation and validate model predictions by $^{13}C$ metabolic flux analysis of *Azorhizobium caulinodans*. Our model provides insights into the fundamental constraints on rhizobial metabolism during symbiotic nitrogen fixation. An improved understanding of metabolic processes in bacteroids is of central importance for ongoing efforts in optimizing existing symbioses and engineering novel plant-microbe interactions for sustainable agriculture (*2*, *33*).

## RESULTS

### Data-based reconstruction of a bacteroid metabolic model

We reconstructed a metabolic model for pea bacteroids of *R. leguminosarum* bv. *viciae* 3841 using bacteroid-specific experimental data. First, we quantified the proteome of unlabeled bacteroids relative to $^{15}N$-labeled free-living bacteria (Fig. 1, figs. S1 and S2, Supplementary Text, and data S1). In addition, genes up-regulated in transcriptional datasets of bacteroids (fig. S3) (*34*) and genes identified as specifically essential for symbiosis formation by insertion sequencing (INSeq) (*35*) were included in the model. The final core metabolic network named *i*CS323 contained 323 genes, 237 metabolites, and 299 reactions, 207 of which are metabolic (excluding transport, demand, and sink reactions) (table S1 and data S2 to S4). Of the 207 metabolic reactions, 177 (86%) are supported by experimental evidence from at least one of the bacteroid-specific datasets. *i*CS323 was evaluated using MEMOTE (*36*), confirming the absence of stoichiometrically balanced cycles, orphan metabolites, and dead-end reactions (data S5).

The main pathways in the final model were central carbon metabolism (TCA cycle, gluconeogenesis, and pentose phosphate pathway), amino acid metabolism, and carbon polymer synthesis. As an initial validation, malate, succinate, and oxygen uptake were constrained according to the flux boundaries defined in a modeling study of *Sinorhizobium meliloti* (*21*), and standard FBA maximizing nitrogenase activity was performed. The obtained flux distribution captured key features of bacteroid metabolism, including use of the TCA cycle, pyruvate synthesis via malic enzyme, and minor activity of gluconeogenesis (fig. S4) (*5*, *8*, *37*).

When comparing gene essentiality predictions to mutant phenotypes identified by INSeq (*35*), we found that 280 (87%) of the genes in the model had an in silico phenotype that agreed with the experimental evidence (data S6). Genes were defined as essential in silico if their deletion prevented flux through the nitrogenase reaction in the model. Of the 31 false negatives (in silico nonessential genes that were essential according to INSeq), 9 were involved in sugar metabolism. In agreement with the results of a recent modeling study of *S. meliloti* (*24*), this may suggest that rhizobia differentiating into bacteroids have access to sugars as a carbon source, whereas nitrogen-fixing bacteroids do not. In addition, four genes involved in serine metabolism were incorrectly predicted to be

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

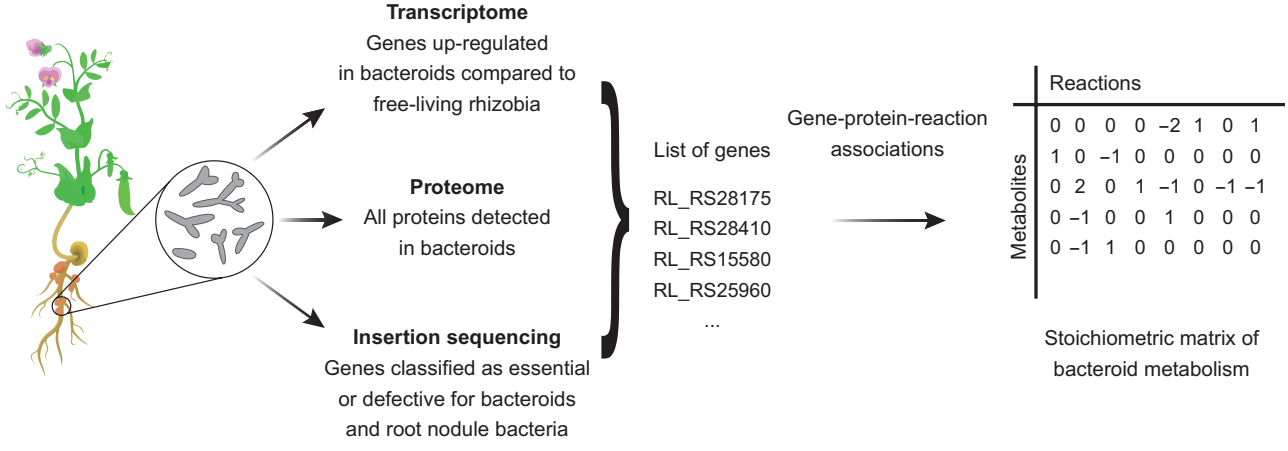



**Fig. 1. Workflow for metabolic model reconstruction.** A metabolic model of *R. leguminosarum* bacteroids was reconstructed using transcriptome, proteome, and gene essentiality data.

nonessential. A possible explanation is the role of serine as a precursor for cysteine biosynthesis. Cysteine plays a role in the synthesis of iron-sulfur clusters for the nitrogenase enzyme (*38*), which is not explicitly accounted for in the model. Pyruvate kinase (PykA) was further predicted to be nonessential in disagreement with the INSeq data. A *pykA* mutant had higher nitrogenase activity than the wild type in plants harvested after 28 days (*5*), indicating that the INSeq phenotype may be a result of developmental delay and the model correctly identifies the gene as nonessential.

Overall, our model shows good predictive quality for gene essentiality. It is important to note that the perfect agreement of in silico predictions with the experimental data is not expected. This is due to our model being limited to central metabolic pathways in bacteroids, neglecting, for example, the synthesis of nucleotides for DNA replication. In addition, genes that are determined to be essential for bacteroids by INSeq may actually cause a growth defect at earlier stages of symbiosis formation and are not necessarily required for nitrogen fixation itself.

### Characterization of bacteroid metabolism using ECMs

Previously published modeling studies of bacteroid metabolism used FBA and maximized a lumped objective reaction comprising ammonia and amino acid export as well as storage polymer synthesis (*19*–*22*). While this approach constrains flux distributions to reflect experimentally observed phenotypes, it also introduces an artificial stoichiometric coupling between the objective metabolites, which precludes the investigation of changes in carbon and nitrogen allocation depending on nutrient availability. To characterize bacteroid metabolism with a minimum number of preset assumptions, we used ecmtool (*32*) to enumerate ECMs of the metabolic network (Fig. 2 and data S7). In the context of ECMs, the overall transformation of nutrients into secreted metabolites and cell components is called a conversion. The set of possible conversions is determined by an organism's metabolism, which is described by a metabolic model. Enumerating an exhaustive set of minimal conversions (ECMs) thus provides a complete overview of all metabolic capabilities in terms of input-output stoichiometry. This approach gives a comprehensive description of metabolism without assuming optimality with respect to a specific objective.

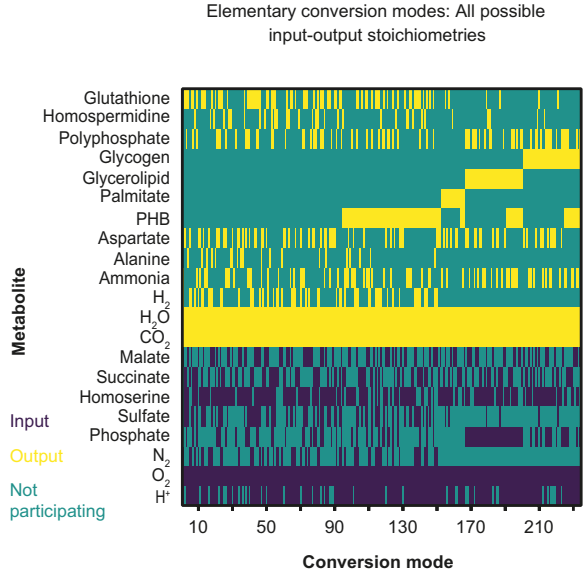



**Fig. 2. ECMs for *i*CS323.** The heatmap represents conversion modes calculated with succinate and malate as carbon sources and carbon polymers (palmitate, PHB, glycerolipid, and glycogen), ammonia, alanine, and aspartate as outputs. Flux values were indexed to represent inputs, outputs, and nonparticipating compounds.

Conversion modes were found for all expected products (ammonia, alanine, aspartate, and storage compounds) with a minimal set of inputs, mainly malate/succinate, oxygen, and $N_2$ (data S8). The presence of conversion modes for storage polymer synthesis without nitrogenase activity suggests that the two processes are not inherently linked, which agrees with reports of storage polymer accumulation before the onset of nitrogen fixation (*12*, *39*). Potential co-catabolism of an amino acid in addition to malate/succinate was investigated by analyzing conversion modes with each of the proteinogenic amino acids or γ-aminobutyric acid (GABA) as an additional input. Conversion modes were found for arginine, cysteine, GABA, glutamine, glutamate, glycine, serine, and threonine. All conversions involving arginine, cysteine, or serine had a negative net nitrogen output and were therefore considered unlikely to be biologically relevant. For the remaining conversion modes with a positive net nitrogen output, no benefit was found for any amino acid in terms of oxygen demand or carbon cost (fig. S5). We therefore limited our analysis to conversion modes using GABA, which is known to be metabolized in pea bacteroids (*40*).

Oxygen demand per carbon uptake was decreased for all conversion modes that produced storage polymers compared to conversions that did not (Fig. 3A and fig. S6). Carbon storage polymers thus function as carbon and redox sinks under oxygen-limiting conditions, enabling electrons to be partitioned to polymer synthesis rather than oxygen as a terminal electron acceptor. The oxygen uptake relative to fixed $N_2$ was increased for conversion modes generating glycogen or lipids (Fig. 3B), consistent with the ATP requirement for producing these storage compounds, which would add to the ATP demand of the nitrogenase reaction. In addition, conversion modes generating carbon polymers increased the carbon cost per nitrogen supplied to the plant (Fig. 3C). This aligns with observations that plant nodule cells accumulate starch when occupied by bacteroid glycogen synthase mutants (*12*); i.e., the plant has excess carbon when bacteroid polymer synthesis is restricted.

The carbon cost determined for conversion modes without polymer production is similar to the theoretical cost of nitrogen fixation (2.5 g of carbon per gram of nitrogen), whereas that for PHB- and some palmitate- and glycerolipid-producing conversion modes is close to the experimentally observed value (8 g of carbon per gram of nitrogen) (*2*). Given that the plant host regulates the supply of nutrients such as carbon and oxygen in response to nitrogen output (*41*, *42*), this would potentially limit excess diversion of carbon into polymers.

Conversions without carbon polymer production could be subgrouped into those generating alanine and/or aspartate and those only producing ammonia, where amino acid secretion reduced oxygen demand per carbon uptake (Fig. 3A) and oxygen demand per nitrogen output (fig. S6). Alanine dehydrogenase catalyzes the NADH-dependent synthesis of alanine from pyruvate and ammonia, making this pathway an oxygen-independent carbon sink for $NAD^+$ regeneration as well as a sink for carbon and protons. Synthesis of alanine in particular enables bacteroids to maintain nitrogen export to the plant in a low-oxygen environment and explains the mixed secretion of ammonia and amino acids observed experimentally (*14*).

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

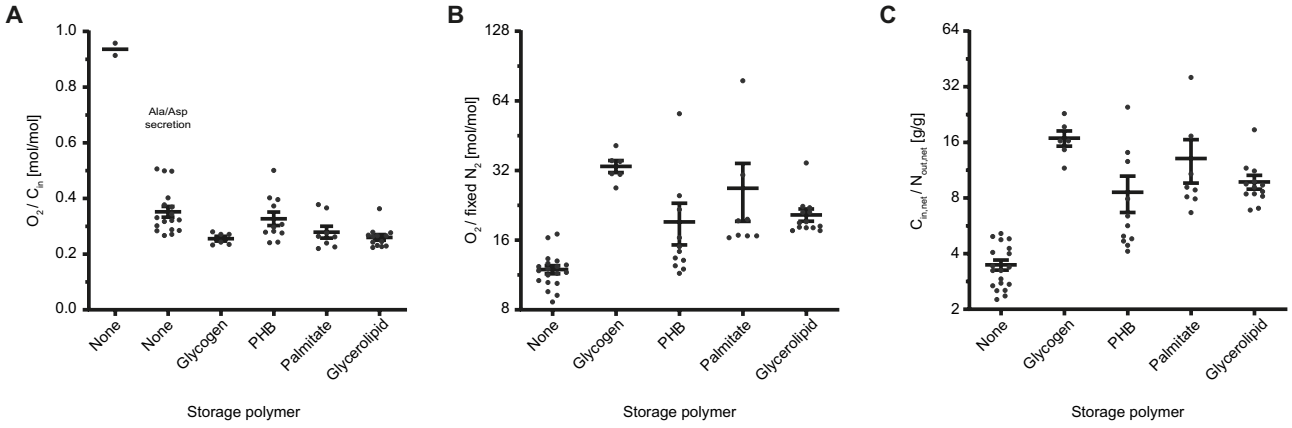



**Fig. 3. Effect of polymer synthesis and amino acid secretion on oxygen demand and carbon cost of nitrogen fixation.** (**A**) Oxygen uptake per carbon uptake, (**B**) oxygen uptake per fixed $N_2$, and (**C**) carbon cost (difference of carbon input and output) per nitrogen secreted (difference of nitrogen input and output) were determined for conversion modes using succinate, malate, and GABA as carbon sources. In (A), ECMs without storage polymer production have been separated into those secreting only ammonia and those secreting alanine and/or aspartate in addition to/instead of ammonia. Each data point represents an individual conversion mode, and lines and bars indicate means ± SEM.

Amino acid secretion could be favored over polymer synthesis under certain conditions because it allows for removal of carbon from the bacteroid rather than intracellular accumulation. The same trends for storage polymer synthesis and amino acid secretion were observed for ECMs of a model for *Sinorhizobium fredii* bacteroids (fig. S7) (*19*), indicating that these principles are likely to govern symbiotic metabolism across rhizobial species.

## Role of oxygen limitation in shaping bacteroid metabolism

We next characterized the network response to different carbon and oxygen availability when optimum nitrogenase activity is maintained. Because uptake fluxes of bacteroids are difficult to determine experimentally, we performed phenotype phase plane analysis (*43*). Avoiding artificial biases of carbon and nitrogen allocation, nitrogenase activity was evaluated without maximizing the production of storage compounds. Four feasible regions with distinct metabolic behavior were identified, with phase I characterized by carbon limitation and increasing oxygen limitation from phase II to most limited in phase IV (Fig. 4A and fig. S8).

Flux variability analysis showed that with increasing oxygen limitation, flux through TCA cycle enzymes, as represented by citrate synthase, decreased (Fig. 4B). Carbon was increasingly channeled into pyruvate, which caused accumulation of PHB and lipids as well as alanine production, with all nitrogen being secreted in the form of alanine instead of ammonia in phase IV. The tendency to increase alanine secretion at low oxygen levels has recently been shown experimentally in *B. japonicum* (*44*). Alanine secretion is thus important to sustain bacteroid metabolism under oxygen limitation, which probably occurs in the natural system (*45*). This is consistent with a 20% reduction in dry weight of peas inoculated with alanine dehydrogenase mutants of *R. leguminosarum* that no longer secrete alanine (*46*).

Independently of maximum nitrogenase activity, general network properties were investigated using ensemble-evolutionary FBA (*47*). The results supported increased PHB and lipid synthesis under oxygen-limiting conditions and decreased activity of the TCA cycle (Fig. 4, C and D). The trend for increased alanine synthesis was not observed, indicating that it is linked to nitrogen fixation rather than being an inherent network property. The identified shifts in metabolic behavior support the role of alanine, PHB, and lipids as sinks for carbon when oxygen is limiting, and usage of the TCA cycle becomes disadvantageous owing to the accumulation of reduced electron carriers (*8*), which agrees with the conversion mode analysis.

## Metabolic constraints on ammonia assimilation by bacteroids

The predicted down-regulation of the TCA cycle would reduce the availability of 2-oxoglutarate required by the GS-GOGAT pathway, which is the sole pathway for ammonia assimilation coupled to growth in rhizobia. It should be noted that rhizobial GS-GOGAT mutants cannot grow on ammonia as a nitrogen source, and alanine dehydrogenase, for example, cannot substitute for GS-GOGAT (*18*). Increased ammonia assimilation into glutamate by bacteroids would require increased TCA cycle activity and hence increase oxygen demand as indicated by a significant positive correlation between oxygen uptake and GS activity in the conversion mode analysis (Fig. 5A). To assess the impact of ammonia assimilation on the metabolic fluxes in bacteroids, we forced flux through a demand reaction for glutamate in the model. In this scenario, flux through the decarboxylating arm of the TCA cycle had to be maintained under oxygen limitation to supply 2-oxoglutarate for ammonia assimilation, which caused an increased oxygen demand (Fig. 5B).

Our modeling results indicate the importance of the carbon source provided to bacteroids to enforce use of the TCA cycle as the main catabolic pathway. It is currently unclear why plants provide bacteroids with $C_4$-dicarboxylates, especially considering that photosynthate is transported to nodules as sucrose (*48*). We therefore compared the effects of using sucrose instead of malate as a carbon source in silico. For a given carbon uptake rate, the model predicted higher nitrogenase activity for sucrose compared to malate (Fig. 5C) in agreement with an integrated plant-bacteroid model for *S. meliloti* (*24*). Furthermore, less oxygen was needed for sucrose (Fig. 5B) (or glucose; fig. S9) catabolism, which is in accordance with experimentally determined oxygen uptake rates of free-living *R. leguminosarum* (Fig. 6A). Catabolism of arabinose, a sugar metabolized via 2-oxoglutarate, induced an oxygen demand similar to growth on

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

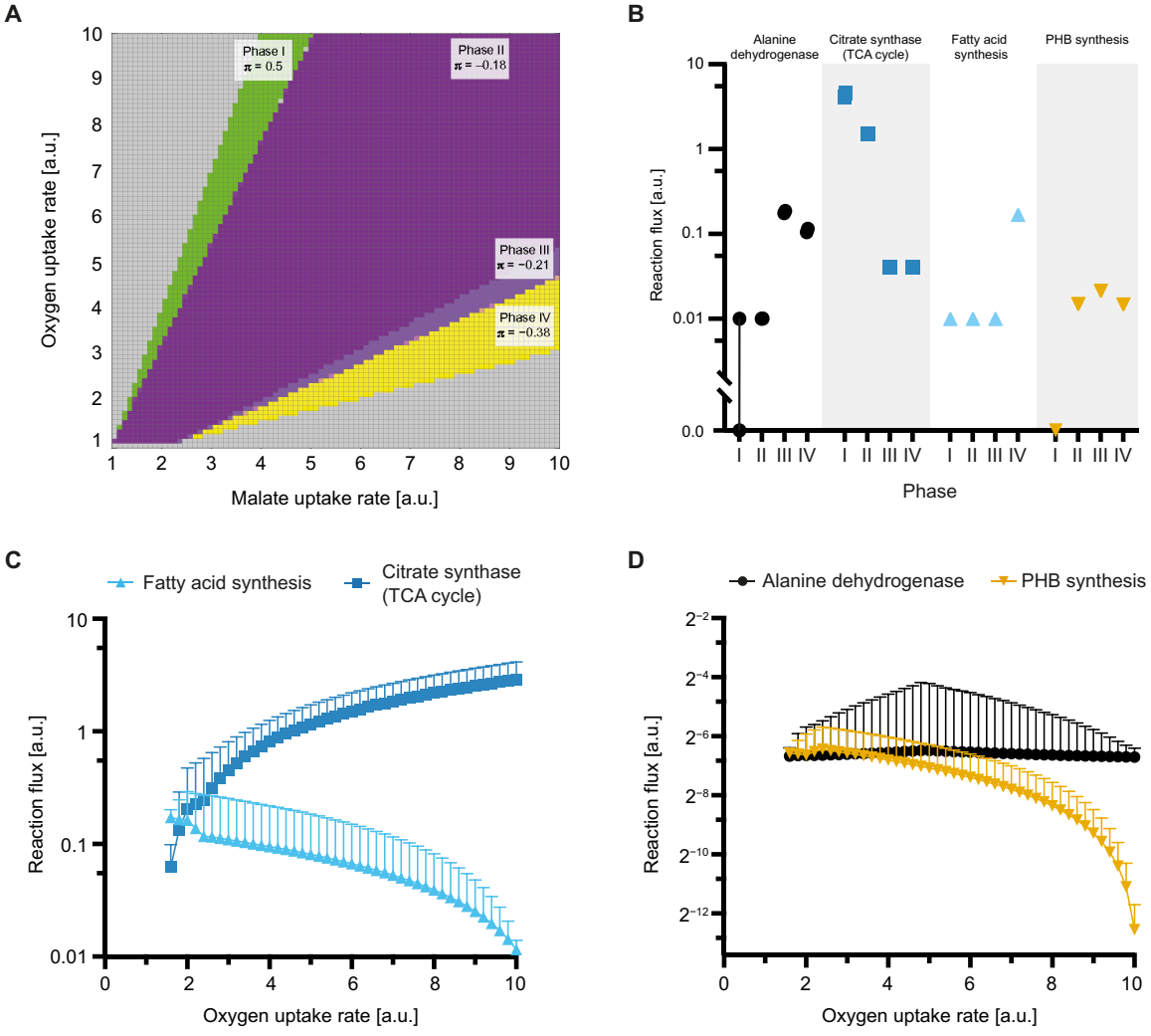



**Fig. 4. Metabolic response of bacteroids to varying carbon and oxygen availability.** Phenotype phase plane analysis with varying malate and oxygen uptake rates was performed for *i*CS323. (**A**) Shadow prices ($\pi$) for oxygen in the four phases defined by phenotype phase plane analysis. (**B**) Flux ranges for alanine dehydrogenase, citrate synthase, fatty acid, and PHB synthesis fluxes determined by flux variability analysis for the four phases defined in (A). Symbols indicate upper and lower bounds of the indicated flux for maximum nitrogenase activity. Note that the symbols for upper and lower bounds are overlapping in most cases. (**C**) Fatty acid synthesis and citrate synthase activity at a fixed malate uptake rate of 4 flux units and varying oxygen uptake rates were determined by ensemble-evolutionary FBA. (**D**) Same as (C) for PHB synthesis and alanine dehydrogenase. Values plotted in (C) and (D) represent mean values and standard deviation for 50,000 random objective functions. a.u., arbitrary units.

succinate. This suggests that catabolism of TCA cycle intermediates generally creates a high oxygen demand and thus causes a more severe growth impairment compared to glucose under low oxygen conditions (Fig. 6C). However, the NADH/$NAD^+$ ratio was similar for growth on glucose and arabinose but significantly higher for succinate catabolism (Fig. 6B). This could be explained not only by 2-oxoglutarate supply from arabinose catabolism partly obviating use of the decarboxylating TCA cycle arm but also by differences in the regulation of sugar versus dicarboxylate uptake and metabolism.

Overall, these findings indicate that supply of dicarboxylates such as succinate to bacteroids creates both a high oxygen demand and a highly reduced redox ratio, which is important for supply of electrons to the nitrogenase enzyme (*8*). The absolute value of the redox ratio did not change significantly for different oxygen concentrations and was also not correlated with the growth rate (Fig. 6C). A slightly higher NADH/$NAD^+$ ratio was measured for cultures grown on succinate at 21% oxygen compared to 1% oxygen. However, this is most likely due to minor effects of sample preparation, which is very time-sensitive in the case of fast-growing cultures with high respiration rates. Thus, the redox state of *R. leguminosarum* is mainly determined by the carbon source, implying that the nature and quantity of the supplied carbon source are of central importance for defining metabolic fluxes in bacteroids, with succinate providing a higher redox value than glucose. Plants provide dicarboxylates as a carbon source to bacteroids although they are less efficient at supporting nitrogen fixation and increase oxygen demand relative to sucrose in the oxygen-limited nodule.

## Metabolic flux analysis of *A. caulinodans* supports model predictions

To validate core findings of our model predictions experimentally, we performed $^{13}C$ metabolic flux analysis of *A. caulinodans* ORS571.

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

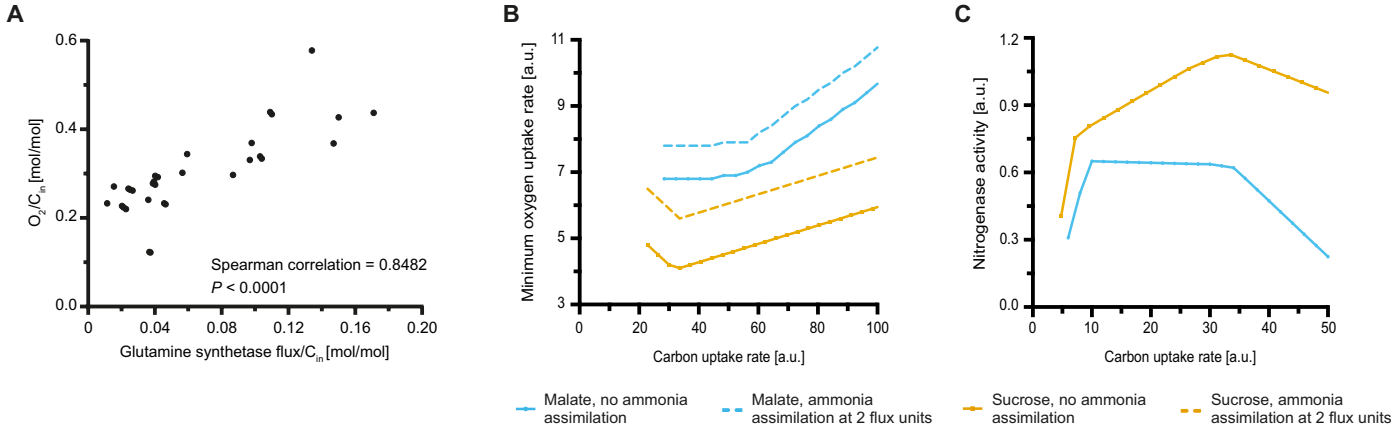



**Fig. 5. Predicted impact of ammonia assimilation on oxygen demand.** (**A**) Scatterplot showing the relationship between oxygen demand and GS activity per carbon uptake predicted by ECM analysis. Each point represents an individual conversion mode, only conversion modes with GS activity are shown. (**B**) Predicted minimum oxygen demand for malate or sucrose as a carbon source without (solid line) and with (dashed line) ammonia assimilation by bacteroids. Oxygen demand is shown for equivalent nitrogenase activity for the respective carbon source. (**C**) Predicted maximum nitrogenase activity for malate or sucrose as a carbon source and a maximum oxygen uptake rate of 4 flux units. Values in (B) and (C) are shown per mole of carbon atoms.

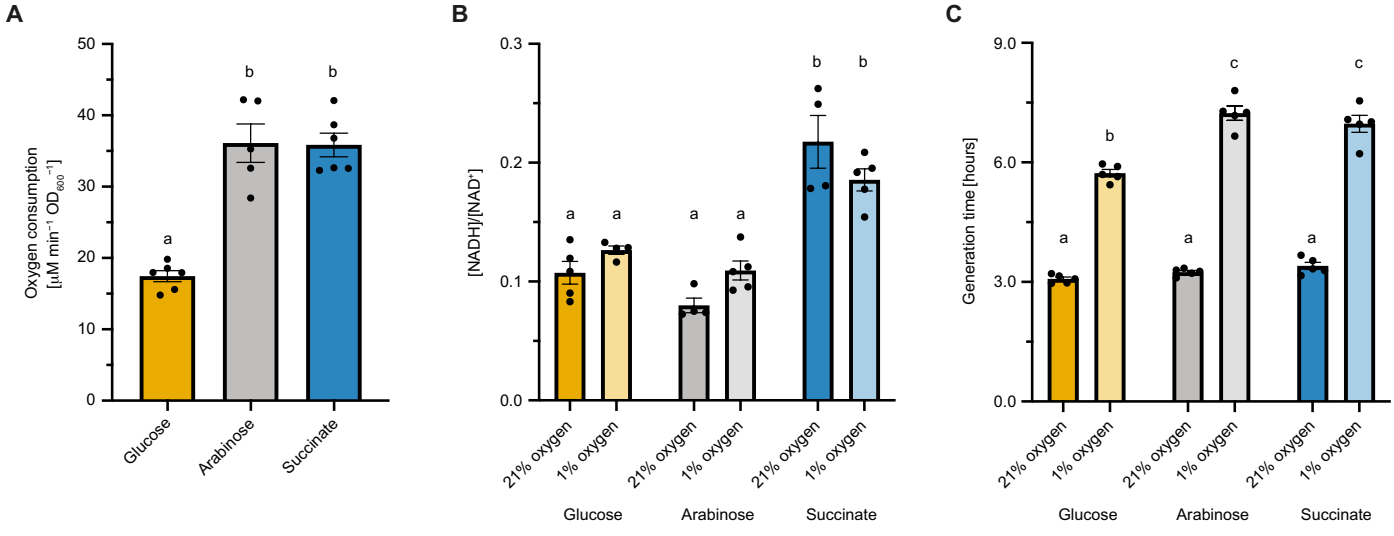



**Fig. 6. Carbon source determines oxygen demand and redox ratio in *R. leguminosarum*.** (**A**) Experimentally determined oxygen consumption rate of *R. leguminosarum* grown in minimal media with $NH_4Cl$ as a nitrogen source and glucose, arabinose, or succinate as the sole carbon source. (**B**) NADH/$NAD^+$ ratio and (**C**) generation time were measured for cultures grown at 21% oxygen and 1% oxygen. Data points represent independent biological replicates with lines and error bars indicating means ± SEM. Lowercase letters indicate significant differences between samples determined by one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons test.

We chose this rhizobial strain for its ability to perform nitrogen fixation in both free-living conditions and in symbiosis with *Sesbania rostrata*. When comparing non-diazotrophic growth of *A. caulinodans* at different oxygen levels, PHB synthesis increased threefold under microaerobic (3% oxygen) conditions compared to aerobic growth and further increased in bacteroids (Fig. 7 and data S9). This agrees with the predicted importance of storage polymer synthesis for balancing carbon allocation under oxygen-limited conditions.

TCA cycle fluxes increased during free-living diazotrophic growth compared to non-diazotrophic growth, which is consistent with an increased demand for reductant for nitrogen fixation as well as for cell growth. However, an overall down-regulation of the TCA cycle was observed when comparing bacteroid metabolism to non-diazotrophic growth under microaerobic conditions. The highest relative down-regulation was observed for the decarboxylating arm of the TCA cycle between citrate and succinate. As predicted by our model, bacteroids would thus have limited capability for 2-oxoglutarate synthesis and would consequently be restricted for ammonia assimilation into glutamate. In agreement with these findings, proteome and transcriptome data showed a strong down-regulation of isocitrate dehydrogenase and the GS-GOGAT system in bacteroids (figs. S2 and S3).

## DISCUSSION

In this study, we combined experimental and computational methods to provide explanations for fundamental properties of bacteroid

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

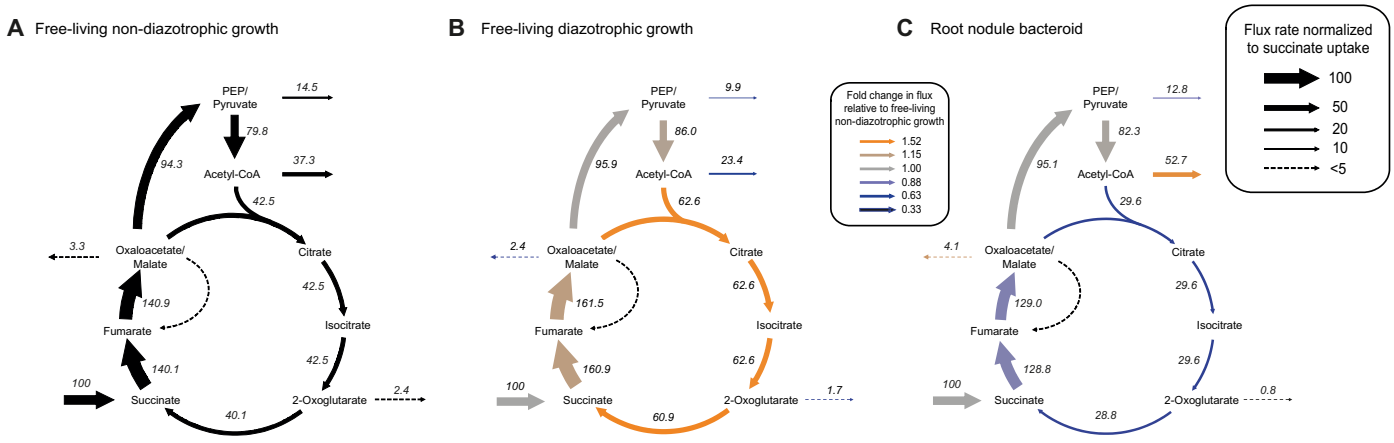



**Fig. 7. Metabolic flux analysis of *A. caulinodans*.** Labeling experiments were performed with [$^{13}C_4$]-succinate for (**A**) *A. caulinodans* grown in free-living conditions at 3% oxygen with $NH_4Cl$ as a nitrogen source, (**B**) *A. caulinodans* grown in free-living diazotrophic conditions, and (**C**) bacteroids isolated from root nodules of *S. rostrata*. Flux values were normalized to a succinate uptake rate of 100 units. Colors in (B) and (C) indicate fold changes in flux values compared to (A). Note that the flux from succinate to fumarate is the sum of the succinate uptake rate and the flux through the decarboxylating arm of the TCA cycle.

metabolism, which have so far mostly been investigated separately. Our modeling results indicate that oxygen limitation is the driving factor behind the observed synthesis of carbon polymers (*8*, *12*) and induces the secretion of alanine, in addition to ammonia. The level of alanine secretion is dependent on the ratio of carbon and oxygen supply, which agrees with experimental studies (*44*). Rhizobial nitrogen fixation is fueled by dicarboxylates (*3*), requires a low-oxygen environment to protect nitrogenase (*10*), and involves down-regulation of ammonia assimilation into glutamate in bacteroids (*18*). Both model predictions and metabolic flux analysis of *A. caulinodans* bacteroids indicate that these factors are interconnected: Catabolism of dicarboxylates induces a highly reduced redox poise and creates a high oxygen demand. In the low-oxygen environment of the nodule, this promotes down-regulation of the decarboxylating TCA cycle arm, which would decrease ammonia assimilation into glutamate by bacteroids. This agrees with glutamate levels being 20-fold lower in bacteroids compared to free-living rhizobia (*8*). In addition to promoting ammonia secretion to the plant, limited ammonia assimilation into glutamate would contribute to the growth arrest of bacteroids. As glutamate is the transamination donor for most other amino acids, this could explain the dependence of bacteroids on supply of amino acids by the plant (*49*, *50*). Intriguingly, the increased oxygen stress on bacteroids and their limited ability to assimilate ammonia into glutamate will force ammonia secretion and is therefore an important mechanism for enforcing a mutualistic relationship with the host plant. Furthermore, the ample supply of reducing equivalents would then be available to satisfy the high electron demand of the nitrogenase reaction, while carbon polymer synthesis and secretion of alanine, in particular, balance the allocation of carbon and regeneration of electron carriers. High oxygen consumption rates induced by dicarboxylate catabolism also contribute to lowering local oxygen levels in bacteroids, which would additionally protect the oxygen-sensitive nitrogenase enzyme.

The proposed metabolic principles provide a general framework for understanding the constraints on bacteroid metabolism that favor ammonia (and amino acid) secretion and contribute to the growth arrest of bacteroids. However, rhizobium-legume symbioses are highly evolved mutualisms. As a result, many factors, such as nodule cysteine-rich peptides, play a role in inducing terminal differentiation of some bacteroids (*51*) and ammonia secretion by bacteroids is further forced by transcriptional regulation of genes involved in nitrogen metabolism (e.g., Ntr) (*52*).

Previous modeling studies of intermicrobial interactions found anoxic conditions to promote mutualistic interactions and increase the diversity of secreted metabolites (*53*, *54*), and oxygen availability was determined to be a better indicator of secreted metabolites than species identity (*54*). It is therefore possible that metabolic principles similar to those outlined in this study have a broader significance in governing metabolite exchanges in interspecies interactions.

## MATERIALS AND METHODS

### General modeling procedures

Standard FBA calculations were performed in MATLAB R2019b (Mathworks) using scripts from the COBRA Toolbox v3.0 (*55*) and the Gurobi 8.0.1 solver (www.gurobi.com). When using the optimizeCbModel function, the Taxicab norm was minimized to avoid loops in the calculated flux distributions. The functions phenotypePhasePlane and fluxVariability were used for phenotype phase plane analysis and flux variability analysis, respectively. For flux variability analysis, loopless solutions were calculated that allowed for maximum nitrogenase activity. In silico gene deletion analysis was performed with the function singleGeneDeletion using the FBA method. A gene was considered essential when its deletion prevented flux through the nitrogenase reaction (data S6). All MATLAB scripts are available on Zenodo (https://doi.org/10.5281/zenodo.4911594). Flux maps were created using Escher (*56*).

### ECM analysis

ECMs were calculated using ecmtool (https://github.com/tjclement/ecmtool) (*32*). Boundary conditions for calculating conversion modes were based on experimental studies and are detailed in data S8. To limit our analysis to biologically meaningful scenarios, only conversion modes with a positive cost lower than 40 g of carbon per gram of nitrogen were considered. We further restricted our analysis to conversion modes using one amino acid at most as an input. The

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

full set of conversion modes calculated with succinate, malate, and amino acids as inputs is available in data S7. For determining the correlation of oxygen demand and GS activity, glutamate was set as an additional output and a virtual metabolite was added to the GS reaction to track the flux through this reaction. ECM analysis for *i*CC541, a metabolic model of *S. fredii* bacteroids (*19*), was performed according to similar principles. To decrease computation time, cofactors were hidden in this analysis (data S8).

### Statistical analysis

Nonparametric Spearman correlation and analysis of variance (ANOVA) followed by Tukey's multiple comparisons test were performed in GraphPad Prism 8.4.3. $P < 0.05$ was considered statistically significant.

### Flux balance analysis

Constraints for FBA-based computations were defined similar to those for conversion mode analysis. For all proteinogenic amino acids except for asparagine, alanine, and aspartate, demand fluxes were constrained to values between 0.01 and 0.05 to mimic low levels of protein synthesis in bacteroids. The upper bound for alanine and aspartate demand reactions was left unconstrained because those amino acids are secreted by bacteroids. All constraints are detailed in data S8.

Ensemble-evolutionary FBA was performed as previously described (*47*, *57*). Briefly, 50,000 objective functions containing a random number of model reactions that were assigned random weights were generated (see fig. S10 for determination of ensemble size). Flux distributions for all objective functions were then calculated using FBA for a fixed malate uptake rate of 4 flux units and varying oxygen uptake rate. Constraints for ensemble-evolutionary FBA were defined as for standard FBA, but with a minimum flux of 0.01 for the nitrogenase reaction and no upper bound on all amino acid demand reactions.

### Model reconstruction

A database of gene-protein-reaction associations was derived from genome annotation of *R. leguminosarum* bv. *viciae* 3841 (Rlv3841). First, an orthology-based reconstruction was obtained using AuReMe (*58*). Because both a genome-scale model [*i*GD1575b (*21*, *59*)] and a highly curated core model [*i*GD726 (*59*)] for the rhizobial strain *S. meliloti* 1021 are available, these were used as templates. Reconstruction in AuReMe was performed for both *S. meliloti* models, resulting in draft networks containing 1304 and 613 reactions for *i*GD1575b and *i*GD726, respectively. Where the same reaction occurred in both reconstructions, the gene-protein-reaction association from the *i*GD726-based reconstruction was selected as it can be expected to be more accurate. To account for gene functions annotated in the Rlv3841 genome but not present in *S. meliloti*, a second draft metabolic model was obtained from KBase (*60*). Briefly, the Rlv3841 genome was annotated with Prokka (v1.12) and the "Build Metabolic Model" function was run without gap-filling. This generated a network with 607 reactions, 98 of which were not contained in the reaction list obtained from the AuReMe reconstructions. A model of bacteroid metabolism in Rlv3841 was then built by manually selecting reactions from the draft reconstructions that are catalyzed by enzymes detected in the bacteroid proteome, as well as those associated with up-regulated (*34*) and essential genes for bacteroids (*35*). Gene-protein-reaction associations were refined by comparison to gene essentiality data for free-living Rlv3841 (*61*). Gaps were filled on the basis of literature evidence and guided by the Kyoto Encyclopedia of Genes and Genomes database (*62*). Because the focus of this study was on central metabolic pathways, biosynthesis of some cofactors was excluded and GTP (guanosine triphosphate) was replaced with ATP where applicable. Some linear pathways, namely, most reactions associated with lipid biosynthesis and *myo*-inositol catabolism, were summarized in lumped reactions. Further details on pathways included in the model and the definition of exchange fluxes are provided in the Supplementary Text.

### Bacterial strains and culture conditions

Oxygen consumption and NADH/$NAD^+$ ratios were measured for Rlv3841. Cultures were grown in universal minimal salts (UMS) (*61*) or acid minimal salts (AMS) (*63*) medium at 28°C. UMS was supplemented with different carbon and nitrogen sources at the following final concentrations: succinate, 20 mM; arabinose, 20 mM; glucose, 10 mM; and $NH_4Cl$, 10 mM. Cultivations were performed in 250-ml Erlenmeyer flasks with an initial filling volume of 50 ml. Low-oxygen cultivations were performed in a glove box (Belle Technology) with the atmosphere adjusted to the desired oxygen concentration by flushing with nitrogen gas.

*A. caulinodans* ORS571 was grown in UMS medium supplemented with 10 mM succinate and 0.3 mM nicotinic acid at 37°C. For diazotrophic growth, cultures were continuously sparged with a gas mixture containing 97% $N_2$ and 3% $O_2$, which had been determined to be the optimal oxygen level for diazotrophic growth.

### Measurement of NADH/$NAD^+$ ratios

NADH/$NAD^+$ ratios were determined using the NAD/NADH-Glo Assay (Promega) according to the manufacturer's instructions. Cells were harvested during exponential phase [$OD_{600}$ (optical density at 600 nm) = 0.4 to 0.6] for all measurements.

### Measurement of oxygen consumption rates

Oxygen consumption rates were measured for liquid cultures in early exponential phase. Cultures were diluted to $OD_{600}$ ~ 0.15, and a 25-ml glass vial containing an OxyDot was filled to the top with liquid culture and sealed. Measurements of oxygen concentration were taken every 15 s using the OxySense 325i system while the culture was stirred with a magnetic stirrer. The data were analyzed with the OxySense Gen III software and oxygen consumption was calculated from measurements obtained between 18 and 15% oxygen concentration.

### Proteomics

Bacteroids were obtained from nodules of pea (*Pisum sativum* cv. Avola) plants inoculated with Rlv3841 and grown in an illuminated environment-controlled growth room as previously described (*8*). Plants were harvested at 28 days after inoculation, with bacteroids extracted from excised root nodules by double Percoll gradient purification (*14*). Bacteroids were isolated from nodules obtained from a total of 21 plants and processed as three independent replicates.

For the free-living cultures, Rlv3841 was grown in AMS medium with 20 mM succinate and 10 mM $^{15}NH_4Cl$. Cultures were grown to late log phase ($OD_{600}$ ~ 0.8) at 28°C on a gyratory shaker at 250 rpm and subcultured into fresh AMS six times, as preliminary trials showed that this yielded >99% $^{15}N$ incorporation into cell proteins.

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

From the sixth subculture, three separate AMS cultures were inoculated and harvested at mid-log phase ($OD_{600}$ ~ 0.4). Cells were harvested by centrifugation, washed in AMS, and stored at −80°C. Bacteroid and cell pellets were resuspended separately in 10 mM HEPES (pH 7.2) buffer and an aliquot of each was taken and lysed by two rounds on a FastPrep FP120 Ribolyser (Bio 101/Savant) at a setting of 6.5 for 30 s, with samples kept on ice for 5 min between each round. The protein content of the extracted aliquots was then determined by Bradford assay, where bovine serum albumin (Pierce) was used to generate a standard curve. These values were then used to mix equivalent proportions of unextracted bacteroid and cells from the original suspensions to yield at least 200 μg/ml of total combined protein. Mixed bacteroid and cell samples were then ribolyzed as described above, another Bradford determination was performed to confirm the protein concentration, and then the equivalent of 50 μg of protein was extracted by methanol-chloroform-water precipitation (*64*).

The protein pellets from the three replicates were dissolved in SDS-gel sample loading buffer, heated at 80°C for 10 min, and loaded onto a Novex gel (10% bis-tris SDS gel, Life Technologies, Carlsbad, CA). After separation and staining with InstantBlue (Expedeon, Harston, UK), the gel lanes were cut into 12 to 15 slices that were washed, reduced and alkylated, and treated with trypsin according to standard procedures. After digestion, peptides were extracted with 5% formic acid/50% acetonitrile, dried, and redissolved in 0.1% trifluoroacetic acid.

Liquid chromatography–tandem mass spectrometry analysis of all gel fractions was performed on an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) coupled with a nanoACQUITY UPLC system (Waters, Manchester, UK). Sample aliquots were loaded onto a trap column (Symmetry$^R$ C18, 5 μm, 180 μm by 20 mm; Waters), and the peptides were then separated on an analytical column (BEH C18, 1.7 μm, 75 μm by 250 mm; Waters) and infused into the mass spectrometer via a 10-μm SilicaTip nanospray emitter (New Objective, Woburn, MA) attached to a nanospray interface (Proxeon, Odense, Denmark).

For separation, the following gradient of solvents A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) was used at a flow rate of 250 nl $min^{-1}$: solvent B at start: 0%; 0 to 3 min: linear ramp to 5% B; 3 to 56 min: ramp to 40% B; 56 to 62 min: ramp to 85% B; 85% B kept for 3 min followed by 100% A for 20 min for reequilibration.

The mass spectrometer was operated in positive ion mode at a capillary temperature of 200°C. The source voltage and focusing voltages were tuned for the transmission of MRFA peptide [mass/charge ratio (*m*/*z*) 524] (Sigma-Aldrich, St. Louis, MO). Data-dependent analysis was carried out in Orbitrap-IT parallel mode using CID fragmentation on the five most abundant ions in each cycle. The full scan mass spectrometry was performed in the Orbitrap at a resolution of 60,000 over the range *m*/*z* 350 to 1800. For the CID-MS2, the mono-isotopic 2+ and 3+ charged precursors were selected with an isolation width of 2 Da. MS2 was triggered by a minimal signal of $10^3$ with an AGC target of $3 \times 10^4$ ions and 150-ms maximum scan time using the chromatography function for peak apex detection. Collision energy was 35, and dynamic exclusion was set to 1 count and 60 s with a mass window of ±20 parts per million (ppm). Mass spectrometry scans were saved in profile mode while MS2 scans were saved in centroid mode.

The analysis of all gel fractions from three replicates resulted in 52 raw files (with replicate 2 run twice). Data were processed using Mascot Distiller 2.7 and Mascot Server 2.7 through Mascot Daemon 2.7 (Matrix Science, London, UK). Peak lists generated by Mascot Distiller were used for a database search using Mascot Server on the GeneDB_Rleguminosarum_Proteins fasta database from www.sanger.ac.uk/resources/downloads/bacteria/rhizobium-leguminosarum.html (May 2020, 7144 entries). For protein annotation, data from https://rhizosphere.org/lab-page/molecular-tools/genomes/rlv3841-genome/ (May 2020, 7288 entries) were used. A small database containing common contaminants (MaxQuant contaminants 2017, 250 entries) was included in the search.

Each raw file was processed and searched separately using the enzyme trypsin with two missed cleavages, a precursor mass tolerance of 10 ppm, and a fragment mass tolerance of 0.6 Da. Carbamidomethyl (C) was set as a fixed modification, and oxidation (M), deamidation (N, Q), and acetylation (protNterm) were set as variable modifications. Intensities for light and heavy $^{15}N$-labeled peptides were extracted using a Mascot Server $^{15}N$ metabolic quantitation method with the following parameters: 99.2% labeling (as determined by Mascot Distiller), Simpson's integration, isotope match rho = 0.6, XIC threshold = 0.1, isolation threshold = 0.5, peptide expect threshold = 0.05, outlier removal auto, and normalization none. The experimental design with three replicates and corresponding fractions was generated in the quantitation table exported via Mascot Daemon. The resulting expression table was used for ratio calculation and statistical analysis in RStudio 1.2.5033 with R version 3.6.3 (*65*). Potential contaminants were removed, and the table was filtered for proteins quantified in all three replicates. Light and heavy protein intensities were $log_{10}$-transformed, and ratios were calculated as light/heavy (bacteroids/*Rhizobium* cells) for each replicate. Those ratios were used for statistical testing using the limma eBayes function in R. The final ratio was calculated as median from the replicates. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*66*) partner repository with the dataset identifier PXD019467.

### Metabolic flux analysis

Metabolic flux analysis of *A. caulinodans* ORS571 was performed as previously described (*8*). Briefly, *A. caulinodans* was grown in UMS medium containing 20% [$^{13}C_4$]-succinate and 80% unlabeled sodium succinate. Bacteroids were isolated from the root nodules of *S. rostrata* as previously described (*67*). Isolated bacteroids were incubated in UMS medium supplemented with 10 mM 20% [$^{13}C_4$]-succinate for 24 hours at ≤1% $O_2$. Nitrogenase activity of free-living cultures and isolated bacteroids during labeling experiments was determined by an acetylene reduction assay (*8*). Free-living cultures were harvested in late exponential phase and bacteroids after 24 hours for metabolite extraction and gas chromatography–mass spectrometry analysis. Mass spectrometry data processing and isotopomer analysis of protein-derived amino acids and hydroxybutyrate obtained by hydrolysis of PHB was done using methods described previously (*8*, *68*).

Metabolic modeling was performed with INCA (Isotopomer Network Compartmental Analysis) using an iterative procedure (*69*). A complete description of the model, including the network carbon atom transitions, and net flux data are provided in data S9. The model along with the mass spectrometry measurements was simulated to obtain an optimized flux pattern in the network, followed by statistical validation of the flux maps. The fluxes were estimated relative to the succinate uptake flux fixed at 100 units. The assessment of goodness of fit for the flux maps with statistically

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

valid sum of squared residuals was performed by the comparison of simulated mass spectrometry measurements with that of experimentally measured values. To assess the precision of flux estimates, parameter continuation was performed in INCA to calculate the lower and upper bounds of the 95% confidence interval for the flux estimates. Comparison of fluxes estimated for the free-living bacteria and bacteroids were based on the confidence intervals (upper and lower limits) for a specific flux. A flux determined under two different conditions was deemed to be significantly different if the confidence intervals for the flux under the two conditions did not overlap.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/31/eabh2433/DC1

View/request a protocol for this paper from *Bio-protocol*.

## REFERENCES AND NOTES


1. D. F. Herridge, M. B. Peoples, R. M. Boddey, Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1–18 (2008).
2. V. C. S. Pankievicz, T. B. Irving, L. G. S. Maia, J.-M. Ané, Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biol.* **17**, 99 (2019).
3. C. W. Ronson, P. Lyttleton, J. G. Robertson, $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4284–4288 (1981).
4. M. J. Mitsch, G. C. diCenzo, A. Cowie, T. M. Finan, Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* **84**, e01561-17 (2018).
5. G. Mulley, M. Lopez-Gomez, Y. Zhang, J. Terpolilli, J. Prell, T. Finan, P. Poole, Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. *J. Bacteriol.* **192**, 4944–4953 (2010).
6. M. F. Dunn, Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. *FEMS Microbiol. Rev.* **22**, 105–123 (1998).
7. L. S. Green, D. W. Emerich, The formation of nitrogen-fixing bacteroids is delayed but not abolished in soybean infected by an α-ketoglutarate dehydrogenase-deficient mutant of *Bradyrhizobium japonicum*. *Plant Physiol.* **114**, 1359–1368 (1997).
8. J. J. Terpolilli, S. K. Masakapalli, R. Karunakaran, I. U. C. Webb, R. Green, N. J. Watmough, N. J. Kruger, R. G. Ratcliffe, P. S. Poole, Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. *J. Bacteriol.* **198**, 2864–2875 (2016).
9. C. P. Vance, G. H. Heichel, Carbon in $N_2$ fixation: Limitation or exquisite adaptation. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **42**, 373–390 (1991).
10. J. R. Gallon, The oxygen sensitivity of nitrogenase: A problem for biochemists and micro-organisms. *Trends Biochem. Sci.* **6**, 19–23 (1981).
11. M. D'Alessio, R. Nordeste, A. C. Doxey, T. C. Charles, Transcriptome analysis of polyhydroxybutyrate cycle mutants reveals discrete loci connecting nitrogen utilization and carbon storage in *Sinorhizobium meliloti*. *mSystems* **2**, e00035-17 (2017).
12. E. M. Lodwig, M. Leonard, S. Marroqui, T. R. Wheeler, K. Findlay, J. A. Downie, P. S. Poole, Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Mol. Plant Microbe Interact.* **18**, 67–74 (2005).
13. J. L. Sachs, K. W. Quides, C. E. Wendlandt, Legumes versus rhizobia: A model for ongoing conflict in symbiosis. *New Phytol.* **219**, 1199–1206 (2018).
14. D. Allaway, E. M. Lodwig, L. A. Crompton, M. Wood, R. Parsons, T. R. Wheeler, P. S. Poole, Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol. Microbiol.* **36**, 508–515 (2000).
15. S. O. Salminen, J. G. Streeter, Labeling of carbon pools in *Bradyrhizobium japonicum* and *Rhizobium leguminosarum* bv *viciae* bacteroids following incubation of intact nodules with $^{14}CO_2$. *Plant Physiol.* **100**, 597–604 (1992).
16. E. J. Patriarca, R. Tatè, M. Iaccarino, Key role of bacterial $NH_4^+$ metabolism in rhizobium-plant symbiosis. *Microbiol. Mol. Biol. Rev.* **66**, 203–222 (2002).
17. C. M. Brown, M. J. Dilworth, Ammonia assimilation by *Rhizobium* cultures and bacteroids. *J. Gen. Microbiol.* **86**, 39–48 (1975).
18. G. Mulley, J. P. White, R. Karunakaran, J. Prell, A. Bourdes, S. Bunnewell, L. Hill, P. S. Poole, Mutation of GOGAT prevents pea bacteroid formation and $N_2$ fixation by globally downregulating transport of organic nitrogen sources. *Mol. Microbiol.* **80**, 149–167 (2011).
19. C. A. Contador, S.-K. Lo, S. H. J. Chan, H.-M. Lam, Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* **5**, e00516-19 (2020).
20. O. Resendis-Antonio, M. Hernández, E. Salazar, S. Contreras, G. M. Batallar, Y. Mora, S. Encarnación, Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling. *BMC Syst. Biol.* **5**, 120 (2011).
21. G. C. diCenzo, A. Checcucci, M. Bazzicalupo, A. Mengoni, C. Viti, L. Dziewit, T. M. Finan, M. Galardini, M. Fondi, Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **7**, 12219 (2016).
22. Y. Yang, X.-P. Hu, B.-G. Ma, Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: A comparison between free-living and symbiotic states. *Mol. Biosyst.* **13**, 607–620 (2017).
23. T. Pfau, N. Christian, S. K. Masakapalli, L. J. Sweetlove, M. G. Poolman, O. Ebenhöh, The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* **8**, 12504 (2018).
24. G. C. diCenzo, M. Tesi, T. Pfau, A. Mengoni, M. Fondi, Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* **11**, 2574 (2020).
25. I. Thiele, B. Ø. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
26. J. Nielsen, Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).
27. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
28. A. M. Feist, B. O. Palsson, The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344–349 (2010).
29. S. Schuster, C. Hilgetag, On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182 (1994).
30. S. Klamt, J. Stelling, Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236 (2002).
31. R. Urbanczik, C. Wagner, Functional stoichiometric analysis of metabolic networks. *Bioinformatics* **21**, 4176–4180 (2005).
32. T. J. Clement, E. B. Baalhuis, B. Teusink, F. J. Bruggeman, R. Planqué, D. H. De Groot, Unlocking elementary conversion modes: Ecmtool unveils all capabilities of metabolic networks. *Cell Patterns* **2**, 100177 (2021).
33. F. Mus, M. B. Crook, K. Garcia, A. G. Costas, B. A. Geddes, E. D. Kouri, P. Paramasivan, M.-H. Ryu, G. E. D. Oldroyd, P. S. Poole, M. K. Udvardi, C. A. Voigt, J.-M. Ané, J. W. Peters, Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).
34. R. Karunakaran, V. K. Ramachandran, J. C. Seaman, A. K. East, B. Mouhsine, T. H. Mauchline, J. Prell, A. Skeffington, P. S. Poole, Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**, 4002–4014 (2009).
35. R. M. Wheatley, B. L. Ford, L. Li, S. T. N. Aroney, H. E. Knights, R. Ledermann, A. K. East, V. K. Ramachandran, P. S. Poole, Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23823–23834 (2020).
36. C. Lieven, M. E. Beber, B. G. Olivier, F. T. Bergmann, M. Ataman, P. Babaei, J. A. Bartell, L. M. Blank, S. Chauhan, K. Correia, C. Diener, A. Dräger, B. E. Ebert, J. N. Edirisinghe, J. P. Faria, A. M. Feist, G. Fengos, R. M. T. Fleming, B. García-Jiménez, V. Hatzimanikatis, W. van Helvoirt, C. S. Henry, H. Hermjakob, M. J. Herrgård, A. Kaafarani, H. U. Kim, Z. King, S. Klamt, E. Klipp, J. J. Koehorst, M. König, M. Lakshmanan, D. Y. Lee, S. Y. Lee, S. Lee, N. E. Lewis, F. Liu, H. Ma, D. Machado, R. Mahadevan, P. Maia, A. Mardinoglu, G. L. Medlock, J. M. Monk, J. Nielsen, L. K. Nielsen, J. Nogales, I. Nookaew, B. O. Palsson, J. A. Papin, K. R. Patil, M. Poolman, N. D. Price, O. Resendis-Antonio, A. Richelle, I. Rocha, B. J. Sánchez, P. J. Schaap, R. S. Malik Sheriff, S. Shoaie, N. Sonnenschein, B. Teusink, P. Vilaça, J. O. Vik, J. A. H. Wodke, J. C. Xavier, Q. Yuan, M. Zakhartsev, C. Zhang, MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
37. M. Udvardi, P. S. Poole, Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).
38. L. Zheng, R. H. White, V. L. Cash, D. R. Dean, Mechanism for the desulfurization of L-cysteine catalyzed by the *nifS* gene product. *Biochemistry* **33**, 4714–4720 (1994).
39. A. S. Paau, J. R. Cowles, D. Raveed, Development of bacteroids in alfalfa (*Medicago sativa*) nodules. *Phys. Ther.* **62**, 526–530 (1978).
40. J. Prell, A. Bourdes, R. Karunakaran, M. Lopez-Gomez, P. Poole, Pathway of γ-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. *J. Bacteriol.* **191**, 2177–2186 (2009).
41. E. T. Kiers, R. A. Rousseau, S. A. West, R. F. Denison, Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003).
42. R. Oono, C. G. Anderson, R. F. Denison, Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B Biol. Sci.* **278**, 2698–2703 (2011).
43. J. S. Edwards, R. Ramakrishna, B. O. Palsson, Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnol. Bioeng.* **77**, 27–36 (2002).


Downloaded from http://advances.sciencemag.org/ on August 8, 2021

44. J. K. Waters, T. P. Mawhinney, D. W. Emerich, Nitrogen assimilation and transport by ex planta nitrogen-fixing *Bradyrhizobium diazoefficiens* bacteroids is modulated by oxygen, bacteroid density and L-malate. *Int. J. Mol. Sci.* **21**, 7542 (2020).
45. M. M. Kuzma, H. Winter, P. Storer, I. Oresnik, C. A. Atkins, D. B. Layzell, The site of oxygen limitation in soybean nodules. *Plant Physiol.* **119**, 399–408 (1999).
46. E. Lodwig, S. Kumar, D. Allaway, A. Bourdes, J. Prell, U. Priefer, P. Poole, Regulation of L-alanine dehydrogenase in *Rhizobium leguminosarum* bv. *viciae* and its role in pea nodules. *J. Bacteriol.* **186**, 842–849 (2004).
47. C. Damiani, D. Pescini, R. Colombo, S. Molinari, L. Alberghina, M. Vanoni, G. Mauri, An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes. *Nat. Comput.* **13**, 321–331 (2014).
48. A. J. Gordon, F. R. Minchin, C. L. James, O. Komina, Sucrose synthase in legume nodules is essential for nitrogen fixation. *Plant Physiol.* **120**, 867–878 (1999).
49. E. M. Lodwig, A. H. F. Hosie, A. Bourdès, K. Findlay, D. Allaway, R. Karunakaran, J. A. Downie, P. S. Poole, Amino-acid cycling drives nitrogen fixation in the legume–*rhizobium* symbiosis. *Nature* **422**, 722–726 (2003).
50. J. Prell, J. P. White, A. Bourdes, S. Bunnewell, R. J. Bongaerts, P. S. Poole, Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12477–12482 (2009).
51. P. Mergaert, T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A.-E. Mausset, F. Barloy-Hubler, F. Galibert, A. Kondorosi, E. Kondorosi, Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5230–5235 (2006).
52. E. Ercolano, R. Mirabella, M. Chiurazzi, M. Merrick, The *Rhizobium leguminosarum glnB* gene is down-regulated during symbiosis. *Mol. Gen. Genet.* **264**, 555–564 (2001).
53. A. Heinken, I. Thiele, Anoxic conditions promote species-specific mutualism between gut microbes *in silico*. *Appl. Environ. Microbiol.* **81**, 4049–4061 (2015).
54. A. R. Pacheco, M. Moel, D. Segrè, Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* **10**, 103 (2019).
55. L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. J. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. El Assal, L. V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. Ben Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. P. Oliveira, P. T. Vuong, L. P. El Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. Aragón Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Ø. Palsson, I. Thiele, R. M. T. Fleming, Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
56. Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, B. O. Palsson, Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Comput. Biol.* **11**, e1004321 (2015).
57. C. Damiani, R. Colombo, D. Gaglio, F. Mastroianni, D. Pescini, H. V. Westerhoff, G. Mauri, M. Vanoni, L. Alberghina, A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The WarburQ effect. *PLOS Comput. Biol.* **13**, e1005758 (2017).
58. M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, M. Latorre, N. Loira, G. V. Markov, A. Maass, A. Siegel, Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLOS Comput. Biol.* **14**, e1006146 (2018).
59. G. C. diCenzo, A. B. Benedict, M. Fondi, G. C. Walker, T. M. Finan, A. Mengoni, J. S. Griffitts, Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLOS Genet.* **14**, e1007357 (2018).
60. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. Dejongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. D. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, D. Yu, KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
61. R. M. Wheatley, V. K. Ramachandran, B. A. Geddes, B. J. Perry, C. K. Yost, P. S. Poole, Role of $O_2$ in the growth of *Rhizobium leguminosarum* bv. viciae 3841 on glucose and succinate. *J. Bacteriol.* **199**, e00572-16 (2017).
62. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
63. P. S. Poole, N. A. Schofield, C. J. Reid, E. M. Drew, D. L. Walshaw, Identification of chromosomal genes located downstream of *dctD* that affect the requirement for calcium and the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* **140**, 2797–2809 (1994).
64. D. Wessel, U. I. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
65. R Core Team, R: A language and environment for statistical computing (2017); www.r-project.org/.
66. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
67. S. Tsukada, T. Aono, N. Akiba, K.-B. Lee, C.-T. Liu, H. Toyazaki, H. Oyaizu, Comparative genome-wide transcriptional profiling of *Azorhizobium caulinodans* ORS571 grown under free-living and symbiotic conditions. *Appl. Environ. Microbiol.* **75**, 5037–5046 (2009).
68. S. K. Masakapalli, N. J. Kruger, R. G. Ratcliffe, The metabolic flux phenotype of heterotrophic *Arabidopsis* cells reveals a complex response to changes in nitrogen supply. *Plant J.* **74**, 569–582 (2013).
69. J. D. Young, INCA: A computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333–1335 (2014).
70. K. Borah (2021), GC-MS datasets for *Azorhizobium caulinodans*, V1, Mendeley Data, https://doi.org/10.17632/y9tmxzv6p9.1.
71. C. Schulte, CarolinSchulte/bacteroid-metabolism: Bacteroid metabolism, v1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.4911594.
72. G. Cheng, R. Karunakaran, A. K. East, O. Munoz-Azcarate, P. S. Poole, Glutathione affects the transport activity of *Rhizobium leguminosarum* 3841 and is essential for efficient nodulation. *FEMS Microbiol. Lett.* **364**, fnx045 (2017).
73. P. Katinakis, R. M. K. Lankhorst, J. Louwerse, A. van Kammen, R. C. van den Bos, Bacteroid-encoded proteins are secreted into the peribacteroid space by *Rhizobium leguminosarum*. *Plant Mol. Biol.* **11**, 183–190 (1988).
74. A. R. Glenn, I. A. McKay, R. Arwas, M. J. Dilworth, Sugar metabolism and the symbiotic properties of carbohydrate mutants of *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **130**, 239–245 (1984).
75. I. A. McKay, A. R. Glenn, M. J. Dilworth, Gluconeogenesis in *Rhizobium leguminosarum* MNF3841. *J. Gen. Microbiol.* **131**, 2067–2073 (1985).
76. P. S. Poole, A. Blyth, C. J. Reid, K. Walters, Myo-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140**, 2787–2795 (1994).
77. J. Fry, M. Wood, P. S. Poole, Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Mol. Plant Microbe Interact.* **14**, 1016–1025 (2001).
78. A. H. F. Hosie, D. Allaway, C. S. Galloway, H. A. Dunsby, P. S. Poole, *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184**, 4071–4080 (2002).
79. W. Małek, M. Kowalski, Auxotrophic mutations related to symbiotic properties of *Rhizobium meliloti* strain L5-30. *Acta Microbiol. Pol.* **26**, 345–350 (1977).
80. P. Aneja, R. Dziak, G.-Q. Cai, T. C. Charles, Identification of an acetoacetyl coenzyme A synthetase-dependent pathway for utilization of L-(+)-3-hydroxybutyrate in *Sinorhizobium meliloti*. *J. Bacteriol.* **184**, 1571–1577 (2002).
81. I. A. McKay, M. J. Dilworth, A. R. Glenn, Carbon catabolism in continuous cultures and bacteroids of *Rhizobium leguminosarum* MNF 3841. *Arch. Microbiol.* **152**, 606–610 (1989).
82. M. F. Dunn, Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Crit. Rev. Microbiol.* **41**, 411–451 (2015).
83. A. S. Craig, K. I. Williamson, Three inclusions of rhizobial bacteroids and their cytochemical character. *Arch. Mikrobiol.* **87**, 165–171 (1972).
84. L. Rosendahl, M. J. Dilworth, A. R. Glenn, Exchange of metabolites across the peribacteroid membrane in pea root nodules. *J. Plant Physiol.* **139**, 635–638 (1992).
85. V. A. Becerra-Rivera, M. F. Dunn, Polyamine biosynthesis and biological roles in rhizobia. *FEMS Microbiol. Lett.* **366**, fnz084 (2019).
86. S. Fujihara, Biogenic amines in rhizobia and legume root nodules. *Microbes Environ.* **24**, 1–13 (2009).
87. I. A. McKay, M. J. Dilworth, A. R. Glenn, $C_4$-dicarboxylate metabolism in free-living and bacteroid forms of *Rhizobium leguminosarum* MNF3841. *J. Gen. Microbiol.* **134**, 1433–1440 (1988).
88. S. Marroquí, A. Zorreguieta, C. Santamaría, F. Temprano, M. Soberón, M. Megías, J. A. Downie, Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. *J. Bacteriol.* **183**, 854–864 (2001).


Downloaded from http://advances.sciencemag.org/ on August 8, 2021

89. J. F. Moran, I. Iturbe-Ormaetxe, M. A. Matamoros, M. C. Rubio, M. R. Clemente, N. J. Brewin, M. Becana, Glutathione and homoglutathione synthetases of legume nodules. Cloning, expression, and subcellular localization. *Plant Physiol.* **124**, 1381–1392 (2000).
90. M. A. Matamoros, N. Fernández-García, S. Wienkoop, J. Loscos, A. Saiz, M. Becana, Mitochondria are an early target of oxidative modifications in senescing legume nodules. *New Phytol.* **197**, 873–885 (2013).
91. L. Wang, M. J. Wise, Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften* **98**, 719–729 (2011).
92. E. Lodwig, P. Poole, Metabolism of *Rhizobium* bacteroids. *Crit. Rev. Plant Sci.* **22**, 37–78 (2003).
93. M. J. Delgado, E. J. Bedmar, J. A. Downie, Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. *Adv. Microb. Physiol.* **40**, 191–231 (1998).
94. O. Geiger, I. M. López-Lara, Rhizobial acyl carrier proteins and their roles in the formation of bacterial cell-surface components that are required for the development of nitrogen-fixing root nodules on legume hosts. *FEMS Microbiol. Lett.* **208**, 153–162 (2002).
95. P. S. Poole, M. Franklin, A. R. Glenn, M. J. Dilworth, The transport of L-glutamate by *Rhizobium leguminosarum* involves a common amino acid carrier. *J. Gen. Microbiol.* **131**, 1441–1448 (1985).
96. D. L. Walshaw, P. S. Poole, The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that also influences efflux of solutes. *Mol. Microbiol.* **21**, 1239–1252 (1996).
97. J. P. White, J. Prell, V. K. Ramachandran, P. S. Poole, Characterization of a γ-aminobutyric acid transport system of *Rhizobium leguminosarum* bv. *viciae* 3841. *J. Bacteriol.* **191**, 1547–1555 (2009).
98. G. Cheng, R. Karunakaran, A. K. East, P. S. Poole, Multiplicity of sulfate and molybdate transporters and their role in nitrogen fixation in *Rhizobium leguminosarum* bv. *viciae* Rlv3841. *Mol. Plant Microbe Interact.* **29**, 143–152 (2016).
99. D. A. Day, P. S. Poole, S. D. Tyerman, L. Rosendahl, Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell. Mol. Life Sci.* **58**, 61–71 (2001).

**Acknowledgments:** We thank B. Teusink at VU Amsterdam for helpful discussions and C. de Oliveira Martins at the John Innes Centre for support with analyzing the proteome data. C. Damiani at University of Milano-Bicocca is acknowledged for providing the ensemble-evolutionary FBA scripts. **Funding:** This work was supported by the Biotechnology and Biological Sciences Research Council (grant numbers BB/F013159/1 and BB/M011224/1 to C.C.M.S., N.C., and P.S.P.), the Natural Environment Research Council (grant number NE/L501530/1 to R.M.W.), NWO VICI grant 865.14.005 (to D.H.d.G.), the Engineering and Physical Sciences Research Council (grant number EP/M002454/1 to A.P.), the Clarendon Fund (Oxford University Press) and Keble College De Breyne Scholarship (to C.C.M.S.), and the Louis Dreyfus Weidenfeld Scholarship for Plant Science, University of Oxford (to K.B.) **Author contributions:** Conceptualization: R.G.R., N.J.K., A.P., and P.S.P. Methodology: C.C.M.S., J.J.T., G.S., D.H.d.G., R.G.R., N.J.K., A.P., and P.S.P. Investigation: C.C.M.S., K.B., R.M.W., J.J.T., G.S., N.C., D.H.d.G., R.G.R., N.J.K., A.P., and P.S.P. Visualization: C.C.M.S. and N.J.K. Supervision: R.G.R., N.J.K., A.P., and P.S.P. Writing—original draft: C.C.M.S. and P.S.P. Writing—review and editing: C.C.M.S., K.B., J.J.T., G.S., N.J.K., R.G.R., A.P., and P.S.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Proteome data are available via ProteomeXchange with identifier PXD019467. GC-MS data for the metabolic flux analysis are available on Mendeley Data (DOI: 10.17632/y9tmxzv6p9.1) (*70*). All code is available on Zenodo (DOI: 10.5281/zenodo.4911594) (*71*).

Submitted 24 February 2021
Accepted 14 June 2021
Published 30 July 2021
10.1126/sciadv.abh2433

**Citation:** C. C. M. Schulte, K. Borah, R. M. Wheatley, J. J. Terpolilli, G. Saalbach, N. Crang, D. H. de Groot, R. G. Ratcliffe, N. J. Kruger, A. Papachristodoulou, P. S. Poole, Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Sci. Adv.* **7**, eabh2433 (2021).

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

Science Advances

# Metabolic control of nitrogen fixation in rhizobium-legume symbioses

Carolin C. M. Schulte, Khushboo Borah, Rachel M. Wheatley, Jason J. Terpolilli, Gerhard Saalbach, Nick Crang, Daan H. de Groot, R. George Ratcliffe, Nicholas J. Kruger, Antonis Papachristodoulou and Philip S. Poole

*Sci Adv* **7** (31), eabh2433.
DOI: 10.1126/sciadv.abh2433

**ARTICLE TOOLS** http://advances.sciencemag.org/content/7/31/eabh2433

**SUPPLEMENTARY MATERIALS** http://advances.sciencemag.org/content/suppl/2021/07/26/7.31.eabh2433.DC1

**REFERENCES** This article cites 96 articles, 26 of which you can access for free
http://advances.sciencemag.org/content/7/31/eabh2433#BIBL

**PERMISSIONS** http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Downloaded from http://advances.sciencemag.org/ on August 8, 2021

Science Advances
AAAS

# Supplementary Materials for

## Metabolic control of nitrogen fixation in rhizobium-legume symbioses

Carolin C. M. Schulte, Khushboo Borah, Rachel M. Wheatley, Jason J. Terpolilli, Gerhard Saalbach, Nick Crang, Daan H. de Groot, R. George Ratcliffe*, Nicholas J. Kruger*, Antonis Papachristodoulou*, Philip S. Poole*

*Corresponding author. Email: george.ratcliffe@plants.ox.ac.uk (R.G.R.); nick.kruger@plants.ox.ac.uk (N.J.K.); antonis@eng.ox.ac.uk (A.P.); philip.poole@plants.ox.ac.uk (P.S.P.)

Published 30 July 2021, *Sci. Adv.* **7**, eabh2433 (2021)
DOI: 10.1126/sciadv.abh2433

**The PDF file includes:**

Supplementary Text
Figs. S1 to S10
Table S1
Legends for datasets S1 to S9
References

**Other Supplementary Material for this manuscript includes the following:**

Datasets S1 to S9

**Supplementary Text**

Evaluation of bacteroid proteome and transcriptome data

The proteome of bacteroids was compared with that of free-living *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) growing on minimal media supplemented with succinate and $^{15}NH_4Cl$ as the sole carbon and nitrogen source, respectively. A total of 1087 proteins were identified and grouped according to their Riley classification. 106 proteins had significantly increased abundance in bacteroids, 426 proteins had decreased abundance and 555 were unchanged.

The adaptation to symbiotic conditions was reflected by 49 and 40 proteins involved in adaptation and regulation, respectively (Figs. S1 and S2A). These were predominantly heat shock proteins and transcriptional regulators. This group also contained enzymes related to glutathione (GSH) metabolism and detoxification of reactive oxygen species (ROS). ROS are produced during the rhizobium-legume symbioses both as signaling molecules and as part of the plant defense response. The ability to detoxify ROS, either enzymatically or through production of antioxidant molecules, such as GSH, is essential for an effective symbiosis (*72*). As expected, the components of nitrogenase (Nif proteins) and associated electron transfer proteins (Fix proteins) were highly increased in abundance (*34*). Despite the non-growing state of bacteroids, 26.3% of the proteins detected were associated with basic cell functions, such as cell envelope, nucleic acid and protein synthesis and degradation, which is consistent with previous reports of protein secretion by bacteroids (*73*) and the known phenomenon of DNA endoreduplication (*51*).

The largest group of proteins detected (34.0%) was broadly classified as metabolic enzymes. Regarding central carbon metabolism, TCA cycle enzymes were identified in accordance with the essentiality of this pathway in Rlv3841 bacteroids (*8*). In addition, enzymes of the pentose phosphate and Entner-Doudoroff pathway were found (*74*, *75*). The presence of several respiratory chain enzymes reflects the obligate aerobic metabolism even under low oxygen conditions inside a nodule (*37*). 79 proteins were involved in amino acid biosynthesis, with 22 of these being downregulated compared to free-living bacteria. This result indicates active synthesis of proteins as well as amino acid exchange with the plant host as demonstrated previously (*14*, *49*). Further enzymes important for amino acid metabolism were found in the group of intermediary metabolism. Notably, this group contained GabT and GabT2, two enzymes of the γ-aminobutyric acid (GABA) pathway, which is highly induced in pea bacteroids (*40*).

The role of storage polymers in rhizobium-legume symbioses has been reported (*8*, *12*), and proteins associated with polyhydroxybutyrate (PHB) (PhaA, PhaB, PhaC, PhaE), glycogen (GlgC) and lipid (various Fab proteins) synthesis were detected. Several proteins involved in *myo*-inositol catabolism (IolABCDE) were present in agreement with the high concentrations of *myo*-inositol in nodules (*76*), even though mutants in inositol catabolism did not show a symbiotic phenotype in Rlv3841 (*77*).

Finally, 8.6% of the detected proteins were putatively involved in transport processes. Apart from the broad-specificity amino acid transporters Aap and Bra (*78*) with known importance for the symbiosis (*49*, *50*), most of the other transport proteins are poorly characterized in Rlv3841. Only one transport protein was significantly upregulated in bacteroids compared to free-living rhizobia, which may be a result of the sample preparation. Targeted purification of bacteroid membrane proteins will be required to further elucidate metabolite exchanges with the plant cell.

Among the 106 proteins showing significantly increased abundance in bacteroids, the Nif and Fix proteins were highly upregulated, along with cytochrome oxidases (Figs. S1 and S2B). Proteins involved in PHB synthesis were highly abundant, emphasizing the importance of this storage polymer, which had initially been thought to be absent from pea bacteroids (*8*).

Apart from GABA metabolism mentioned above, a homoserine O-acetyltransferase was strongly upregulated among the enzymes involved in amino acid metabolism. Starting from homoserine, an abundant and probably plant-provided metabolite in bacteroids (*8*), this enzyme catalyzes the first step towards methionine synthesis. HisI, a protein involved in histidine biosynthesis, also showed highly increased abundance. Together with the presence of five other His proteins (HisB, HisC1, HisD1, HisH, HisF), this indicates active biosynthesis of histidine in bacteroids, which agrees with studies on histidine auxotrophs of *Rhizobium meliloti* that could not be rescued by the plant host (*79*). A propionate CoA-transferase (pRL100119) also showed significantly increased abundance. This gene has high similarity to general acyl CoA:acetate/3-ketoacid CoA-transferases in various bacteria and is thus likely involved in lipid metabolism.

To further support the selection of pathways to include in the metabolic model, data from a previously published transcriptome study of Rlv3841 bacteroids (*34*) were evaluated. For model reconstruction purposes, we only considered the 136 genes found to be at least 3-fold upregulated in pea bacteroids at 28 d post inoculation relative to the free-living state. A large fraction of upregulated genes (34.6%) could not be assigned a function (Fig. S3A) and in agreement with the

proteome data, cell growth- and maintenance-associated genes were largely decreased in expression (Fig. S3B) due to the non-growing state of bacteroids.

The genes/proteins identified in both datasets supported the inclusion of PHB, methionine and histidine metabolism. Further upregulated genes identified in the transcriptome dataset included a phenylalanine-4-hydroxylase, consistent with significant labeling of phenylalanine in bacteroids of pea plants incubated in $^{15}N_2$ (*40*). An acetyl-CoA synthetase (pRL100121) showed a 4-fold increase in expression. This gene has high similarity to a characterized *acsA1* gene in *S. meliloti* (*80*) and is an important connection between anabolic and catabolic reactions.

Pathways included in *i*CS323

*Central carbon metabolism*: The TCA cycle was included as the main route for metabolizing dicarboxylic acids provided by the plant. Activity of all TCA cycle enzymes has been detected in bacteroids of *R. leguminosarum* (*81*), the genes have been found to be transcriptionally upregulated and corresponding proteins identified in the proteome data. Enzymes involved in glycolysis and gluconeogenesis, such as pyruvate kinase (PykA) and phosphoenolpyruvate carboxykinase (PckA), were detected in the bacteroid proteome in accordance with previous studies (*74*) and insertions in the *pckA* as well as fructokinase (*frk*) genes had deleterious effects on the symbiosis (*35*). Even though these pathways are expected to be less important in bacteroids than dicarboxylic acid metabolism, they are required e.g., for nucleotide and glycogen synthesis. *myo*-inositol is abundant in pea nodules (*76*) and may be an additional carbon source provided by the plant. All proteins encoded by the *iolABCDE* genes were detected in bacteroids.

*Amino acid metabolism*: Despite the non-growing state of bacteroids, protein turnover is still occurring, as evidenced by labeling studies (*40*) and the detection of ribosomal subunits in the proteome. It is important to account for metabolism of amino acids provided by the plant host (*49*, *50*) and synthesis of the exported amino acids alanine and aspartate (*14*, *15*). According to the experimental datasets used in this study, homoserine O-acetyltransferase was one of the most highly upregulated enzymes in bacteroids. Homoserine is probably provided by the plant host (*8*) and functions as a precursor for methionine synthesis. Multiple proteins involved in histidine, arginine and tryptophan biosynthesis were detected in bacteroids. Biosynthetic pathways for all amino acids except for asparagine, which is not synthesized in its free form by Rlv3841 (*8*), were

therefore included in the model. In addition, GABA catabolism, which has been shown to be strongly induced in bacteroids compared to free-living bacteria (*40*), was included.

*Carbon polymers*: Pathways for the biosynthesis of PHB and glycogen were included since both polymers are produced by bacteroids of *R. leguminosarum* (*8*, *12*) and especially proteins involved in PHB synthesis showed significantly increased abundance in bacteroids. Lipids have further been identified as an important carbon sink in bacteroids (*8*), and several Fab proteins involved in lipid metabolism were identified in the proteome. A representative species for free fatty acids and glycerolipids was therefore included in the model.

*Other polymers*: Homospermidine and polyphosphate have been shown to be present in bacteroids (*82*, *83*) and pathways for their synthesis were added to the model.

Nutrient exchange between pea plants and *Rhizobium leguminosarum* bacteroids

Based on a review of experimental evidence, we defined the boundary conditions for nutrient exchange between bacteroid and plant host.

*L-Malate:* It is well established that dicarboxylic acids, such as malate, succinate and fumarate, are the main carbon sources fueling symbiotic nitrogen fixation. A recent study found that malate is sufficient as the main carbon source to support efficient nitrogen fixation by Rlv3841 (*4*).

*L-Leucine:* Transport of branched-chain amino acids is required at low levels to support an efficient symbiosis, with leucine being the most likely to be required from the plant host (*50*).

*Homoserine:* Homoserine levels are high in pea bacteroids (*8*) and although catabolic genes (*hom1*, *hom2*) were not upregulated, a putative homoserine O-acetyltransferase (MetX, encoded by pRL100137) was highly increased in abundance compared to free-living cells, indicating an important role for homoserine as a precursor for amino acid synthesis.

*Other amino acids*: While the requirement for branched-chain amino acids has been explicitly demonstrated (*50*), this does not preclude provision of other amino acids by the plant. For example, enzymes involved in GABA metabolism are highly active in bacteroids, even though this pathway does not seem to be essential for efficient nitrogen fixation (*40*). Elementary conversion mode analysis was therefore performed for all proteinogenic amino acids and GABA as inputs.

While it is generally accepted that ammonia is the main nitrogenous compound supplied to the plant by the bacteroid, experimental evidence suggests that amino acids such as alanine and aspartate may play a role as additional secretion products (*14*, *84*). Demand reactions for those amino acids were therefore also included.

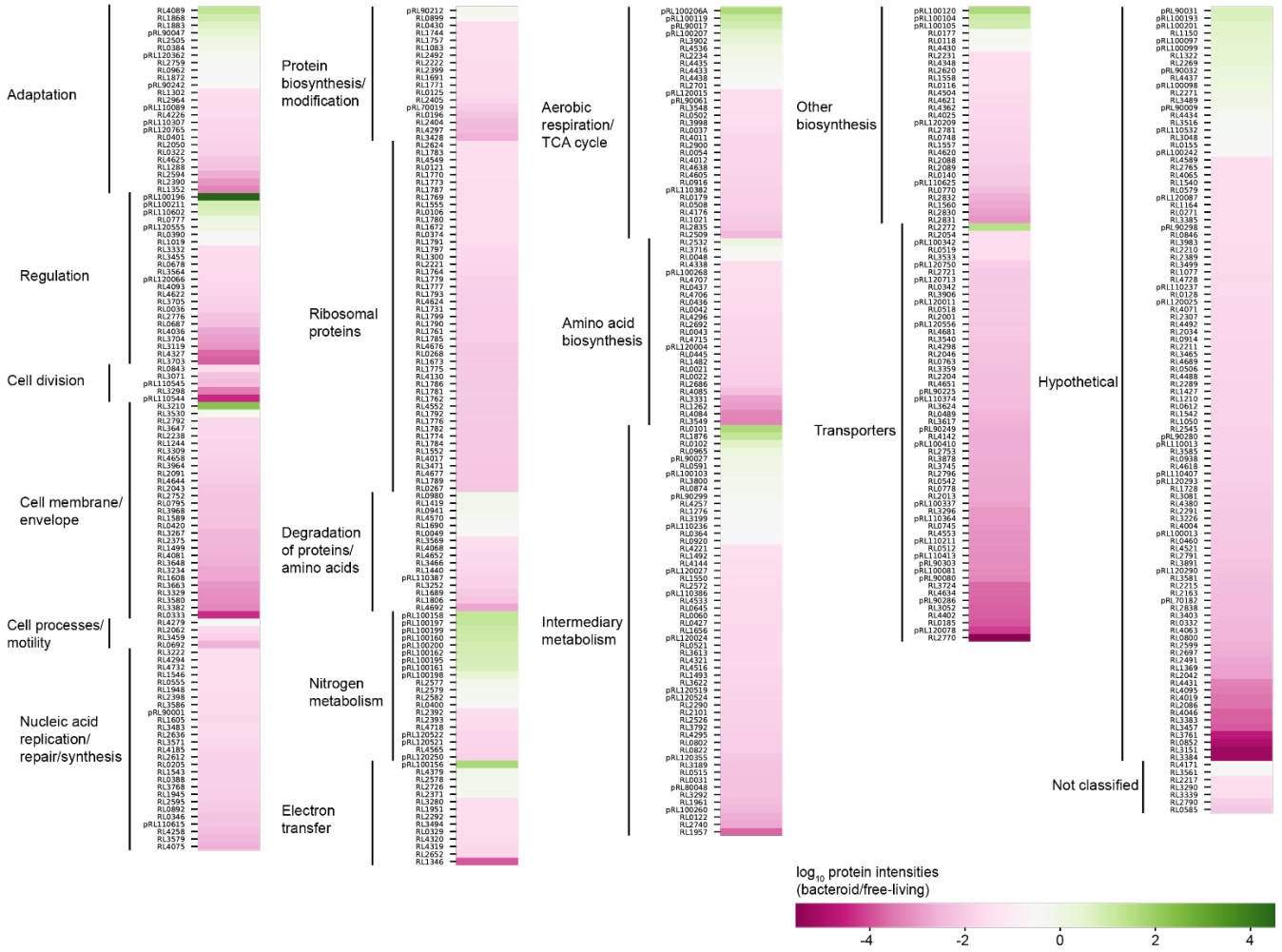



**Fig. S1. Changes in the proteome of bacteroids compared to free-living *Rhizobium leguminosarum* bv. *viciae* 3841.** The proteome of bacteroids at 28 d post inoculation and free-living bacteria grown on succinate and ammonia was compared. Heatmap showing the median of the $\log_{10}$ ratio bacteroid/free-living for the intensities of all proteins whose abundance changed significantly between the two conditions. Genes are grouped according to their Riley classification.

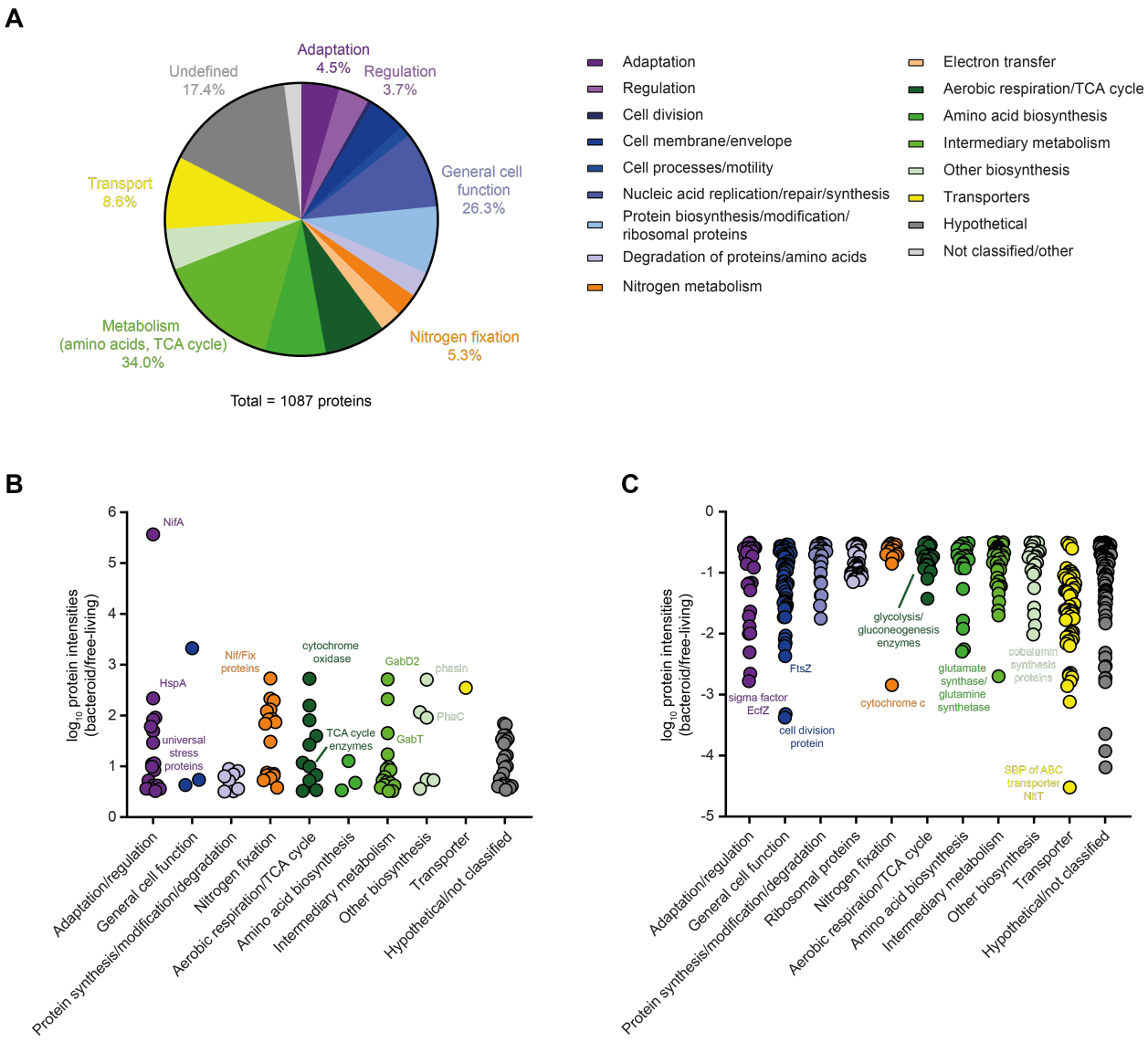



**Fig. S2. Comparative proteomics of bacteroids and free-living *Rhizobium leguminosarum* bv. *viciae* 3841.** The proteome of bacteroids at 28 d post inoculation and free-living bacteria grown on succinate and ammonia was compared. (**A**) Pie chart showing the functional distribution of all proteins detected in bacteroids according to their Riley classification. (**B**) Scatter plot showing the median of the $\log_{10}$ ratio bacteroid/free-living for the intensities of the 106 proteins that were significantly increased in abundance in bacteroids. (**C**) Scatter plot showing the median of the $\log_{10}$ ratio bacteroid/free-living for the intensities of the 426 proteins that were significantly decreased in abundance in bacteroids.

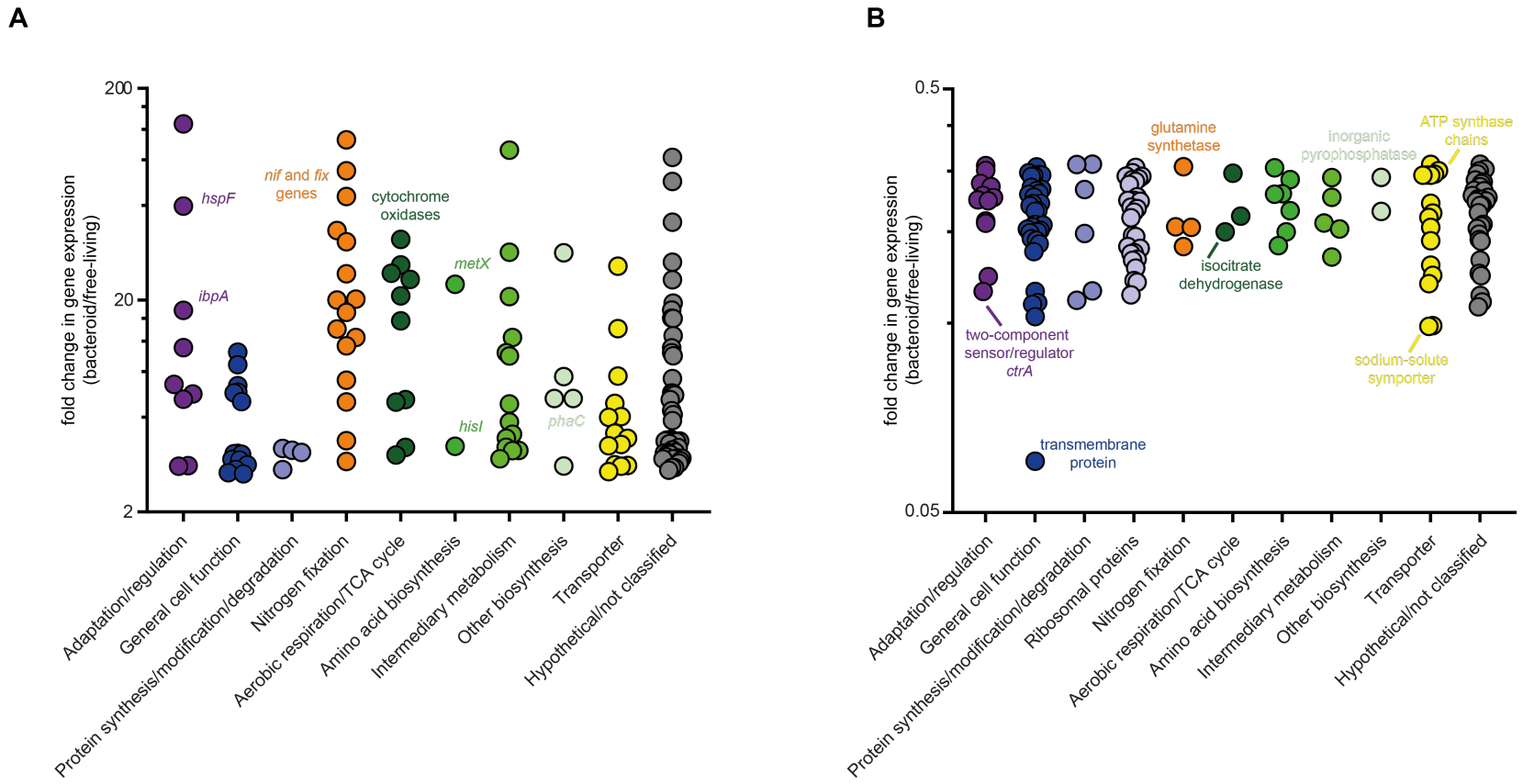



**Fig. S3. Comparative transcriptomics of bacteroids and free-living *Rhizobium leguminosarum* bv. *viciae* 3841**. The transcriptome of bacteroids at 28 d post inoculation and free-living bacteria grown on succinate and ammonia *(34)* was compared. (**A**) Scatter plot showing the fold change (bacteroid/free-living) in expression for genes upregulated in bacteroids. (**B**) Scatter plot showing the fold change (bacteroid/free-living) in expression for genes downregulated in bacteroids. Only genes with at least 3-fold up- or downregulation are shown.

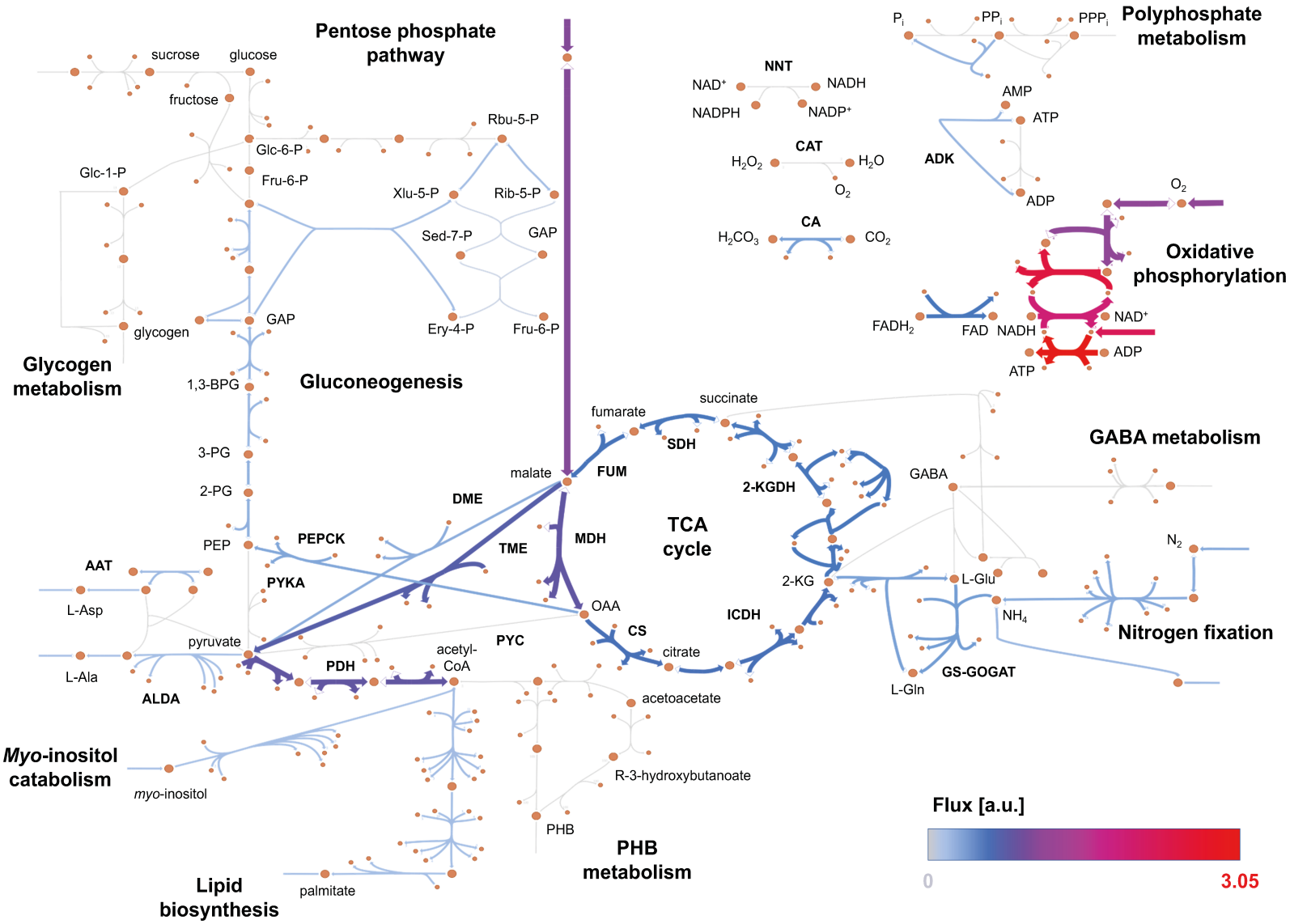

Pentose phosphate pathway
sucrose
glucose
fructose
Rbu-5-P
Glc-6-P
Glc-1-P
Fru-6-P
Xlu-5-P
Rib-5-P
Sed-7-P
GAP
glycogen
GAP
Ery-4-P
Fru-6-P
Glycogen metabolism
1,3-BPG
Gluconeogenesis
3-PG
2-PG
PEP
PEPCK
DME
TME
MDH
malate
FUM
fumarate
SDH
succinate
2-KGDH
TCA cycle
AAT
L-Asp
PYKA
OAA
CS
citrate
ICDH
2-KG
pyruvate
PYC
L-Ala
ALDA
PDH
acetyl-CoA
acetoacetate
R-3-hydroxybutanoate
PHB
PHB metabolism
myo-inositol catabolism
myo-inositol
Lipid biosynthesis
palmitate
NNT
NAD+
NADH
NADPH
NADP+
CAT
H2O2
H2O
O2
CA
H2CO3
CO2
Pi
PPi
PPPi
Polyphosphate metabolism
AMP
ATP
ADK
ADP
O2
Oxidative phosphorylation
FADH2
FAD
NADH
NAD+
ADP
ATP
GABA metabolism
GABA
N2
L-Glu
NH4
Nitrogen fixation
GS-GOGAT
L-Gln
Flux [a.u.]
0
3.05

**Fig. S4. Summary map of *i*CS323, a model of bacteroid metabolism in *Rhizobium leguminosarum* bv. *viciae* 3841.** A metabolic network of pea bacteroids was reconstructed from proteome, transcriptome and gene essentiality data. The map summarizes reactions in the model with colors corresponding to a flux distribution for maximum nitrogenase activity with malate, succinate and oxygen uptake rates constrained according to boundary conditions used in a modeling study of *Sinorhizobium meliloti* bacteroids *(21)*. Color and arrow thickness correspond to flux values with thin grey arrows indicating inactive reactions. Some pathways for amino acid biosynthesis have been omitted for clarity.

AAT: aspartate aminotransferase, ADK: adenylate kinase, ALDA: alanine dehydrogenase, CA: carbonic anhydrase, CAT: catalase, CS: citrate synthase, DME: NAD-dependent malic enzyme, FUM: fumarase, GS-GOGAT: glutamine synthetase-glutamine oxoglutarate aminotransferase, ICDH: isocitrate dehydrogenase, 2-KGDH: 2-ketoglutarate dehydrogenase, MDH: malate dehydrogenase, NNT: nicotinamide nucleotide transhydrogenase, PYC: pyruvate carboxylase, PDH: pyruvate dehydrogenase, PEPCK: phosphoenolpyruvate carboxykinase, PYKA: pyruvate kinase, SDH: succinate dehydrogenase, TME: NADP-dependent malic enzyme; 1,3-BPG: 1,3-bisphosphoglycerate, Ery-4-P: erythrose 4-phosphate, Fru-6-P: fructose 6-phosphate, GAP: glyceraldehyde 3-phosphate, Glc-1-P: glucose 1-phosphate, Glc-6-P: glucose 6-phosphate, 2-KG: 2-ketoglutarate, OAA: oxaloacetate, 2-PG: 2-phosphoglycerate, 3-PG: 3-phosphoglycerate, PEP: phosphoenolpyruvate, PHB: polyhydroxybutyrate, Rib-5-P: ribose 5-phosphate, Rbu-5-P: ribulose 5-phosphate, Sed-7-P: sedoheptulose 7-phosphate, Xlu-5-P: xylulose 5-phosphate

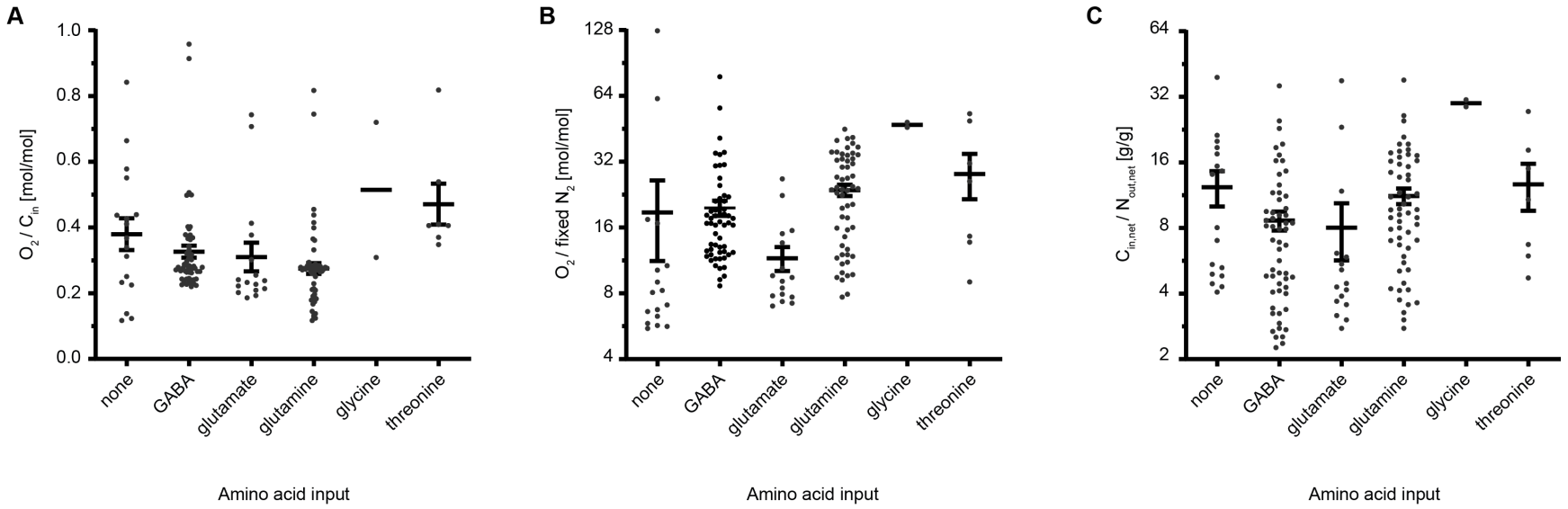





**Fig. S5. Effect of different amino acid inputs on oxygen demand and carbon cost of nitrogen fixation.** Elementary conversion modes were calculated with malate, succinate and different amino acids as inputs and carbon polymers (palmitate, PHB, glycerolipid, glycogen), ammonia, alanine and aspartate as outputs. (**A**) Oxygen uptake per carbon uptake, (**B**) oxygen uptake per fixed $N_2$ and (**C**) carbon cost (difference of carbon input and output) per nitrogen secreted (difference of nitrogen input and output). Conversion modes are grouped according to the amino acid in the input. Each data point represents an individual conversion mode, lines and bars indicate mean ± SEM.

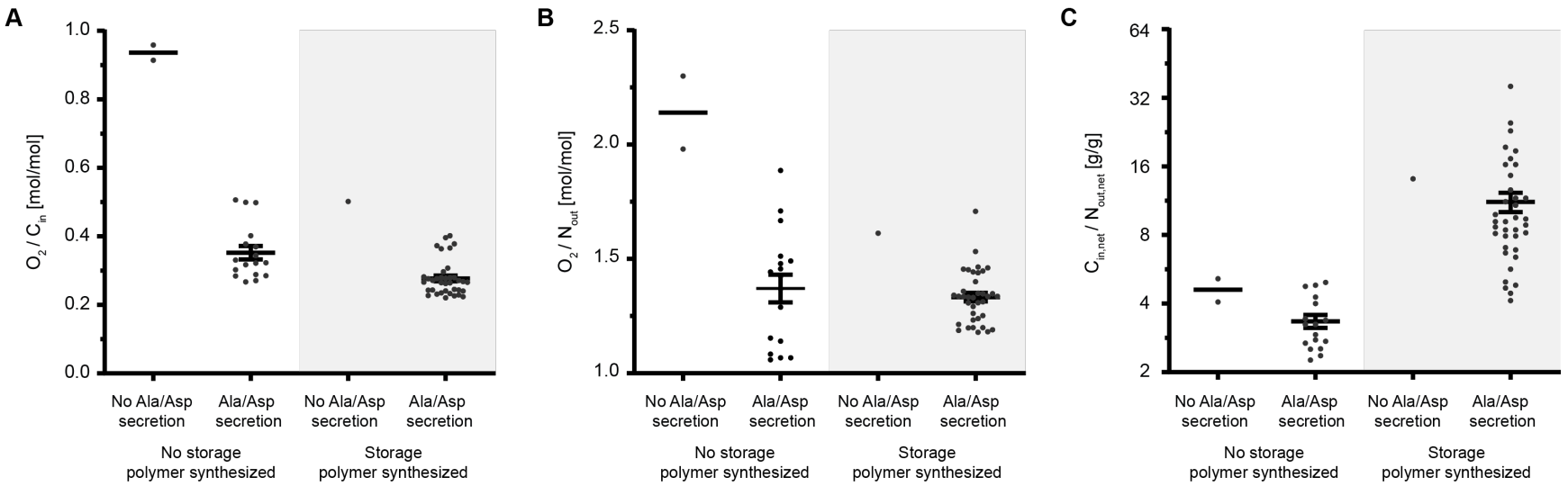





**Fig. S6. Effect of polymer synthesis and amino acid secretion on oxygen demand and carbon cost of nitrogen fixation.** Elementary conversion modes were calculated with malate, succinate and GABA as carbon sources and carbon polymers (palmitate, PHB, glycerolipid, glycogen), ammonia, alanine and aspartate as outputs. (**A**) Oxygen uptake per carbon uptake, (**B**) oxygen uptake per nitrogen output and (**C**) carbon cost (difference of carbon input and output) per nitrogen secreted (difference of nitrogen input and output). Each data point represents an individual conversion mode, lines and bars indicate mean ± SEM.

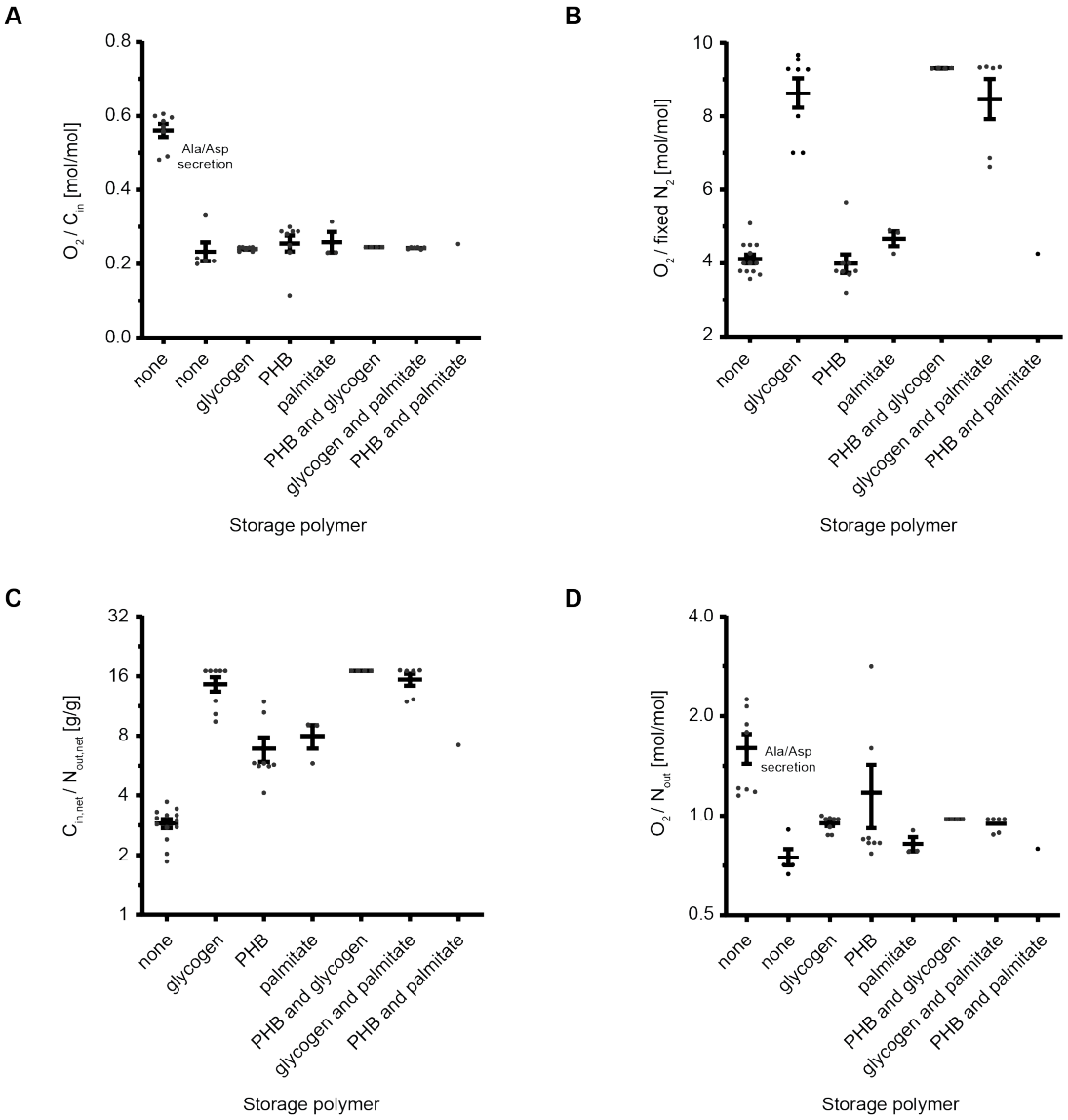



**Fig. S7. Elementary conversion mode analysis of *Sinorhizobium fredii* bacteroids.** Conversion modes for the metabolic model *i*CC541 *(19)* were calculated with succinate, malate, and glutamate as the main carbon inputs and carbon polymers (PHB, glycogen, palmitate), ammonia, alanine and aspartate as outputs. (**A**) Oxygen uptake per carbon uptake, (**B**) oxygen uptake per fixed $N_2$, (**C**) carbon cost (difference of carbon input and output) per nitrogen secreted (difference of nitrogen input and output) and (**D**) oxygen uptake per nitrogen output. In (**A**) and (**D**), ECMs without storage polymer production have been separated into those secreting only ammonia and those secreting alanine and/or aspartate in addition to/instead of ammonia. Each data point represents an individual conversion mode, lines and bars indicate mean ± SEM.

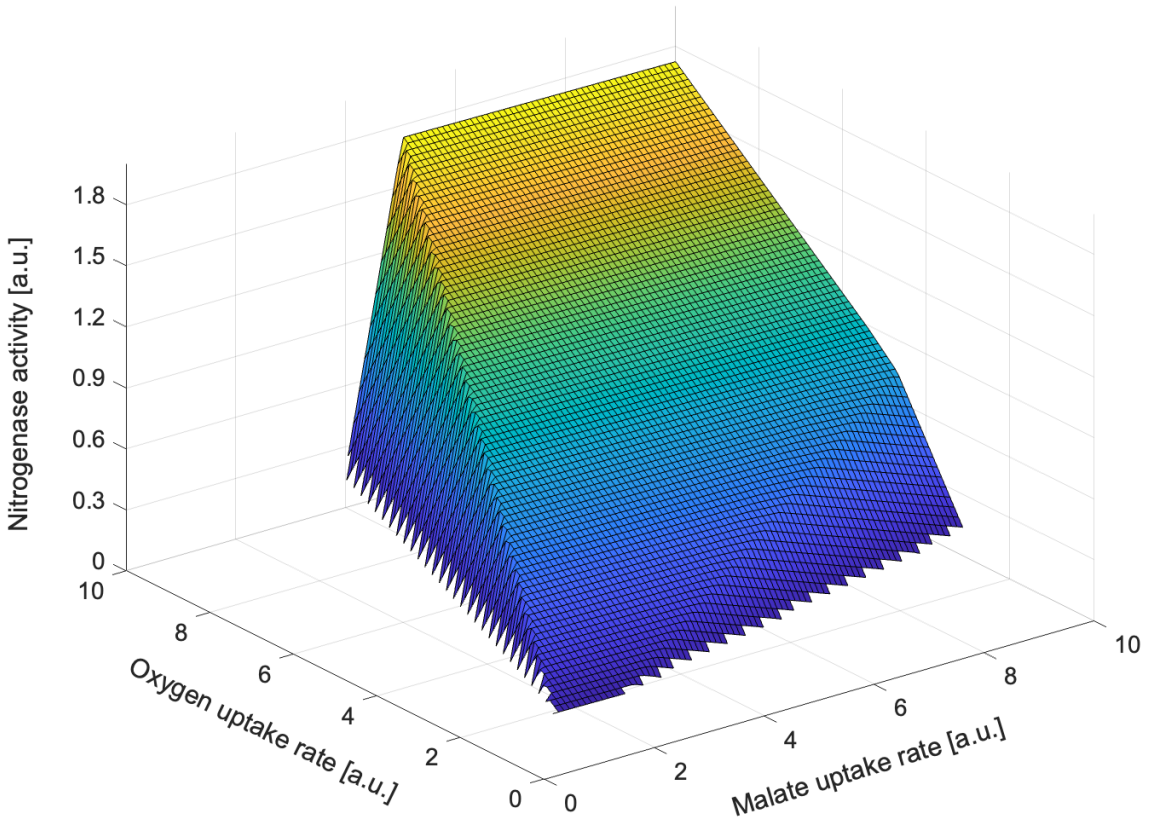



**Fig. S8. Phenotype phase plane analysis of *i*CS323.** Three-dimensional representation of the phenotype phase plane analysis of *i*CS323. Colors correspond to nitrogenase activity.

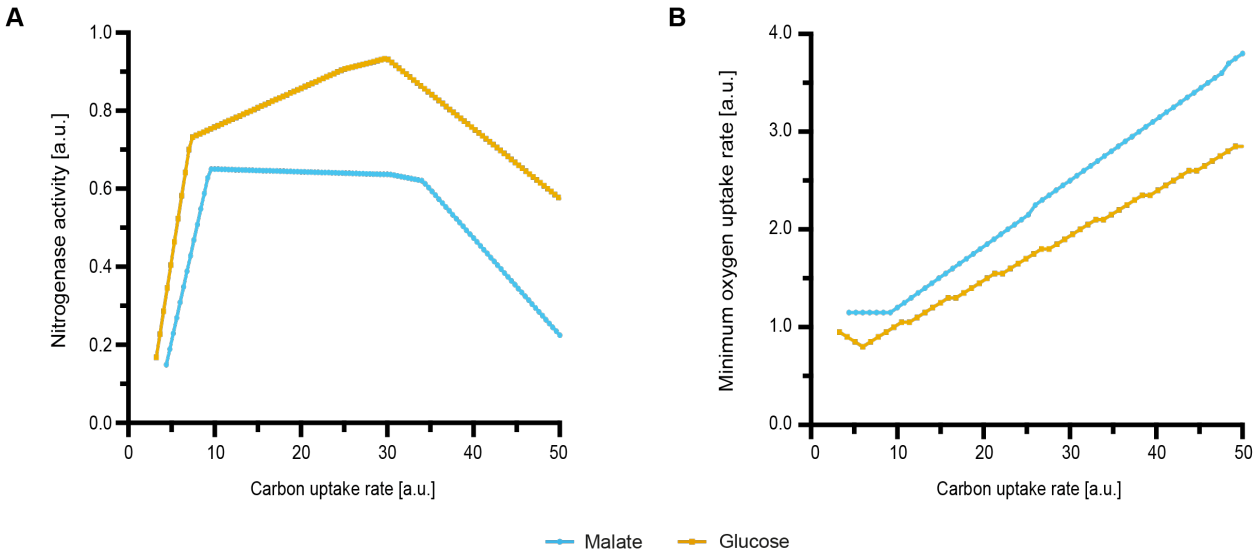



**Fig. S9. Predicted effect of malate and glucose catabolism on nitrogenase activity and oxygen demand.** (**A**) Maximum nitrogenase activity for malate or glucose as a carbon source with maximum oxygen uptake of 4 flux units. (**B**) Minimum oxygen demand for nitrogenase activity with malate or glucose as a carbon source. Values are shown per mol of carbon uptake.

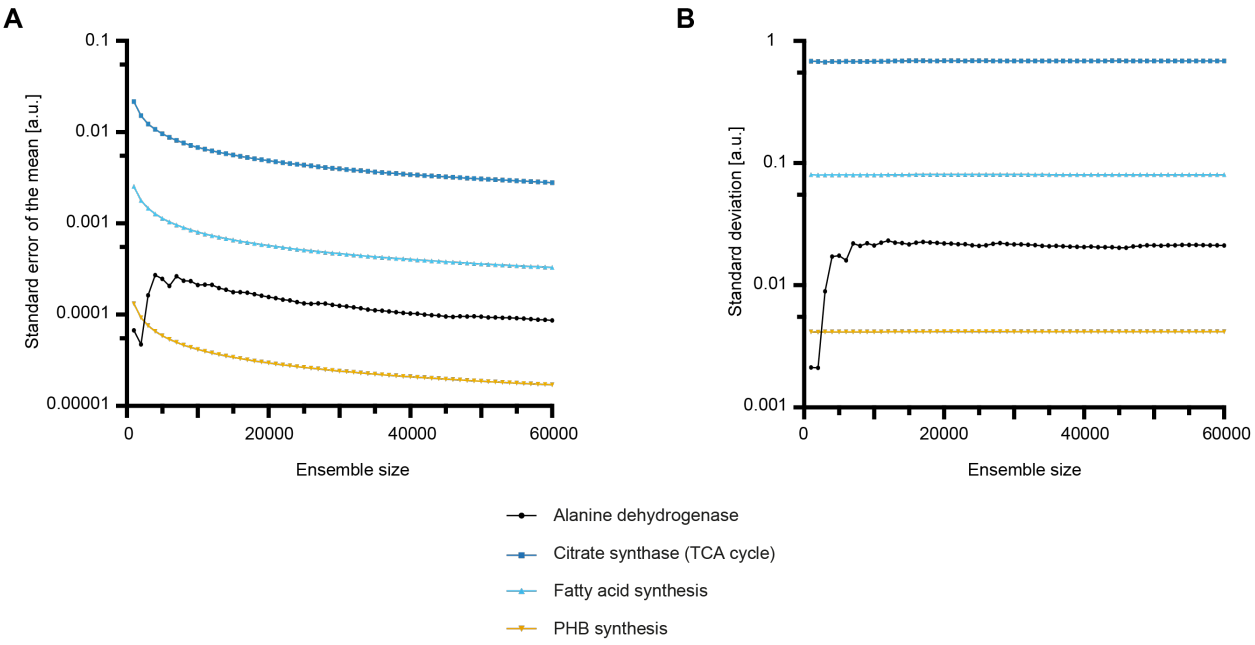



**Fig. S10. Determination of ensemble size for ensemble-evolutionary flux balance analysis.** To define the ensemble size (number of random objective functions) used in ensemble-evolutionary flux balance analysis computations, (**A**) the standard error of the mean and (**B**) the standard deviation for different ensemble sizes were calculated for the reactions of interest.

**Table S1. Properties of the experimental datasets and the bacteroid model *i*CS323.**

| **Bacteroid datasets** | |
|---|---|
| Total identified proteins | 1087 |
| Involved in metabolism | 489 |
| Included in model | 218 |
| Upregulated genes | 136 |
| Involved in metabolism | 49 |
| Corresponding protein detected | 52 |
| Included in model | 28 |
| Essential genes for bacteroids and nodule bacteria | 258 |
| Involved in metabolism | 69 |
| Corresponding protein detected | 70 |
| Gene upregulated | 11 |
| Included in model | 33 |
| | |
| ***i*CS323** | |
| Genes | 323 |
| Metabolites | 237 |
| Unique EC identifiers | 176 |
| Reactions | 299 |
| Metabolic reactions | 207 |
| Gene-associated metabolic reactions | 203 |
| Transport reactions | 32 |
| Exchange and demand reactions | 60 |
| Metabolic reactions supported by experimental evidence | 177 |

**Dataset S1 (separate file).** Proteome data comparing unlabeled bacteroids of *Rhizobium leguminosarum* bv. *viciae* 3841 with $^{15}N$ labeled free-living bacteria.

**Dataset S2 (separate file).** Excel file for *i*CS323.

**Dataset S3 (separate file).** sbml file for *i*CS323.

**Dataset S4 (separate file).** MATLAB structure for *i*CS323.

**Dataset S5 (separate file).** MEMOTE report for *i*CS323.

**Dataset S6 (separate file).** Comparison of *in silico* gene essentiality predictions with gene essentiality determined by insertion sequencing.

**Dataset S7 (separate file).** Full set of elementary conversion modes for *i*CS323 and *i*CC541.

**Dataset S8 (separate file).** List of input and output metabolites used in elementary conversion mode enumeration and boundary conditions for general modeling procedures.

**Dataset S9 (separate file).** Model and measured flux values for $^{13}C$ metabolic flux analysis of *Azorhizobium caulinodans* ORS571.

## REFERENCES AND NOTES


1. D. F. Herridge, M. B. Peoples, R. M. Boddey, Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1–18 (2008).

2. V. C. S. Pankievicz, T. B. Irving, L. G. S. Maia, J.-M. Ané, Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biol.* **17**, 99 (2019).

3. C. W. Ronson, P. Lyttleton, J. G. Robertson, $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4284–4288 (1981).

4. M. J. Mitsch, G. C. diCenzo, A. Cowie, T. M. Finan, Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* **84**, e01561-17 (2018).

5. G. Mulley, M. Lopez-Gomez, Y. Zhang, J. Terpolilli, J. Prell, T. Finan, P. Poole, Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. *J. Bacteriol.* **192**, 4944–4953 (2010).

6. M. F. Dunn, Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. *FEMS Microbiol. Rev.* **22**, 105–123 (1998).

7. L. S. Green, D. W. Emerich, The formation of nitrogen-fixing bacteroids is delayed but not abolished in soybean infected by an α-ketoglutarate dehydrogenase-deficient mutant of *Bradyrhizobium japonicum*. *Plant Physiol.* **114**, 1359–1368 (1997).

8. J. J. Terpolilli, S. K. Masakapalli, R. Karunakaran, I. U. C. Webb, R. Green, N. J. Watmough, N. J. Kruger, R. G. Ratcliffe, P. S. Poole, Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. *J. Bacteriol.* **198**, 2864–2875 (2016).

9. C. P. Vance, G. H. Heichel, Carbon in $N_2$ fixation: Limitation or exquisite adaptation. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **42**, 373–390 (1991).

10. J. R. Gallon, The oxygen sensitivity of nitrogenase: A problem for biochemists and micro-organisms. *Trends Biochem. Sci.* **6**, 19–23 (1981).

11. M. D'Alessio, R. Nordeste, A. C. Doxey, T. C. Charles, Transcriptome analysis of polyhydroxybutyrate cycle mutants reveals discrete loci connecting nitrogen utilization and carbon storage in *Sinorhizobium meliloti*. *mSystems* **2**, e00035-17 (2017).

12. E. M. Lodwig, M. Leonard, S. Marroqui, T. R. Wheeler, K. Findlay, J. A. Downie, P. S. Poole, Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Mol. Plant Microbe Interact.* **18**, 67–74 (2005).

13. J. L. Sachs, K. W. Quides, C. E. Wendlandt, Legumes versus rhizobia: A model for ongoing conflict in symbiosis. *New Phytol.* **219**, 1199–1206 (2018).

14. D. Allaway, E. M. Lodwig, L. A. Crompton, M. Wood, R. Parsons, T. R. Wheeler, P. S. Poole, Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol. Microbiol.* **36**, 508–515 (2000).

15. S. O. Salminen, J. G. Streeter, Labeling of carbon pools in *Bradyrhizobium japonicum* and *Rhizobium leguminosarum* bv *viciae* bacteroids following incubation of intact nodules with $^{14}CO_2$. *Plant Physiol.* **100**, 597–604 (1992).

16. E. J. Patriarca, R. Tatè, M. Iaccarino, Key role of bacterial $NH_4^+$ metabolism in rhizobium-plant symbiosis. *Microbiol. Mol. Biol. Rev.* **66**, 203–222 (2002).

17. C. M. Brown, M. J. Dilworth, Ammonia assimilation by *Rhizobium* cultures and bacteroids. *J. Gen. Microbiol.* **86**, 39–48 (1975).

18. G. Mulley, J. P. White, R. Karunakaran, J. Prell, A. Bourdes, S. Bunnewell, L. Hill, P. S. Poole, Mutation of GOGAT prevents pea bacteroid formation and $N_2$ fixation by globally downregulating transport of organic nitrogen sources. *Mol. Microbiol.* **80**, 149–167 (2011).

19. C. A. Contador, S.-K. Lo, S. H. J. Chan, H.-M. Lam, Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* **5**, e00516-19 (2020).

20. O. Resendis-Antonio, M. Hernández, E. Salazar, S. Contreras, G. M. Batallar, Y. Mora, S. Encarnación, Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling. *BMC Syst. Biol.* **5**, 120 (2011).

21. G. C. diCenzo, A. Checcucci, M. Bazzicalupo, A. Mengoni, C. Viti, L. Dziewit, T. M. Finan, M. Galardini, M. Fondi, Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **7**, 12219 (2016).

22. Y. Yang, X.-P. Hu, B.-G. Ma, Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: A comparison between free-living and symbiotic states. *Mol. Biosyst.* **13**, 607–620 (2017).

23. T. Pfau, N. Christian, S. K. Masakapalli, L. J. Sweetlove, M. G. Poolman, O. Ebenhöh, The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* **8**, 12504 (2018).

24. G. C. diCenzo, M. Tesi, T. Pfau, A. Mengoni, M. Fondi, Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* **11**, 2574 (2020).

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

26. J. Nielsen, Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).

27. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).

28. A. M. Feist, B. O. Palsson, The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344–349 (2010).

29. S. Schuster, C. Hilgetag, On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182 (1994).

30. S. Klamt, J. Stelling, Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236 (2002).

31. R. Urbanczik, C. Wagner, Functional stoichiometric analysis of metabolic networks. *Bioinformatics* **21**, 4176–4180 (2005).

32. T. J. Clement, E. B. Baalhuis, B. Teusink, F. J. Bruggeman, R. Planqué, D. H. De Groot, Unlocking elementary conversion modes: Ecmtool unveils all capabilities of metabolic networks. *Cell Patterns* **2**, 100177 (2021).

33. F. Mus, M. B. Crook, K. Garcia, A. G. Costas, B. A. Geddes, E. D. Kouri, P. Paramasivan, M.-H. Ryu, G. E. D. Oldroyd, P. S. Poole, M. K. Udvardi, C. A. Voigt, J.-M. Ané, J. W. Peters, Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).

34. R. Karunakaran, V. K. Ramachandran, J. C. Seaman, A. K. East, B. Mouhsine, T. H. Mauchline, J. Prell, A. Skeffington, P. S. Poole, Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**, 4002–4014 (2009).

35. R. M. Wheatley, B. L. Ford, L. Li, S. T. N. Aroney, H. E. Knights, R. Ledermann, A. K. East, V. K. Ramachandran, P. S. Poole, Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23823–23834 (2020).

36. C. Lieven, M. E. Beber, B. G. Olivier, F. T. Bergmann, M. Ataman, P. Babaei, J. A. Bartell, L. M. Blank, S. Chauhan, K. Correia, C. Diener, A. Dräger, B. E. Ebert, J. N. Edirisinghe, J. P. Faria, A. M. Feist, G. Fengos, R. M. T. Fleming, B. García-Jiménez, V. Hatzimanikatis, W. van Helvoirt, C. S. Henry, H. Hermjakob, M. J. Herrgård, A. Kaafarani, H. U. Kim, Z. King, S. Klamt, E. Klipp, J. J. Koehorst, M. König, M. Lakshmanan, D. Y. Lee, S. Y. Lee, S. Lee, N. E. Lewis, F. Liu, H. Ma, D. Machado, R. Mahadevan, P. Maia, A. Mardinoglu, G. L. Medlock, J. M. Monk, J. Nielsen, L. K.

Nielsen, J. Nogales, I. Nookaew, B. O. Palsson, J. A. Papin, K. R. Patil, M. Poolman, N. D. Price, O. Resendis-Antonio, A. Richelle, I. Rocha, B. J. Sánchez, P. J. Schaap, R. S. Malik Sheriff, S. Shoaie, N. Sonnenschein, B. Teusink, P. Vilaça, J. O. Vik, J. A. H. Wodke, J. C. Xavier, Q. Yuan, M. Zakhartsev, C. Zhang, MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).

37. M. Udvardi, P. S. Poole, Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).

38. L. Zheng, R. H. White, V. L. Cash, D. R. Dean, Mechanism for the desulfurization of L-cysteine catalyzed by the *nifS* gene product. *Biochemistry* **33**, 4714–4720 (1994).

39. A. S. Paau, J. R. Cowles, D. Raveed, Development of bacteroids in alfalfa (*Medicago sativa*) nodules. *Phys. Ther.* **62**, 526–530 (1978).

40. J. Prell, A. Bourdes, R. Karunakaran, M. Lopez-Gomez, P. Poole, Pathway of γ-aminobutyrate metabolism in *Rhizobium leguminosarum* 3841 and its role in symbiosis. *J. Bacteriol.* **191**, 2177–2186 (2009).

41. E. T. Kiers, R. A. Rousseau, S. A. West, R. F. Denison, Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003).

42. R. Oono, C. G. Anderson, R. F. Denison, Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. R. Soc. B Biol. Sci.* **278**, 2698–2703 (2011).

43. J. S. Edwards, R. Ramakrishna, B. O. Palsson, Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnol. Bioeng.* **77**, 27–36 (2002).

44. J. K. Waters, T. P. Mawhinney, D. W. Emerich, Nitrogen assimilation and transport by ex planta nitrogen-fixing *Bradyrhizobium diazoefficiens* bacteroids is modulated by oxygen, bacteroid density and L-malate. *Int. J. Mol. Sci.* **21**, 7542 (2020).

45. M. M. Kuzma, H. Winter, P. Storer, I. Oresnik, C. A. Atkins, D. B. Layzell, The site of oxygen limitation in soybean nodules. *Plant Physiol.* **119**, 399–408 (1999).

46. E. Lodwig, S. Kumar, D. Allaway, A. Bourdes, J. Prell, U. Priefer, P. Poole, Regulation of L-alanine dehydrogenase in *Rhizobium leguminosarum* bv. *viciae* and its role in pea nodules. *J. Bacteriol.* **186**, 842–849 (2004).

47. C. Damiani, D. Pescini, R. Colombo, S. Molinari, L. Alberghina, M. Vanoni, G. Mauri, An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes. *Nat. Comput.* **13**, 321–331 (2014).

48. A. J. Gordon, F. R. Minchin, C. L. James, O. Komina, Sucrose synthase in legume nodules is essential for nitrogen fixation. *Plant Physiol.* **120**, 867–878 (1999).

49. E. M. Lodwig, A. H. F. Hosie, A. Bourdès, K. Findlay, D. Allaway, R. Karunakaran, J. A. Downie, P. S. Poole, Amino-acid cycling drives nitrogen fixation in the legume–*rhizobium* symbiosis. *Nature* **422**, 722–726 (2003).

50. J. Prell, J. P. White, A. Bourdes, S. Bunnewell, R. J. Bongaerts, P. S. Poole, Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12477–12482 (2009).

51. P. Mergaert, T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A.-E. Mausset, F. Barloy-Hubler, F. Galibert, A. Kondorosi, E. Kondorosi, Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5230–5235 (2006).

52. E. Ercolano, R. Mirabella, M. Chiurazzi, M. Merrick, The *Rhizobium leguminosarum glnB* gene is down-regulated during symbiosis. *Mol. Gen. Genet.* **264**, 555–564 (2001).

53. A. Heinken, I. Thiele, Anoxic conditions promote species-specific mutualism between gut microbes *in silico. Appl. Environ. Microbiol.*, **81**, 4049–4061 (2015).

54. A. R. Pacheco, M. Moel, D. Segrè, Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* **10**, 103 (2019).

55. L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. J. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. El Assal, L. V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. Ben Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. P. Oliveira, P. T. Vuong, L. P. El Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. Aragón Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Ø. Palsson, I. Thiele, R. M. T. Fleming, Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).

56. Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, B. O. Palsson, Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLOS Comput. Biol.* **11**, e1004321 (2015).

57. C. Damiani, R. Colombo, D. Gaglio, F. Mastroianni, D. Pescini, H. V. Westerhoff, G. Mauri, M. Vanoni, L. Alberghina, A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The WarburQ effect. *PLOS Comput. Biol.* **13**, e1005758 (2017).

58. M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, M. Latorre, N. Loira, G. V. Markov, A. Maass, A. Siegel, Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLOS Comput. Biol.* **14**, e1006146 (2018).

59. G. C. diCenzo, A. B. Benedict, M. Fondi, G. C. Walker, T. M. Finan, A. Mengoni, J. S. Griffitts, Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLOS Genet.* **14**, e1007357 (2018).

60. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. Dejongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. D. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, D. Yu, KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).

61. R. M. Wheatley, V. K. Ramachandran, B. A. Geddes, B. J. Perry, C. K. Yost, P. S. Poole, Role of $O_2$ in the growth of *Rhizobium leguminosarum* bv. viciae 3841 on glucose and succinate. *J. Bacteriol.* **199**, e00572-16 (2017).

62. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

63. P. S. Poole, N. A. Schofield, C. J. Reid, E. M. Drew, D. L. Walshaw, Identification of chromosomal genes located downstream of *dctD* that affect the requirement for calcium and the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* **140**, 2797–2809 (1994).

64. D. Wessel, U. I. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).

65. R Core Team, R: A language and environment for statistical computing (2017); www.r-project.org/.

66. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A.

Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

67. S. Tsukada, T. Aono, N. Akiba, K.-B. Lee, C.-T. Liu, H. Toyazaki, H. Oyaizu, Comparative genome-wide transcriptional profiling of *Azorhizobium caulinodans* ORS571 grown under free-living and symbiotic conditions. *Appl. Environ. Microbiol.* **75**, 5037–5046 (2009).

68. S. K. Masakapalli, N. J. Kruger, R. G. Ratcliffe, The metabolic flux phenotype of heterotrophic *Arabidopsis* cells reveals a complex response to changes in nitrogen supply. *Plant J.* **74**, 569–582 (2013).

69. J. D. Young, INCA: A computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333–1335 (2014).

70. K. Borah (2021), GC-MS datasets for *Azorhizobium caulinodans*, V1, Mendeley Data, https://doi.org/10.17632/y9tmxzv6p9.1.

71. C. Schulte, CarolinSchulte/bacteroid-metabolism: Bacteroid metabolism, v1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.4911594.

72. G. Cheng, R. Karunakaran, A. K. East, O. Munoz-Azcarate, P. S. Poole, Glutathione affects the transport activity of *Rhizobium leguminosarum* 3841 and is essential for efficient nodulation. *FEMS Microbiol. Lett.* **364**, fnx045 (2017).

73. P. Katinakis, R. M. K. Lankhorst, J. Louwerse, A. van Kammen, R. C. van den Bos, Bacteroid-encoded proteins are secreted into the peribacteroid space by *Rhizobium leguminosarum*. *Plant Mol. Biol.* **11**, 183–190 (1988).

74. A. R. Glenn, I. A. McKay, R. Arwas, M. J. Dilworth, Sugar metabolism and the symbiotic properties of carbohydrate mutants of *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **130**, 239–245 (1984).

75. I. A. McKay, A. R. Glenn, M. J. Dilworth, Gluconeogenesis in *Rhizobium leguminosarum* MNF3841. *J. Gen. Microbiol.* **131**, 2067–2073 (1985).

76. P. S. Poole, A. Blyth, C. J. Reid, K. Walters, Myo-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140**, 2787–2795 (1994).

77. J. Fry, M. Wood, P. S. Poole, Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Mol. Plant Microbe Interact.* **14**, 1016–1025 (2001).

78. A. H. F. Hosie, D. Allaway, C. S. Galloway, H. A. Dunsby, P. S. Poole, *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184**, 4071–4080 (2002).

79. W. Małek, M. Kowalski, Auxotrophic mutations related to symbiotic properties of *Rhizobium meliloti* strain L5-30. *Acta Microbiol. Pol.* **26**, 345–350 (1977).

80. P. Aneja, R. Dziak, G.-Q. Cai, T. C. Charles, Identification of an acetoacetyl coenzyme A synthetase-dependent pathway for utilization of L-(+)-3-hydroxybutyrate in *Sinorhizobium meliloti*. *J. Bacteriol.* **184**, 1571–1577 (2002).

81. I. A. McKay, M. J. Dilworth, A. R. Glenn, Carbon catabolism in continuous cultures and bacteroids of *Rhizobium leguminosarum* MNF 3841. *Arch. Microbiol.* **152**, 606–610 (1989).

82. M. F. Dunn, Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Crit. Rev. Microbiol.* **41**, 411–451 (2015).

83. A. S. Craig, K. I. Williamson, Three inclusions of rhizobial bacteroids and their cytochemical character. *Arch. Mikrobiol.* **87**, 165–171 (1972).

84. L. Rosendahl, M. J. Dilworth, A. R. Glenn, Exchange of metabolites across the peribacteroid membrane in pea root nodules. *J. Plant Physiol.* **139**, 635–638 (1992).

85. V. A. Becerra-Rivera, M. F. Dunn, Polyamine biosynthesis and biological roles in rhizobia. *FEMS Microbiol. Lett.* **366**, fnz084 (2019).

86. S. Fujihara, Biogenic amines in rhizobia and legume root nodules. *Microbes Environ.* **24**, 1–13 (2009).

87. I. A. McKay, M. J. Dilworth, A. R. Glenn, $C_4$-dicarboxylate metabolism in free-living and bacteroid forms of *Rhizobium leguminosarum* MNF3841. *J. Gen. Microbiol.* **134**, 1433–1440 (1988).

88. S. Marroquí, A. Zorreguieta, C. Santamaría, F. Temprano, M. Soberón, M. Megías, J. A. Downie, Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. *J. Bacteriol.* **183**, 854–864 (2001).

89. J. F. Moran, I. Iturbe-Ormaetxe, M. A. Matamoros, M. C. Rubio, M. R. Clemente, N. J. Brewin, M. Becana, Glutathione and homoglutathione synthetases of legume nodules. Cloning, expression, and subcellular localization. *Plant Physiol.* **124**, 1381–1392 (2000).

90. M. A. Matamoros, N. Fernández-García, S. Wienkoop, J. Loscos, A. Saiz, M. Becana, Mitochondria are an early target of oxidative modifications in senescing legume nodules. *New Phytol.* **197**, 873–885 (2013).

91. L. Wang, M. J. Wise, Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften* **98**, 719–729 (2011).

92. E. Lodwig, P. Poole, Metabolism of *Rhizobium* bacteroids. *Crit. Rev. Plant Sci.* **22**, 37–78 (2003).

93. M. J. Delgado, E. J. Bedmar, J. A. Downie, Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. *Adv. Microb. Physiol.* **40**, 191–231 (1998).

94. O. Geiger, I. M. López-Lara, Rhizobial acyl carrier proteins and their roles in the formation of bacterial cell-surface components that are required for the development of nitrogen-fixing root nodules on legume hosts. *FEMS Microbiol. Lett.* **208**, 153–162 (2002).

95. P. S. Poole, M. Franklin, A. R. Glenn, M. J. Dilworth, The transport of L-glutamate by *Rhizobium leguminosarum* involves a common amino acid carrier. *J. Gen. Microbiol.* **131**, 1441–1448 (1985).

96. D. L. Walshaw, P. S. Poole, The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that also influences efflux of solutes. *Mol. Microbiol.* **21**, 1239–1252 (1996).

97. J. P. White, J. Prell, V. K. Ramachandran, P. S. Poole, Characterization of a γ-aminobutyric acid transport system of *Rhizobium leguminosarum* bv. *viciae* 3841. *J. Bacteriol.* **191**, 1547–1555 (2009).

98. G. Cheng, R. Karunakaran, A. K. East, P. S. Poole, Multiplicity of sulfate and molybdate transporters and their role in nitrogen fixation in *Rhizobium leguminosarum* bv. *viciae* Rlv3841. *Mol. Plant Microbe Interact.* **29**, 143–152 (2016).

99. D. A. Day, P. S. Poole, S. D. Tyerman, L. Rosendahl, Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell. Mol. Life Sci.* **58**, 61–71 (2001).

## Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

| | |
|---|---|
| Title of Paper | Metabolic control of nitrogen fixation in rhizobium-legume symbioses |
| Publication Status | Published |
| Publication Details | C. C. M. Schulte*, K. Borah*, R. M. Wheatley, J. J. Terpolilli, G. Saalbach, N. Crang, D. H. de Groot, R. G. Ratcliffe, N. J. Kruger, A. Papachristodoulou, P. S. Poole, Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Science Advances* **7** (31), eabh2433 (2021). *equal contribution |

### Student Confirmation

| | | | |
|---|---|---|---|
| Student Name: | Carolin C. M. Schulte | | |
| Contribution to the Paper | performed model reconstruction and analysis, performed measurements of growth rates, oxygen uptake rates and NADH/$NAD^+$ ratios, wrote manuscript | | |
| Signature | Carolin Schulte | Date | 9th August 2021 |

### Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

| | | |
|---|---|---|
| Supervisor name and title: Professor Philip Poole | | |
| Supervisor comments<br><br>A truly original and undoubtably brilliant piece of work, entirely developed by Carolin. | | |
| Signature | Date | 10th August 2021 |

This completed form should be included in the thesis, at the end of the relevant chapter.

# 5. Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses

## Summary

While the focus in Chapter 3 and Chapter 4 is on metabolic modelling of the well-investigated strain Rlv3841, in this chapter I present four additional models for *Rhizobium* and *Sinorhizobium* strains. Models are reconstructed based on homology to the Rlv3841 model presented in Chapter 3 and a published model for *S. meliloti* 1021. After performing extensive reconciliation of model reaction contents, I use gene expression data to obtain bacteroid models for all strains. Model predictions are then compared to investigate differences in nutrient uptake and reaction content for terminally and reversibly differentiated bacteroids. Because of its importance for developing inocula for agricultural application, special consideration is given to the predicted symbiotic efficiency in the different symbioses, which describes the nitrogen output from bacteroids relative to the carbon supply from the plant. The manuscript for this work is in preparation.

Manuscript in preparation

# Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses

Carolin C. M. Schulte[1,2], Antonis Papachristodoulou[2], Philip S. Poole[1]

[1]Department of Plant Sciences, University of Oxford, Oxford, UK

[2]Department of Engineering Science, University of Oxford, Oxford, UK

## Abstract

Rhizobia are soil bacteria that form symbioses with certain legume plants, where rhizobia differentiate into bacteroids that convert atmospheric $N_2$ into ammonia. Although bacteroids across symbioses share several physiological and metabolic characteristics, the types of nodules formed, the reversibility of bacteroid differentiation, and the metabolic flux patterns vary between host plants and rhizobial strains. Unified genome-scale metabolic models for *Rhizobium* and *Sinorhizobium* strains were reconstructed and integrated with gene expression data for bacteroids, resulting in models for six different symbioses. Comparison of model predictions revealed higher metabolic activity and higher symbiotic efficiency in terminally differentiated bacteroids compared to reversibly differentiated bacteroids, which is consistent with experimental evidence. A further rhizobial strain- and host plant-specific difference was identified regarding pathways for pyruvate synthesis from dicarboxylates. The bacteroid models presented herein provide a key tool for investigating metabolic properties of efficient rhizobium-legume interactions to inform the design of rhizobial inocula.

Manuscript in preparation

## Introduction

The symbiosis between rhizobia and legumes provides an estimated 40 million tonnes of bioavailable nitrogen per year and is of major importance for sustainable agriculture (1). Rhizobia induce the formation of nodules on plant roots and differentiate into bacteroids that convert atmospheric $N_2$ gas into ammonia, which is secreted to the plant (2–4). In exchange, legumes provide bacteroids with dicarboxylates as carbon sources (5–7), as well as small amounts of cofactors such as homocitrate (8) and branched-chain amino acids (9). Nodules can be classified into two major groups based on their morphology: (i) the spherical determinate nodules that are predominantly formed by (sub)tropical legumes and (ii) the elongated and often branched indeterminate nodules found on the roots of temperate legumes. Indeterminate nodules have a persistent meristem and are comprised of an infection zone containing undifferentiated rhizobia, a nitrogen fixation zone containing bacteroids, and a senescent zone. Determinate nodules do not have a persistent meristem and are only inhabited by bacteroids performing nitrogen fixation (10, 11). In addition to the distinct nodule morphologies, some legumes induce terminal differentiation in their symbionts. Terminally differentiated (TD) bacteroids are enlarged compared to free-living rhizobia, they contain multiple copies of their genome, and they are unable to resume growth. In contrast, reversibly differentiated (RD) bacteroids have a physiology similar to their free-living counterparts and can be cultivated after extraction from the nodule (12). Both the type of nodule formed and the terminal differentiation of bacteroids are determined by the host plant. Bacteroids of the same rhizobial strain can therefore inhabit both determinate and indeterminate

48 nodules (13) and undergo terminal differentiation in symbiosis with some plants but
49 not with others (14).
50 Irrespective of the rhizobial strain and host plant, bacteroids are non-dividing and
51 catabolize dicarboxylates to perform nitrogen fixation at very low oxygen levels. These
52 shared properties result in several common features of bacteroid metabolism as
53 reviewed in our recent meta-analysis of transcriptome and proteome datasets for
54 various symbioses (4). However, both mutant and transcriptome studies have
55 revealed differences in metabolic requirements and/or pathway usage between
56 symbioses (13, 15–19), warranting careful evaluation of the generality of results
57 obtained for a specific rhizobium-legume pairing.
58 To complement experimental efforts, computational approaches such as metabolic
59 modelling are increasingly used to investigate rhizobium-legume symbioses (4, 20–
60 26). Metabolic models are knowledge bases containing the reactions that can occur in
61 an organism, where each reaction is associated with one or multiple genes encoding
62 the enzyme that catalyses the reaction (27, 28). By defining a set of metabolites that
63 can be exchanged with the environment as well as an objective function, metabolic
64 flux distributions can be calculated using flux balance analysis (29). While early studies
65 using metabolic models mainly focused on investigating properties of single species
66 in defined growth conditions, the range of applications has expanded to include
67 microbial communities (30–33), host-symbiont interactions (26, 34, 35) and
68 comparative studies, both on the species and the strain level (36–38). A commonly
69 encountered issue when comparing predictions of metabolic models for related
70 species is the presence of differences that are introduced as part of the model
71 reconstruction process, but do not have biological relevance. Possible reasons for

these artefactual differences include the use of different reconstruction algorithms and reference databases, or choices made as part of curation and gap-filling steps (39). Performing a meaningful comparison of closely related organisms therefore requires extensive model reconciliation (36, 37).

Several metabolic models have been developed to investigate bacteroids of various rhizobial species, namely *Rhizobium etli* (23, 24, 40), *Sinorhizobium meliloti* (25, 26, 41), *Bradyrhizobium japonicum* (42), *Sinorhizobium fredii* (21), and *Rhizobium leguminosarum* (22, 43). However, these reconstructions vary in their scope, the annotation of genes, reactions and metabolites, and the level of manual curation that has been performed, hindering a model-based comparison of rhizobial metabolism in the different symbioses. Based on two published genome-scale metabolic models (GSMs) for different rhizobia, we reconstructed and curated models for four *Rhizobium* and two *Sinorhizobium* strains. We unified the scope and nomenclature for all models to minimise the occurrence of differences stemming from the reconstruction process and subsequently constrained the GSMs with RNA-Seq data to obtain models specific for bacteroids. This allowed us to compare bacteroid metabolism in six different rhizobium-legume pairings, highlighting common features across symbioses as well as differences regarding metabolic pathway usage and symbiotic efficiency. We anticipate that the bacteroid models presented herein will provide a valuable resource for the interpretation of strain- and host plant-specific mutant studies.

## Results

### Genome comparison of *Rhizobium* and *Sinorhizobium* strains

The strains chosen for model reconstruction in this study were *S. meliloti* 1021 (Sm1021), *R. leguminosarum* bv. *viciae* 3841 (Rlv3841), *R. etli* CFN 42 (Re42), *S. fredii* NGR234 (Sf234), *R. leguminosarum* bv. *viciae* A34 (RlvA34) and *R. leguminosarum* bv. *phaseoli* 4292 (Rlp4292) (Table 1). The choice of strains was made based on several factors: (i) genome sequences are available for all strains (15, 44–47) (Table S1), (ii) among rhizobial strains, Sm1021, Rlv3841 and Re42 are well-investigated and a substantial body of literature describing their metabolism is available, (iii) curated GSMs are available for Sm1021 (26) and Rlv3841 (43) and (iv) gene expression data for bacteroids of these strains are available. Importantly, Sf234 bacteroids can inhabit determinate and indeterminate nodules depending on the plant host (13), but do not terminally differentiate in either symbiosis. RlvA34 and Rlp4292 differ only in their symbiosis plasmid but form indeterminate and determinate nodules inhabited by TD and RD bacteroids, respectively, thus enabling direct comparison between the different bacteroid forms (15). To assess the feasibility of template-based model reconstruction, we determined genomic similarity between the strains by performing genome comparison based on clusters of orthologous genes using OrthoVenn2 (48). We found a core of 2882 clusters of orthologous genes that were shared between all rhizobial strains accounting for between 44% of clusters in Rlp4292 and 61% of clusters in Sf234 (Fig. 1). A further 2228 clusters were shared between at least four strains, and 644 clusters shared between the *Sinorhizobium* strains Sm1021 and Sf234 (Fig. S1). No enrichment of gene ontology (GO) categories was observed for the clusters shared between *Rhizobium* strains, whereas the clusters shared by

the two *Sinorhizobium* strains were significantly enriched in polysaccharide synthesis (GO:0000271), phenylacetate catabolism (GO:0010124), pyrroquinoline quinone synthesis (GO:0018189) and one-carbon metabolism (GO:0006730), and clusters shared by *R. leguminosarum* strains were enriched in transmembrane transport processes (GO:0055085). Sf234 and Sm1021 had the highest number of proteins for which no ortholog was found, accounting for 18% of all proteins annotated in their genomes, followed by 12% for Rlv3841, 8% for Re42, 4% for RlvA34 and 2% for Rlp4292. 39-40% of the genes without orthologs in Sf234 and Sm1021 were either annotated as hypothetical proteins or involved in transport processes. Overall, this analysis indicated sufficient genomic similarity between the strains to perform homology-based reconstruction of metabolic models.

**Model reconstruction and evaluation**

We next reconstructed GSMs for all rhizobial strains investigated in this study. Briefly, starting from curated GSMs for Sm1021 (26) and Rlv3841 (43), we reconstructed models for all strains based on bidirectional best hit BLAST searches as described previously (49). Reactions from a bacteroid model for Re42 (24) were also evaluated and added if appropriate. The models were then curated to ensure *in silico* growth in minimal media conditions, consistency with experimentally determined carbon source utilization (Fig. 2A) and unified nomenclature, stoichiometry and reversibility of all reactions as described in the Methods section (Table 2). All models were evaluated using MEMOTE (50) and scored 99-100% in the consistency category, indicating mass and charge balance as well as stoichiometric consistency (Supplementary Data 1). Jaccard similarity indices were calculated based on model reactions for all model

pairs (Fig. 2B), showing a clear separation between *Rhizobium* and *Sinorhizobium* models. The reaction content of the GSMs for RlvA34 and Rlp4292 is identical, which was expected because the genomes of these two strains only differ in their symbiosis plasmids (15). The six models contained a total of 1110 unique metabolic reactions, excluding transport and exchange reactions as well as biomass-related reactions such as DNA or RNA synthesis. In agreement with the high similarity on a genomic level, the six strain-specific models shared a core of 826 metabolic reactions. A further 36 reactions were shared among all *Rhizobium* strains and 140 reactions were specific to the two *Sinorhizobium* strains. Apart from a few reactions involved in amino acid and carbon metabolism, most differences between the *Rhizobium* and *Sinorhizobium* models were related to the biosynthesis of cell envelope components, such as lipopolysaccharides or exopolysaccharides, which is consistent with the GO enrichment determined by genome analysis. Comparing carbon source utilization data obtained by phenotype microarray analysis (51) for Rlv3841, Re42, Sm1021 and Sf234 further confirmed high metabolic similarity (Fig. 2A). Carbon source utilization agreed for at least three of the four strains for 74 out of 81 tested carbon sources. Comparing model predictions with experimentally determined growth phenotypes resulted in accuracies of 90.7%, 93.7%, 90.7%, 82.1% for Rlv3841, Re42, Sm1021 and Sf234, respectively (Fig. 2C). The comparatively low accuracy for Sf234 was due to growth of this strain on several unusual carbon sources, such as melezitose or *p*-hydroxyphenylacetate, for which catabolic pathways have not been elucidated so far.

## Reconstruction of bacteroid-specific models

To develop bacteroid-specific models based on gene-expression data, we limited our analysis to strains for which RNA-Seq data for bacteroids are available, which excluded Re42. Similar to our previous study of Rlv3841 (43), we used the RIPTiDe algorithm (52) to obtain bacteroid models by integrating gene expression data with the GSMs. Flux sampling was then performed to analyse nutrient uptake and metabolic pathway usage, and a reaction was generally considered active if it had a non-zero median flux value across 500 flux samples. Due to the different availability of nodule metabolome information for the investigated strains, we set the boundary conditions for all strains to allow for unlimited uptake of metabolites which are essential for bacteroids across symbioses. In particular, this included the dicarboxylates malate and succinate as carbon sources (5–7) as well as various trace metals and cofactors (Table S2). All other exchange reactions were constrained to allow for low levels of uptake of the remaining external metabolites for which transporters were present in the models. Bacteroid models were then obtained using published RNA-Seq data for bacteroids of the different strains (Table S3). In addition, removal of genes that were upregulated in bacteroids compared to free-living bacteria was prevented during data integration. For Sf234, RNA-Seq data were used for determinate nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena leucocephala*, and the two different bacteroid models are henceforth referred to as Sf234_Vu and Sf234_Ll, respectively. While all bacteroid models were reduced in size compared to the full GSMs, the bacteroid models for RD bacteroids retained a higher fraction of the original reaction content (52%, 54% and 58% for Rlp4292, Sf234_Ll and Sf234_Vu, respectively) compared to those for TD bacteroids (18%, 29% and 42% for Sm1021,

188 Rlv3841 and RlvA34, respectively). This difference may be related to the higher
189 metabolic complexity of RD bacteroids, which show physiological characteristics
190 similar to free-living bacteria, in contrast to the enlarged TD bacteroids (53). The
191 pattern of Jaccard similarity indices for the bacteroid models differed noticeably from
192 the original GSMs (Fig. 3A). The two models for Sf234 bacteroids had the highest
193 similarity index of 0.83, which was expected since they were obtained from the same
194 GSM and both describe RD bacteroids. The Jaccard similarity of the models for
195 RlvA34 and Rlp4292, which also originated from GSMs with identical reaction content,
196 decreased to 0.71, indicating substantial differences between the models describing
197 RD and TD bacteroids. Similarly, the Rlp4292 model had higher similarity to the Sf234
198 models than the RlvA34 model and the reaction content of the Sm1021 model was
199 now more similar to the models for Rlv3841 and RlvA34 than the models for Sf234.
200 We next evaluated the similarity of predicted metabolic fluxes for all models by
201 performing principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity of flux
202 samples for those reactions that were shared between all models (Fig. 3B). Although
203 flux samples for the different bacteroid models overlapped in many cases, the models
204 for TD bacteroids of *Rhizobium* strains (Rlv3841 and RlvA34) were more similar than
205 the Rlp4292 model for RD bacteroids. The flux samples for Sm1021 had some overlap
206 with those for Sf234_Vu, but were completely separated from those for Sf234_Ll.
207 While the flux samples for Sf234_Vu overlapped with those of the *Rhizobium* models,
208 flux samples for the Sm1021 and the Sf234_Ll models clustered largely separately
209 from those for the *Rhizobium* models, indicating some metabolic differences between
210 *Rhizobium* and *Sinorhizobium* bacteroids that were captured by the models. The
211 sampled flux distributions for RD bacteroids were more spread out compared to the

TD bacteroid models, which can be explained by the higher number of reactions in those models giving rise to a higher number of possible flux distributions. Overall, these results indicate successful creation of condition-specific models whose reaction content differs according on the type of bacteroid differentiation and the host plant.

**Shared metabolic features of bacteroid models**

To analyse metabolic features of the bacteroid models, we first compared nutrient uptake and secretion reactions that were predicted to be active across flux samples. As expected, all models predicted dicarboxylates as the main carbon source. Uptake of malate was favoured over succinate, which is consistent with experimental data showing that mutants in Rlv3841 and Sm1021 unable to take up succinate are not impaired in symbiosis (6). In addition to dicarboxylates, low levels of sugar uptake were predicted only by the models for indeterminate nodules. Indeterminate nodules have characteristic developmental zones and contain both undifferentiated bacteria and nitrogen-fixing bacteroids, whereas bacteroids in determinate nodules are physiologically identical (53, 54). Both experimental and computational studies have indicated that undifferentiated bacteria in indeterminate nodules utilise sugars instead of or in addition to dicarboxylates as a carbon source (5, 26, 43). The predicted sugar uptake by bacteroids in indeterminate nodules may thus be an artefact related to the isolation of RNA from whole nodules, which would contain rhizobia at a range of developmental stages. In addition to dicarboxylates, all models predicted uptake of oxygen, $N_2$ gas, trace metals and homocitrate, as well as secretion of $CO_2$, ammonia and/or alanine and $H_2$. Thiamine was generally predicted to be taken up by bacteroids, however, the corresponding biosynthetic pathway was retained in the Rlp4292 model.

While Rlv3841 only has a salvage pathway for thiamine synthesis (55) and is therefore probably dependent on plant supply of this vitamin, expression of thiamine biosynthetic genes has been linked to expression of the *fixNOQP* genes in *R. etli* (56). Thiamine biosynthesis genes were also highly expressed in Sf234 bacteroids, especially in symbiosis with *V. unguiculata* (13). This suggests that thiamine supply either varies depending on the host plant, or that RD bacteroids have a higher requirement for thiamine. All models further predicted uptake of at least one intermediate of purine metabolism, such as adenosine, guanosine and inosine. According to our previously reported metabolome data (43), several metabolites involved in purine metabolism are abundant in the nodule cytosol of pea plants. In addition, levels of guanine, adenine and 2'-deoxyinosine are 5-, 14- and 16-fold higher in bacteroids compared to free-living rhizobia. While this could be related to DNA endoreduplication occurring as part of the terminal differentiation process in bacteroids (53), increased DNA synthesis was not accounted for in the models and no difference between purine uptake rates was observed between the models for TD and RD bacteroids. Supply of purine intermediates may thus have a yet unknown function in the symbiosis. Rhizobial purine auxotrophs are generally defective in symbiosis formation, however synthesis of purine metabolites such as AICAR seems to be required at early stages of infection in many symbioses (57–59), and it is therefore unclear to what extent purine synthesis is required in mature bacteroids. Apart from transport of nutrients, metabolic reactions predicted to be active in all models mainly included nitrogen fixation and the tricarboxylic acid (TCA) cycle. In addition, several biosynthetic pathways for heme, cobalamin, pyridoxin and riboflavin were predicted to be active by all models, in

agreement with mutant studies indicating a requirement for these cofactors in bacteroids (60–62).

**Distinct symbiotic efficiency of TD and RD bacteroids**

We next evaluated differences in symbiotic efficiency between symbioses. Both absolute nitrogenase activity and the carbon input required from the plant to obtain a certain level of ammonia output from the bacteroid are critical parameters when developing efficient rhizobial inocula for agricultural applications (63, 64). As shown in Figs. 4A and 4B, TD bacteroids were generally predicted to have higher uptake rates for dicarboxylates and oxygen compared to RD bacteroids of the same species. Endoreduplication, i.e., genome replication without cell division, leads to higher metabolic activity in plant cells (65), and a similar effect may occur in TD bacteroids (66). It has further been suggested that the higher density of RD bacteroids in nodules could restrict their access to oxygen, thus limiting the achievable rate of nitrogenase activity (67). Accordingly, predicted nitrogenase activity was also significantly higher for TD bacteroids (Fig. 4C). Symbiotic efficiency was determined as the ratio of catabolized dicarboxylates and nitrogenase activity in each flux sample, thus giving an indication of the carbon investment per nitrogen output that is required from the plant (Fig. 4D). Significantly lower symbiotic efficiency was predicted for Rlp4292 compared to RlvA34, which agrees with experimental measurements of symbiotic efficiency for RlvA34 in symbiosis with pea and bean (17). Similarly, both Sf234 symbioses with RD bacteroids were predicted to be less efficient than the TD Sm1021 bacteroids. The higher efficiency of nitrogen fixation predicted for TD bacteroids was mainly due to reduced diversion of carbon into polymers such as polyhydroxybutyrate and glycogen,

but also cell envelope components, compared to RD bacteroids (Fig. 5). Electron micrographs of both Rlp4292 and Sf234 bacteroids show substantial accumulation of PHB (13, 69), which is not observed in TD bacteroids (69–71). The optimization of metabolism for maximum nitrogen provision to the plant may be one of the major selective advantages of terminal differentiation (17, 66) and could explain why this trait evolved multiple times independently (68). While the predicted nitrogenase activity was lower for Sf234_Ll compared to Sf234_Vu, the symbiotic efficiency of both strains did not differ significantly, indicating that the differentiation state of bacteroids rather than the nodule type determines the efficiency of the symbiosis. In addition, the broad host range strain Sf234 has been found to form less efficient nodules on *Lotus japonicus* compared to *Mesorhizobium loti*, which is the natural symbiont of *Lotus* spp. and also forms RD bacteroids (72). Apart from reduced metabolic activity and polymer synthesis in RD bacteroids, the low nitrogenase activity predicted for Sf234 may thus also be a result of a lower degree of adaptation to a specific plant host.

**Differences in central carbon metabolism**

Finally, we determined differences in predicted metabolic flux distributions using the AUC-Random Forest algorithm (73) for pairwise comparison of the flux samples of all bacteroid models. The mean decrease accuracy was evaluated to determine which reactions most distinguished each pair of models (Supplementary Data 2). This analysis confirmed the differences in metabolic activity described in the previous section. Several TCA cycle enzymes were differentially used between TD and RD bacteroids, where TD bacteroids consistently had higher predicted fluxes through these enzymes. Two reactions of central carbon metabolism that discriminated the

Rlv3841 from the Sm1021 model, both of which describe TD bacteroids, were malic enzyme and pyruvate kinase, where Rlv3841 had higher fluxes through the pyruvate kinase reaction, but lower fluxes through the malic enzyme reaction. It is known that *R. leguminosarum* bacteroids synthesize pyruvate from dicarboxylates using both malic enzyme and the combined action of phosphoenolpyruvate carboxykinase and pyruvate kinase, whereas *S. meliloti* bacteroids only express malic enzyme (16). The higher flux through malic enzyme for Sm1021 can thus be explained because it is the only pathway for pyruvate synthesis in contrast to the two alternative pathways in Rlv3841. The phosphoenolpyruvate carboxykinase reaction was correctly pruned from the Sm1021 bacteroid model. Malic enzyme was predicted to be the preferred pathway for pyruvate synthesis for Sf234 in symbiosis with *V. unguiculata*, but not with *L. leucocephala*. Although Sf234 can express the phosphoenolpyruvate carboxykinase/pyruvate kinase pathway, its contribution to pyruvate formation in symbiosis varies greatly depending on the plant host. In agreement with the model predictions, *V. unguiculata* inoculated with a malic enzyme (*dme*) mutant of Sf234 had a 65% reduction in dry weight compared to plants inoculated with the wild-type, while only a 21% reduction was observed for *L. leucocephala* (74).

Reactions distinguishing TD RlvA34 and RD Rlp4292 bacteroids particularly included synthesis of cell membrane components such as lipopolysaccharides, which carried higher fluxes in Rlp4292. This is consistent with the 70-fold upregulation of *lpxH* and *lpxQ* in Rlp4292 bacteroids (15). A distinguishing feature of the Sf234_Vu bacteroid model was the activity of acetate kinase, consistent with the *ackA* gene being 182-fold upregulated in bacteroids (13). Acetate kinase activity has been detected in *B. japonicum* bacteroids (75), but mutants in *S. meliloti* were unaffected in symbiosis

with alfalfa (76). The operon containing *ackA* in Sf234 is organized in a similar way to that of *S. meliloti*, which is induced in response to phosphate starvation (76), and upregulation of acetate kinase could therefore also be a result of phosphate limitation, which would agree with the increased expression of phosphate transporters in Sf234 bacteroids (13).

## Discussion

In this study, we reconstructed four GSMs for *Rhizobium* and *Sinorhizobium* strains. The models were based on curated GSMs for Rlv3841 and Sm1021 and extensive unification of reaction content and nomenclature was performed to enable direct comparison of model predictions. As the number of available reconstruction tools and databases increases, the need for reconciliation of model contents is increasingly recognized, especially when performing comparative studies (37, 77). Several models that have been reconstructed for rhizobia so far exclusively focus on the bacteroid stage (21–24, 41) and therefore have a limited scope which is biased towards the inclusion of reactions for which experimental evidence exists. Model curation has also been performed to varying extents and different objective functions have been used, which has hindered direct comparisons. Addressing these limitations, we obtained bacteroid-specific models by integrating RNA-Seq data with unified GSMs using the same constraints for all rhizobial strains investigated in this study. Reaction content and distribution of metabolic fluxes showed that integration of gene expression data was sufficient to reveal differences between bacteroids in symbiosis with different plants even when identical GSMs were used for data integration. When comparing metabolic properties for TD and RD bacteroids of closely related strains, we found that

355 models for RD bacteroids retained a higher number of reactions, in particular synthesis
356 pathways for cell envelope components and storage polymers. At the same time, our
357 analysis revealed both reduced overall metabolic activity and reduced symbiotic
358 efficiency of RD compared to TD bacteroids. These results are consistent with an
359 experimental study comparing nitrogenase activity relative to respiratory activity of
360 bacteroids for TD and RD bacteroids of the same rhizobial strains nodulating different
361 hosts (17) and further support the hypothesis that plant-induced terminal differentiation
362 of bacteroids evolved multiple times due to its benefits regarding symbiotic efficiency
363 (68). Machine learning was able to reveal experimentally supported differences in
364 central carbon metabolism, identifying malic enzyme as the sole pathway for pyruvate
365 synthesis in Sm1021 and correctly predicting varying contributions of
366 phosphoenolpyruvate carboxykinase and pyruvate kinase in Sf234 on different host
367 plants (74). While our analysis identified several differences between TD and RD
368 bacteroids which are consistent with experimental evidence, more extensive studies
369 comparing a larger number of rhizobial strains and host plants will be required to
370 investigate metabolic characteristics based on sufficiently large sample sizes. In
371 addition, to ensure comparability of RNA-Seq data across plants and rhizobial strains,
372 it would be important to establish standardized plant growth conditions to avoid non-
373 meaningful differences that could arise from different phosphate and trace metal
374 availability in the plant growth medium used.
375 Rhizobial inocula for agricultural applications should ideally have both high activity and
376 efficiency, i.e., the maximum possible amount of nitrogen should be fixed in return for
377 the carbon investment by the plant. Symbiotic efficiency can be difficult to determine
378 based on measuring nitrogenase activity alone, since this only provides a snapshot of

379 nitrogen fixation activity, which is dependent on the time of the measurement and does
380 not indicate the amount of carbon supplied by the plant (17, 74). The good agreement
381 of our model predictions with experimental studies suggests that model-based
382 comparison of metabolic activity in bacteroids may be a useful tool to identify and
383 engineer rhizobial strains with optimum symbiotic efficiency.

## Methods

### Phenotype Microarray analysis

Re42 was grown at 28 °C in universal minimal salts (UMS) media supplemented with 10 mM glucose and 10 mM ammonium chloride as a carbon and nitrogen source, respectively, as well as thiamine, biotin and pantothenic acid (78). When the culture reached an optical density at 600 nm ($OD_{600}$) of 0.6, the cells were spun down and washed three times in UMS media without supplements. Cells were then resuspended in UMS supplemented with ammonium chloride and vitamins and incubated for 6 h to ensure carbon starvation. The culture was then adjusted to an $OD_{600}$ of 0.1 and 100 μL of cell culture were added to each well of a PM1 Phentotype MicroArray (Biolog, Hayward, CA). 10 μL of a 0.1% (w/v) solution of 2,3,5-triphenyltetrazolium chloride in water were then added to each well and the plate was incubated in an Omega FluoStar plate reader at 28 °C with double orbital shaking at 500 rpm. The absorbance at 505 nm was measured every 15 min. Data analysis was performed using DuctApe (79), and a carbon source was considered to support growth when the activity value was higher than the activity value of the water control. Phenotype microarray data for Re42 are provided as Supplementary Data 3.

### Analysis of orthologous genes

All genome sequences were obtained from NCBI RefSeq, apart from RlvA34, a synthetic strain whose annotated genome sequence was obtained from a recent transcriptome study (15) (Table S1). Clusters of orthologous genes were determined using the web interface of OrthoVenn2 (48), with e-value $1e^{-5}$ and inflation 1.5.

408 **General modelling procedures**

409 General model analysis was performed using scripts from the COBRA Toolbox v.3.0
410 (80) in MATLAB R2020b. The Python implementation of RIPTiDe
411 (https://github.com/mjenior/riptide) was used as previously described (52). All models,
412 scripts and Supplementary Data are available on Github
413 (https://github.com/CarolinSchulte/StrainComparison).
414
415 **Statistical methods**
416 AUC-Random Forest was implemented using the AUCRF package (73) in R v4.1.0
417 (81). PCoA was performed in R and Kruskal-Wallis test was performed in GraphPad
418 Prism v9.0.0.
419
420 **Model reconstruction**
421 Models were reconstructed based on a protocol for multi-strain model reconstruction
422 (49). Briefly, bidirectional protein BLAST searches were performed for Sm1021 versus
423 all other strains and Rlv3841 versus all other strains. Using the GSMs iGD1348 (26)
424 and iCS1224 (43) for Sm1021 and Rlv3841 as references, draft models for all other
425 strains were then obtained by determining the presence or absence of a reaction
426 based on the presence of a bidirectional best hit with at least 60% identity. Due to a
427 lack of information on strain-specific biomass composition, the biomass reaction from
428 iGD1348 was used for Sf234 and the biomass reaction from iCS1224 was used for all
429 *Rhizobium* strain models. The composition of protein, DNA and RNA was adjusted for
430 each strain based on its genome sequence. The reconstructions obtained from
431 iCS1224 and iGD1348 were then merged for each strain. iCS1224 was originally

reconstructed based on iGD1348 and the models therefore share many reactions. However, several reaction identifiers as well as constraints on reaction directionality had to be reconciled, which was done using the ModelSEED database as a reference (82). Next, reactions in a previously published model for Re42 (24) were added to the Re42 model when an equivalent reaction was not already present. Bidirectional BLAST searches were again performed for all strains versus the Re42 genome and reactions were added to the models if appropriate. The gene associations for DNA, RNA, protein and biomass synthesis in iCS1224 and iGD1348 are based on gene essentiality data, which were not available for the remaining strains. Gene associations for these reactions were therefore removed to prevent expression of those genes from skewing the results of the data integration. The models were then individually curated to ensure flux through the biomass reaction in simulated UMS media (78) and further curation was performed based on carbon source utilization data determined by phenotype microarray experiments. All models were evaluated using MEMOTE (50).

**Data integration using RIPTiDe**

Bacteroid models were obtained by integrating the GSMs for each strain with published RNA-Seq datasets for bacteroids (Table S3). The lower bounds of exchange reactions for all metabolites with universal importance across rhizobium-legume symbioses were set to -1000 to allow for unlimited uptake of these metabolites (Table S2). In addition, lower bounds for all other exchange reactions were set to -1 in each model and the nitrogenase reaction was used as the objective function. Transcriptome data were then integrated using the RIPTiDe algorithm, while a list of genes that were

upregulated in bacteroids compared to free-living rhizobia was specified for each model to prevent removal of those genes from the bacteroid model. The only exception was the model for Sm1021, where no solution with significant correlation between transcriptome data and predicted fluxes was found when upregulated genes were included. For Rlv3841, no RNA-Seq data for free-living bacteria were available, and we therefore used a list of proteins that were increased in abundance in bacteroids compared to free-living bacteria (22). During data integration, constraining flux through the objective reaction to values between 20% and 95% of its maximum was tested, and the bacteroid model with the highest correlation between transcriptome data and flux predictions was used in further analyses.

1. Herridge DF, Peoples MB, Boddey RM, Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 311:1–18.
2. Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* 64:781–805.
3. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* 16:291–303.
4. Ledermann R, Schulte CCM, Poole PS. 2021. How rhizobia adapt to the nodule environment. *J Bacteriol* 203:e00539–20.
5. Ronson CW, Lyttleton P, Robertson JG. 1981. $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc Natl Acad Sci U S A* 78:4284–4288.
6. Mitsch MJ, DiCenzo GC, Cowie A, Finan TM. 2018. Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. *Appl Environ Microbiol* 84:e01561-17.
7. Finan TM, Wood JM, Jordan DC. 1983. Symbiotic properties of $C_4$-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J Bacteriol* 154:1403–1413.
8. Hakoyama T, Niimi K, Watanabe H, Tabata R, Matsubara J, Sato S, Nakamura Y, Tabata S, Jichun L, Matsumoto T, Tatsumi K, Nomura M, Tajima S, Ishizaka M, Yano K, Imaizumi-Anraku H, Kawaguchi M, Kouchi H, Suganuma N. 2009. Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* 462:514–517.
9. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009.

490 Legumes regulate *Rhizobium* bacteroid development and persistence by the
491 supply of branched-chain amino acids. *Proc Natl Acad Sci U S A* 106:12477–
492 12482.
493 10. Hirsch AM. 1992. Developmental biology of legume nodulation. *New Phytol*
494 122:211–237.
495 11. Brewin NJ. 1991. Development of the legume root nodule. *Annu Rev Cell Biol*
496 7:191–226.
497 12. Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE,
498 Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control
499 on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis.
500 *Proc Natl Acad Sci U S A* 103:5230–5235.
501 13. Li Y, Tian CF, Chen WF, Wang L, Sui XH, Chen WX. 2013. High-resolution
502 transcriptomic analyses of *Sinorhizobium* sp. NGR234 bacteroids in determinate
503 nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena*
504 *leucocephala*. *PLoS One* 8:e70531.
505 14. Sen D, Weaver RW. 1980. Nitrogen fixing activity of rhizobial strain 32H1 in
506 peanut and cowpea nodules. *Plant Sci Lett* 18:315–318.
507 15. Green RT, East AK, Karunakaran R, Downie JA, Poole PS. 2019.
508 Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate
509 and indeterminate nodules. *Microb Genomics* 5:e000254.
510 16. Mulley G, Lopez-Gomez M, Zhang Y, Terpolilli J, Prell J, Finan T, Poole P. 2010.
511 Pyruvate is synthesized by two pathways in pea bacteroids with different
512 efficiencies for nitrogen fixation. *J Bacteriol* 192:4944–4953.
513 17. Oono R, Denison RF. 2010. Comparing symbiotic efficiency between swollen

versus nonswollen rhizobial bacteroids. *Plant Physiol* 154:1541–1548.

18. Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS. 2009. Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J Bacteriol* 191:4002–4014.

19. Prell J, Bourdès A, Kumar S, Lodwig E, Hosie A, Kinghorn S, White J, Poole P. 2010. Role of symbiotic auxotrophy in the rhizobium-legume symbioses. *PLoS One* 5:e13933.

20. DiCenzo GC, Zamani M, Checcucci A, Fondi M, Griffitts JS, Finan TM, Mengoni A. 2019. Multidisciplinary approaches for studying rhizobium–legume symbioses. *Can J Microbiol* 65:1–33.

21. Contador CA, Lo S-K, Chan SHJ, Lam H-M. 2020. Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* 5:e00516–19.

22. Schulte CCM, Borah K, Wheatley RM, Terpolilli JJ, Saalbach G, Crang N, de Groot DH, Ratcliffe RG, Kruger NJ, Papachristodoulou A, Poole PS. 2021. Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Sci Adv* 7:eabh2433.

23. Resendis-Antonio O, Reed JL, Encarnación S, Collado-Vides J, Palsson B. 2007. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput Biol* 3:1887–1895.

24. Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. 2011. Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based

538 modeling. *BMC Syst Biol* 5:120.
539 25. DiCenzo GC, Checcucci A, Bazzicalupo M, Mengoni A, Viti C, Dziewit L, Finan
540 TM, Galardini M, Fondi M. 2016. Metabolic modelling reveals the specialization
541 of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat*
542 *Commun* 7:1–10.
543 26. DiCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale
544 metabolic reconstruction of the symbiosis between a leguminous plant and a
545 nitrogen-fixing bacterium. *Nat Commun* 11:1–11.
546 27. Thiele I, Palsson B. 2010. A protocol for generating a high-quality genome-scale
547 metabolic reconstruction. *Nat Protoc* 5:93–121.
548 28. Nielsen J. 2017. Systems biology of metabolism. *Annu Rev Biochem* 86:245–
549 275.
550 29. Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nat*
551 *Biotechnol* 28:245-248.
552 30. Succurro A, Ebenhöh O. 2018. Review and perspective on mathematical
553 modeling of microbial ecosystems. *Biochem Soc Trans* 46:403–412.
554 31. Heinken A, Basile A, Thiele I. 2021. Advances in constraint-based modelling of
555 microbial communities. *Curr Opin Syst Biol*, in press.
556 32. Heinken A, Thiele I. 2015. Anoxic conditions promote species-specific
557 mutualism between gut microbes *in silico*. *Appl Environ Microbiol* 81:4049–
558 4061.
559 33. Pacheco AR, Moel M, Segrè D. 2019. Costless metabolic secretions as drivers
560 of interspecies interactions in microbial ecosystems. *Nat Commun* 10:1–12.
561 34. Pfau T, Christian N, Masakapalli SK, Sweetlove LJ, Poolman MG, Ebenhöh O.

562 2018. The intertwined metabolism during symbiotic nitrogen fixation elucidated
563 by metabolic modelling. *Sci Rep* 8:12504.
564 35. Ankrah NYD, Luan J, Douglas AE. 2017. Cooperative metabolism in a three-
565 partner insect-bacterial symbiosis revealed by metabolic modeling. *J Bacteriol*
566 199:e00872–16.
567 36. Bartell JA, Yen P, Varga JJ, Goldberg JB, Papin JA. 2014. Comparative
568 metabolic systems analysis of pathogenic *Burkholderia*. *J Bacteriol* 196:210–
569 226.
570 37. Oberhardt MA, Puchałka J, Martins dos Santos VAP, Papin JA. 2011.
571 Reconciliation of genome-scale metabolic reconstructions for comparative
572 systems analysis. *PLoS Comput Biol* 7:e1001116.
573 38. Lee D-S, Burd H, Liu J, Almaas E, Wiest O, Barabási A-L, Oltvai ZN, Kapatral
574 V. 2009. Comparative genome-scale metabolic reconstruction and flux balance
575 analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial
576 drug targets. *J Bacteriol* 191:4015–4024.
577 39. Bernstein DB, Sulheim S, Almaas E, Segrè D. 2021. Addressing uncertainty in
578 genome-scale metabolic model reconstruction and analysis. *Genome Biol*
579 22:64.
580 40. Resendis-Antonio O, Hernández M, Mora Y, Encarnación S. 2012. Functional
581 modules, structural topology, and optimal activity in metabolic networks. *PLoS*
582 *Comput Biol* 8:e1002720.
583 41. Zhao H, Li M, Fang K, Chen W, Wang J. 2012. *In silico* insights into the symbiotic
584 nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PLoS*
585 *One* 7:e31287.

586 42. Yang Y, Hu X-P, Ma B-G. 2017. Construction and simulation of the
587 *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison
588 between free-living and symbiotic states. *Mol BioSyst* 13:607.
589 43. Schulte CCM, Ramachandran VK, Papachristodoulou A, Poole PS. 2021.
590 Genome-scale metabolic modelling of lifestyle changes in *Rhizobium*
591 *leguminosarum*. *bioRxiv* https://doi.org/10.1101/2021.07.28.454262.
592 44. Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F,
593 Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J,
594 Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA,
595 Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M,
596 Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML,
597 Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC,
598 Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P,
599 Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J.
600 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*.
601 *Science* 293:668–672.
602 45. Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH,
603 Wexler M, Curson ARJ, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA,
604 Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A,
605 Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K,
606 Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S,
607 Parkhill J. 2006. The genome of *Rhizobium leguminosarum* has recognizable
608 core and accessory components. *Genome Biol* 7:R34.
609 46. Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quere A, Wollherr A,

610 Heinemeyer I, Morgenstern B, Pommerening-Röser A, Flores M, Palacios R,
611 Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW,
612 Streit WR. 2009. *Rhizobium* sp. strain NGR234 possesses a remarkable number
613 of secretion systems. *Appl Environ Microbiol* 75:4035–4045.
614 47. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto
615 A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-
616 Vides J, Dávila G. 2006. The partitioned *Rhizobium etli* genome: genetic and
617 metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A*
618 103:3834–3839.
619 48. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr
620 D, Xia Q, Wang Y. 2019. OrthoVenn2: a web server for whole-genome
621 comparison and annotation of orthologous clusters across multiple species.
622 *Nucleic Acids Res* 47:W52–W58.
623 49. Norsigian CJ, Fang X, Seif Y, Monk JM, Palsson BO. 2020. A workflow for
624 generating multi-strain genome-scale metabolic models of prokaryotes. *Nat*
625 *Protoc* 15:1–14.
626 50. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell
627 JA, Blank LM, Chauhan S, Correia K, Diener C, Dräger A, Ebert BE, Edirisinghe
628 JN, Faria JP, Feist AM, Fengos G, Fleming RMT, García-Jiménez B,
629 Hatzimanikatis V, van Helvoirt W, Henry CS, Hermjakob H, Herrgård MJ,
630 Kaafarani A, Kim HU, King Z, Klamt S, Klipp E, Koehorst JJ, König M,
631 Lakshmanan M, Lee DY, Lee SY, Lee S, Lewis NE, Liu F, Ma H, Machado D,
632 Mahadevan R, Maia P, Mardinoglu A, Medlock GL, Monk JM, Nielsen J, Nielsen
633 LK, Nogales J, Nookaew I, Palsson BO, Papin JA, Patil KR, Poolman M, Price

634 ND, Resendis-Antonio O, Richelle A, Rocha I, Sánchez BJ, Schaap PJ, Malik
635 Sheriff RS, Shoaie S, Sonnenschein N, Teusink B, Vilaça P, Vik JO, Wodke
636 JAH, Xavier JC, Yuan Q, Zakhartsev M, Zhang C. 2020. MEMOTE for
637 standardized genome-scale metabolic model testing. *Nat Biotechnol* 38:272–
638 276.
639 51. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-
640 throughput phenotypic testing and assay of gene function. *Genome Res*
641 11:1246–1255.
642 52. Jenior ML, Moutinho Jr. TJ, Dougherty B V, Papin JA. 2020. Transcriptome-
643 guided parsimonious flux analysis improves predictions with metabolic networks
644 in complex environments. *PLoS Comput Biol* 16:e1007099.
645 53. Kondorosi E, Mergaert P, Kereszt A. 2013. A paradigm for endosymbiotic life:
646 cell differentiation of rhizobium bacteria provoked by host plant factors. *Annu*
647 *Rev Microbiol* 67:611–628.
648 54. Vasse J, de Billy F, Camut S, Truchet G. 1990. Correlation between
649 ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa
650 nodules. *J Bacteriol* 172:4295–4306.
651 55. Karunakaran R, Ebert K, Harvey S, Leonard ME, Ramachandran V, Poole PS.
652 2006. Thiamine is synthesized by a salvage pathway in *Rhizobium*
653 *leguminosarum* bv. *viciae* strain 3841. *J Bacteriol* 188:6661–6668.
654 56. Miranda-Ríos J, Morera C, Taboada H, Dávalos A, Encarnación S, Mora J,
655 Soberón M. 1997. Expression of thiamin biosynthetic genes (*thiCOGE*) and
656 production of symbiotic terminal oxidase cbb3 in *Rhizobium etli*. *J Bacteriol*
657 179:6887–6893.

658 57. Newman JD, Diebold RJ, Schultz BW, Noel KD. 1994. Infection of soybean and
659 pea nodules by *Rhizobium* spp. purine auxotrophs in the presence of 5-
660 aminoimidazole-4-carboxamide riboside. *J Bacteriol* 176:3286–3294.
661 58. Soberón M, Morera C, Kondorosi A, Lopez O, Miranda J. 2001. A purine-related
662 metabolite negatively regulates *fixNOQP* expression in *Sinorhizobium meliloti*
663 by modulation of *fixK* expression. *Mol Plant Microbe Interact* 14:572–576.
664 59. Noel KD, Diebold RJ, Cava JR, Brink BA. 1988. Rhizobial purine and pyrimidine
665 auxotrophs: nutrient supplementation, genetic analysis, and the symbiotic
666 requirement for the novo purine biosynthesis. *Arch Microbiol* 149:499–506.
667 60. Pankhurst CE, Schwinghamer EA, Bergersen FJ. 1972. The structure and
668 acetylene-reducing activity of root nodules formed by a riboflavin-requiring
669 mutant of *Rhizobium trifolii*. *Microbiology* 70:161–177.
670 61. Taga ME, Walker GC. 2010. *Sinorhizobium meliloti* requires a cobalamin-
671 dependent ribonucleotide reductase for symbiosis with its plant host. *Mol Plant*
672 *Microbe Interact* 23:1643–1654.
673 62. Leong SA, Ditta GS, Helinski DR. 1982. Heme biosynthesis in *Rhizobium*.
674 Identification of a cloned gene coding for delta-aminolevulinic acid synthetase
675 from *Rhizobium meliloti*. *J Biol Chem* 257:8724–8730.
676 63. Mendoza-Suárez MA, Geddes BA, Sánchez-Cañizares C, Ramírez-González
677 RH, Kirchhelle C, Jorrin B, Poole PS. 2020. Optimizing *Rhizobium*–legume
678 symbioses by simultaneous measurement of rhizobial competitiveness and $N_2$
679 fixation in nodules. *Proc Natl Acad Sci U S A* 117:9822–9831.
680 64. Checcucci A, DiCenzo GC, Bazzicalupo M, Mengoni A. 2017. Trade, diplomacy,
681 and warfare: the quest for elite rhizobia inoculant strains. *Front Microbiol* 8:2207.

682 65. Kondorosi E, Kondorosi A. 2004. Endoreduplication and activation of the
683 anaphase-promoting complex during symbiotic cell development. *FEBS Lett*
684 567:152–157.
685 66. Kereszt A, Mergaert P, Kondorosi E. 2011. Bacteroid development in legume
686 nodules: evolution of mutual benefit or of sacrificial victims? *Mol Plant Microbe*
687 *Interact* 24:1300–1309.
688 67. Sen D, Weaver RW. 1984. A basis for different rates of $N_2$-fixation by the same
689 strains of *Rhizobium* in peanut and cowpea root nodules. *Plant Sci Lett* 34:239–
690 246.
691 68. Oono R, Schmitt I, Sprent JI, Denison RF. 2010. Multiple evolutionary origins of
692 legume traits leading to extreme rhizobial differentiation. *New Phytol* 187:508–
693 520.
694 69. Lodwig EM, Leonard M, Marroqui S, Wheeler TR, Findlay K, Downie JA, Poole
695 PS. 2005. Role of polyhydroxybutyrate and glycogen as carbon storage
696 compounds in pea and bean bacteroids. *Mol Plant-Microbe Interact* 18:67–74.
697 70. Terpolilli JJ, Masakapalli SK, Karunakaran R, Webb IUC, Green R, Watmough
698 NJ, Kruger NJ, Ratcliffe RG, Poole PS. 2016. Lipogenesis and redox balance in
699 nitrogen-fixing pea bacteroids. *J Bacteriol* 198:2864–2875.
700 71. Wang C, Saldanha M, Sheng X, Shelswell KJ, Walsh KT, Sobral BWS, Charles
701 TC. 2007. Roles of poly-3-hydroxybutyrate (PHB) and glycogen in symbiosis of
702 *Sinorhizobium meliloti* with *Medicago* sp. *Microbiology* 153:388–398.
703 72. Schumpp O, Crèvecoeur M, Broughton WJ, Deakin WJ. 2009. Delayed
704 maturation of nodules reduces symbiotic effectiveness of the *Lotus japonicus*–
705 *Rhizobium* sp. NGR234 interaction. *J Exp Bot* 60:581–590.

706 73. Calle ML, Urrea V, Boulesteix A-L, Malats N. 2011. AUC-RF: A new strategy for
707 genomic profiling with random forest. *Hum Hered* 72:121–132.
708 74. Zhang Y, Aono T, Poole P, Finan TM. 2012. NAD(P)$^+$-malic enzyme mutants of
709 *Sinorhizobium* sp. strain NGR234, but not *Azorhizobium caulinodans* ORS571,
710 maintain symbiotic $N_2$ fixation capabilities. *Appl Environ Microbiol* 78:2803–
711 2812.
712 75. Preston GG, Zeiher C, Wall JD, Emerich DW. 1989. Acetate-activating enzymes
713 of *Bradyrhizobium japonicum* bacteroids. *Appl Environ Microbiol* 55:165–170.
714 76. Summers ML, Denton MC, McDermott TR. 1999. Genes coding for
715 phosphotransacetylase and acetate kinase in *Sinorhizobium meliloti* are in an
716 operon that is inducible by phosphate stress and controlled by PhoB. *J Bacteriol*
717 181:2217–2224.
718 77. Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M. 2014.
719 Reconciliation of metabolites and biochemical reactions for metabolic networks.
720 *Brief Bioinform* 15:123–135.
721 78. Wheatley RM, Ramachandran VK, Geddes BA, Perry BJ, Yost CK, Poole PS.
722 2017. Role of $O_2$ in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on
723 glucose and succinate. *J Bacteriol* 199:e00572–16.
724 79. Galardini M, Mengoni A, Biondi EG, Semeraro R, Florio A, Bazzicalupo M,
725 Benedetti A, Mocali S. 2014. DuctApe: A suite for the analysis and correlation
726 of genomic and OmniLog™ Phenotype Microarray data. *Genomics* 103:1–10.
727 80. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A,
728 Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY,
729 Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT,

730 Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel L V., Apaolaza I,
731 Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM,
732 Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP,
733 Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ,
734 Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD,
735 Lewis NE, Sauter T, Palsson BØ, Thiele I, Fleming RMT. 2019. Creation and
736 analysis of biochemical constraint-based models using the COBRA Toolbox
737 v.3.0. *Nat Protoc* 14:639–702.
738 81. R Core Team. 2017. R: A language and environment for statistical computing.
739 R Foundation for Statistical Computing, Vienna, Austria.
740 82. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, Mundy M,
741 Chia N, Noor E, Beber ME, Best AA, DeJongh M, Kimbrel JA, D'haeseleer P,
742 McCorkle SR, Bolton JR, Pearson E, Canon S, Wood-Charlson EM, Cottingham
743 RW, Arkin AP, Henry CS. 2021. The ModelSEED biochemistry database for the
744 integration of metabolic annotations and the reconstruction, comparison and
745 analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res*
746 49:D575–D588.
747 83. Fagorzi C, Ilie A, Decorosi F, Cangioli L, Viti C, Mengoni A, diCenzo GC. 2020.
748 Symbiotic and nonsymbiotic members of the genus *Ensifer* (syn. *Sinorhizobium*)
749 are separated into two clades based on comparative genomics and high-
750 throughput phenotyping. *Genome Biol Evol* 12:2521–2534.

**Table 1. Rhizobial strains investigated in this study.**

| Rhizobial strain | Host plant | Nodule type | Terminal differentiation |
|---|---|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 | *Pisum sativum* (pea) | indeterminate | Yes |
| *Rhizobium leguminosarum* bv. *viciae* A34 | *Pisum sativum* (pea) | indeterminate | Yes |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 | *Phaseolus vulgaris* (bean) | determinate | No |
| *Rhizobium etli* CFN 42 | *Phaseolus vulgaris* (bean) | determinate | No |
| *Sinorhizobium meliloti* 1021 | *Medicago sativa* (alfalfa) | indeterminate | Yes |
| *Sinorhizobium fredii* NGR234 | *Vigna unguiculata* (cowpea) | determinate | No |
| | *Leucaena leucocephala* (white leadtree) | indeterminate | No |

**Table 2. Properties of genome-scale models.**

| Rhizobial strain | Genes | Metabolites | Reactions |
|---|---|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 | 1101 | 971 | 1254 |
| *Rhizobium leguminosarum* bv. *viciae* A34 | 981 | 984 | 1282 |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 | 979 | 984 | 1282 |
| *Rhizobium etli* CFN 42 | 976 | 952 | 1230 |
| *Sinorhizobium meliloti* 1021 | 1227 | 1163 | 1453 |
| *Sinorhizobium fredii* NGR234 | 992 | 1132 | 1406 |

**Table 3. Properties of bacteroid models.**

| | Genes | Metabolites | Reactions |
|---|---|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 | 377 | 348 | 366 |
| *Rhizobium leguminosarum* bv. *viciae* A34 | 475 | 528 | 540 |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 | 562 | 623 | 664 |
| *Sinorhizobium meliloti* 1021 | 311 | 252 | 258 |
| *Sinorhizobium fredii* NGR234 (*Leucaena leucocephala*) | 580 | 744 | 760 |
| *Sinorhizobium fredii* NGR234 (*Vigna unguiculata*) | 641 | 791 | 819 |

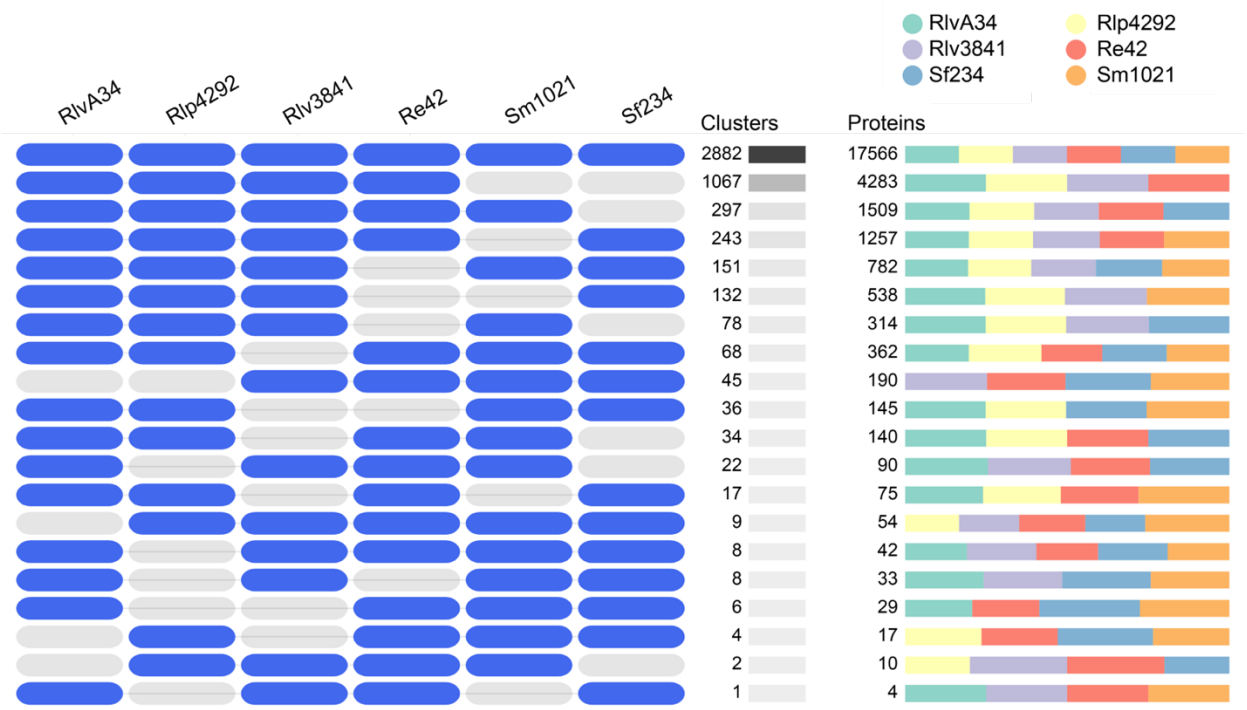



**Fig. 1. Analysis of orthologous gene clusters in the investigated rhizobial species.** The table on the left indicates presence/absence of shared gene clusters for each strain. Only gene clusters that occur in at least four of the six rhizobial species are shown. The number of shared clusters is indicated in the middle column. The table on the right indicates for each species the number of proteins that are contained in the shared clusters. The bars are scaled to represent for each strain the percentage of proteins in the shared clusters rather than the absolute number of proteins. This figure was created with OrthoVenn2 (48).

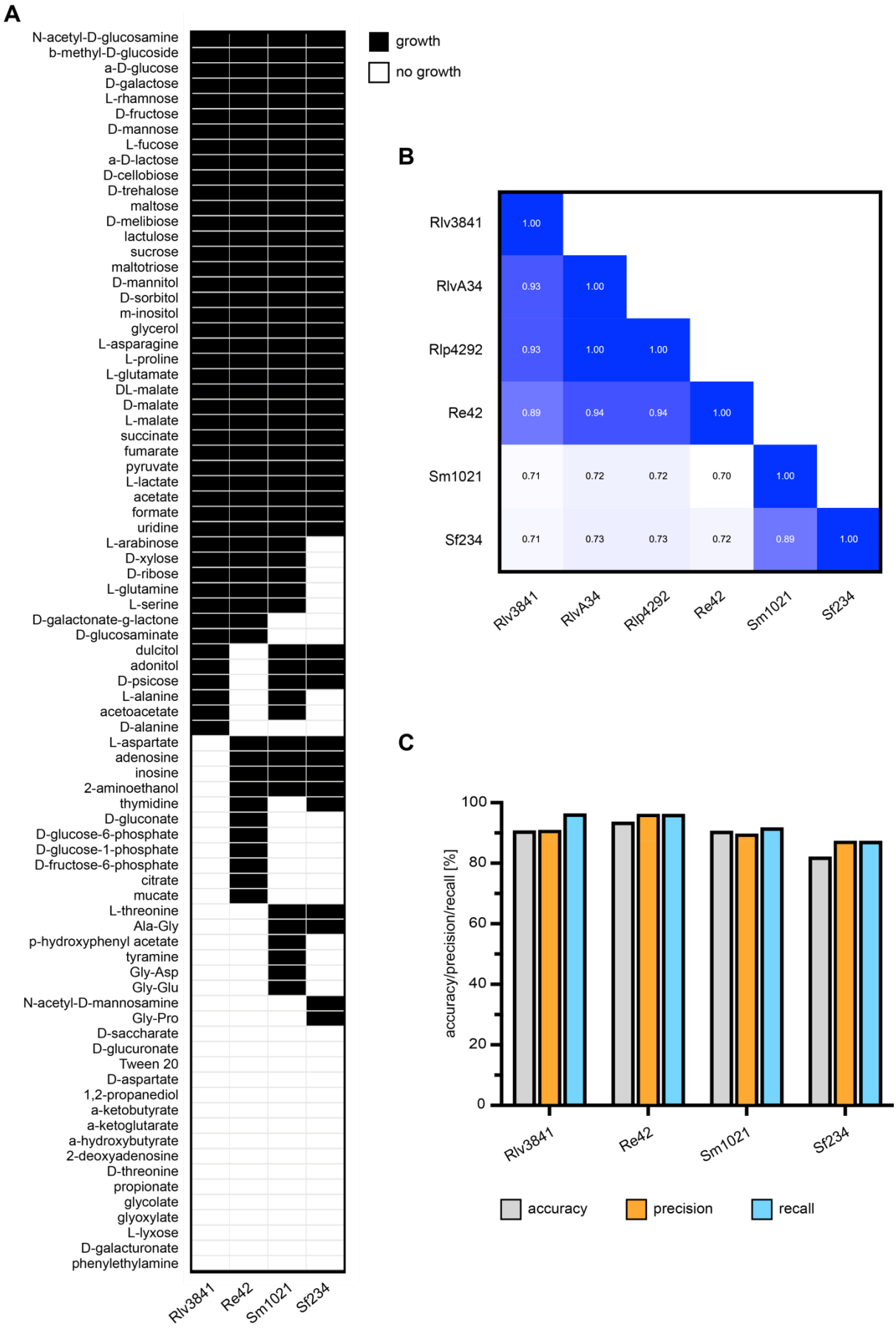

A
growth
no growth
N-acetyl-D-glucosamine
b-methyl-D-glucoside
a-D-glucose
D-galactose
L-rhamnose
D-fructose
D-mannose
L-fucose
a-D-lactose
D-cellobiose
D-trehalose
maltose
D-melibiose
lactulose
sucrose
maltotriose
D-mannitol
D-sorbitol
m-inositol
glycerol
L-asparagine
L-proline
L-glutamate
DL-malate
D-malate
L-malate
succinate
fumarate
pyruvate
L-lactate
acetate
formate
uridine
L-arabinose
D-xylose
D-ribose
L-glutamine
L-serine
D-galactonate-g-lactone
D-glucosaminate
dulcitol
adonitol
D-psicose
L-alanine
acetoacetate
D-alanine
L-aspartate
adenosine
inosine
2-aminoethanol
thymidine
D-gluconate
D-glucose-6-phosphate
D-glucose-1-phosphate
D-fructose-6-phosphate
citrate
mucate
L-threonine
Ala-Gly
p-hydroxyphenyl acetate
tyramine
Gly-Asp
Gly-Glu
N-acetyl-D-mannosamine
Gly-Pro
D-saccharate
D-glucuronate
Tween 20
D-aspartate
1,2-propanediol
a-ketobutyrate
a-ketoglutarate
a-hydroxybutyrate
2-deoxyadenosine
D-threonine
propionate
glycolate
glyoxylate
L-lyxose
D-galacturonate
phenylethylamine
Rlv3841
Re42
Sm1021
Sf234
B
Rlv3841
RlvA34
Rlp4292
Re42
Sm1021
Sf234
1.00
0.93 1.00
0.93 1.00 1.00
0.89 0.94 0.94 1.00
0.71 0.72 0.72 0.70 1.00
0.71 0.73 0.73 0.72 0.89 1.00
C
accuracy/precision/recall [%]
0
20
40
60
80
100
Rlv3841
Re42
Sm1021
Sf234
accuracy
precision
recall

**Fig. 2. Metabolic similarity of the rhizobial strains. A** Comparison of carbon source utilization determined by phenotype microarray analysis. Data for Rlv3841, Sm1021 and Sf234 were previously published (25, 43, 83). Only carbon sources where data were available for all four strains are shown. **B** Matrix showing the Jaccard similarity index for all model pairs based on model reaction content. **C** Model predictive performance for carbon source utilization. All available carbon source utilization data were used to determine accuracy, precision and recall for the model predictions.

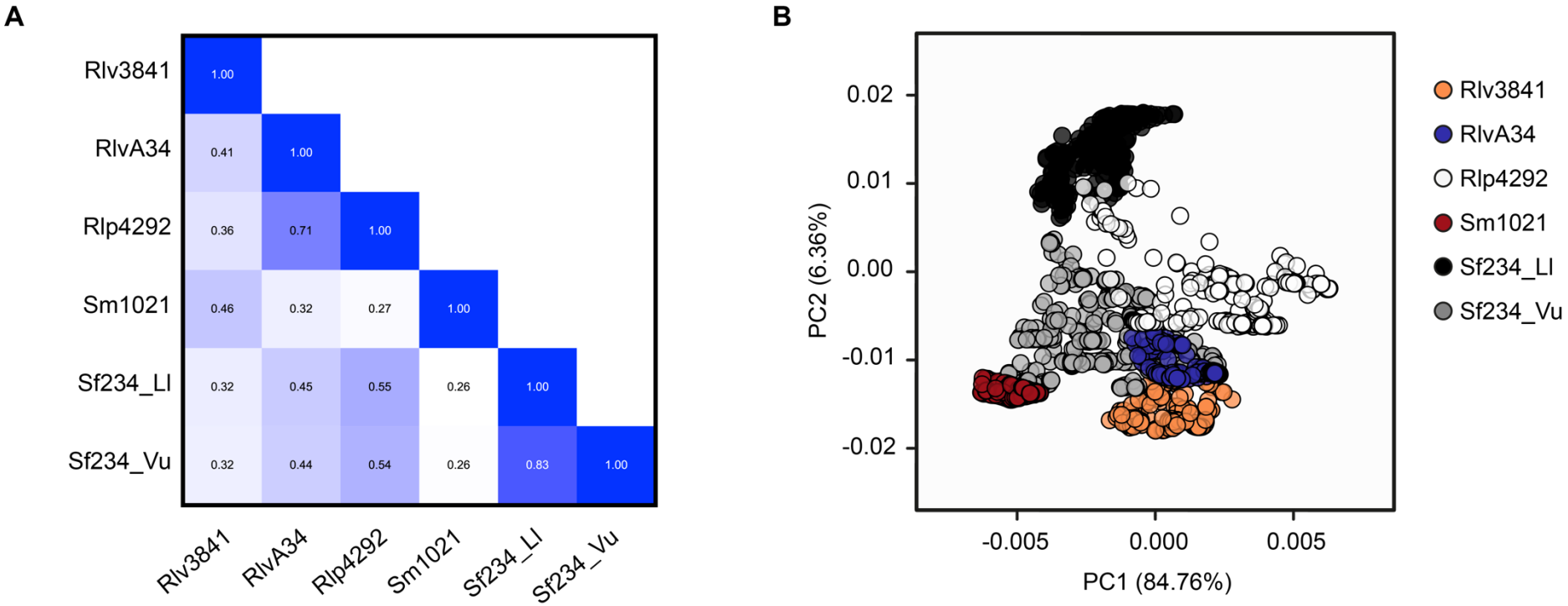



**Fig. 3. Reaction content of bacteroid models. A** Matrix showing the Jaccard similarity index for all pairs of bacteroid models based on model reactions. **B** Principal coordinate analysis of the Bray-Curtis dissimilarity of flux samples for those reactions that were shared between all bacteroid models.

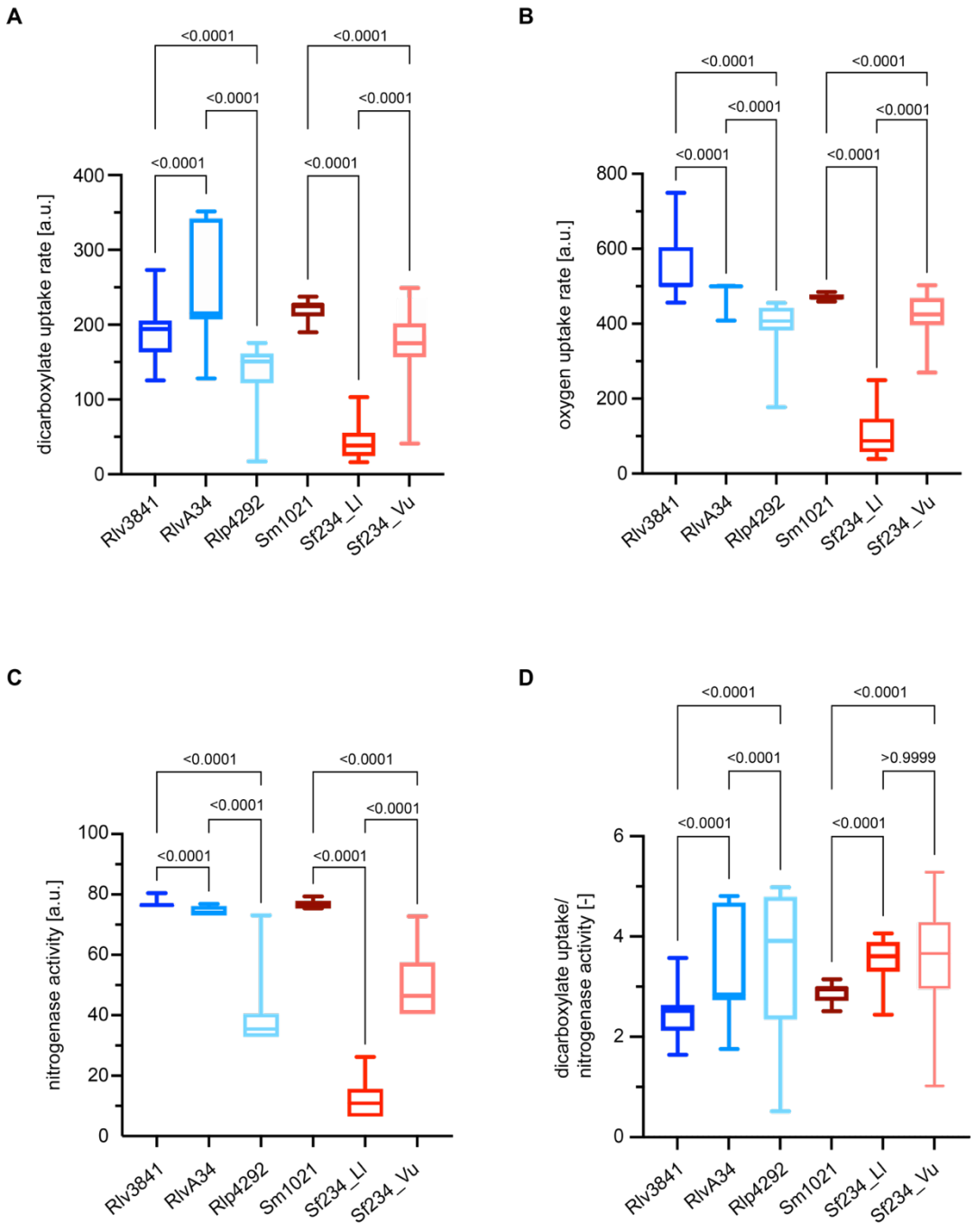



**Fig. 4. Comparison of symbiotic efficiency.** Box-whisker plots showing predicted values for **A** dicarboxylate uptake rate, **B** oxygen uptake rate, **C** nitrogenase activity, and **D** symbiotic efficiency as determined by dicarboxylate uptake required per nitrogenase activity. Values are based on n=500 flux samples for each strain. *P* values were calculated with Kruskal-Wallis test followed by Dunn's multiple comparisons test.

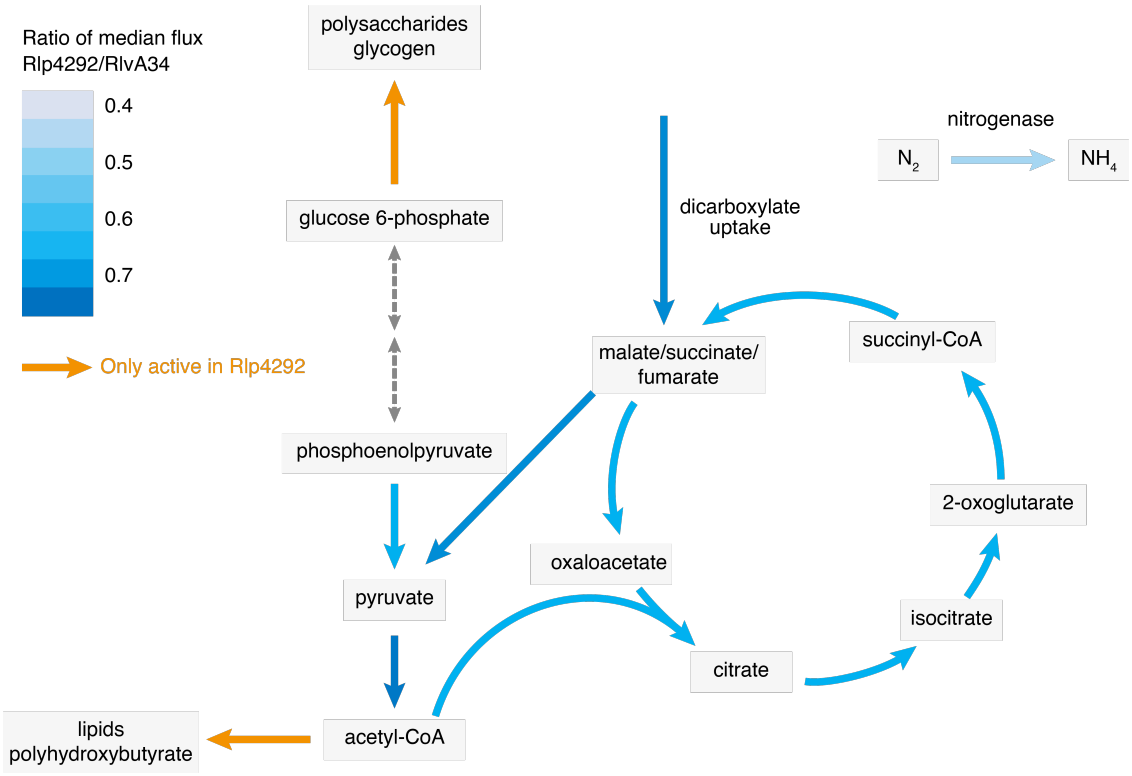



**Fig. 5. Metabolic flux distributions in reversibly differentiated and terminally differentiated bacteroids.** Flux map showing the ratio of median flux values for the Rlp4292 bacteroid model relative to the RlvA34 bacteroid model. Note that arrow thickness does not correspond to the magnitude of flux. The ratio of flux values is not indicated for reactions involved in glycolysis/gluconeogenesis because of the predicted sugar uptake by RlvA34, which is probably a result of sample preparation for RNA extraction as described in the main text.

Manuscript in preparation

# Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses

Carolin C. M. Schulte, Antonis Papachristodoulou, Philip S. Poole

## Supplementary Information

**Supplementary Data 1.** MEMOTE reports for the metabolic models developed in this study.

**Supplementary Data 2.** Mean decrease accuracies for reactions shared between bacteroid models determined by AUC-Random Forest.

**Supplementary Data 3.** Phenotype microarray data for *Rhizobium etli* CFN 42.

**Table S1. Sources of genome sequences for the strains investigated in this study.**

| Strain | Genome sequence |
|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 | NCBI RefSeq assembly GCF_000009265.1 |
| *Rhizobium leguminosarum* bv. *viciae* A34 | (15) |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 | NCBI RefSeq assembly GCF_000379005.1 |
| *Rhizobium etli* CFN 42 | NCBI RefSeq assembly GCF_000092045.1 |
| *Sinorhizobium meliloti* 1021 | NCBI RefSeq assembly GCF_000006965.1 |
| *Sinorhizobium fredii* NGR234 | NCBI RefSeq assembly GCF_000018545.1 |

**Table S2. Metabolites available to bacteroid models.** The table lists all metabolites for which unlimited uptake was allowed in all bacteroid models.

| Metabolite | SEED ID |
|---|---|
| L-malate | cpd00130 |
| succinate | cpd00036 |
| $O_2$ | cpd00007 |
| $N_2$ | cpd00528 |
| $H^+$ | cpd00067 |
| thiamine | cpd00305 |
| homocitrate | cpd00919 |
| $PO_4^{2-}$ | cpd00009 |
| $SO_4^{2-}$ | cpd00048 |
| $Cl^-$ | cpd00099 |
| $Mn^{2+}$ | cpd00030 |
| $Cu^{2+}$ | cpd00058 |
| $Ca^{2+}$ | cpd00063 |
| $Mg^{2+}$ | cpd00254 |
| $Zn^{2+}$ | cpd00034 |
| $Co^{2+}$ | cpd00149 |
| $K^+$ | cpd00205 |
| $Na^+$ | cpd00971 |
| $Fe^{2+}$ | cpd10515 |
| $Fe^{3+}$ | cpd10516 |
| $MoO_4^{2-}$ | cpd11574 |

**Table S3. RNA-Seq datasets used for bacteroid models.**

| Strain | RNA-Seq data |
|---|---|
| *Rhizobium leguminosarum* bv. *viciae* 3841 | NCBI BioProject PRJNA667846 |
| *Rhizobium leguminosarum* bv. *viciae* A34 | NCBI BioProject PRJEB28599 |
| *Rhizobium leguminosarum* bv. *phaseoli* 4292 | NCBI BioProject PRJEB28599 |
| *Sinorhizobium meliloti* 1021 | NCBI BioProject PRJNA401532 |
| *Sinorhizobium fredii* NGR234 | (13) |

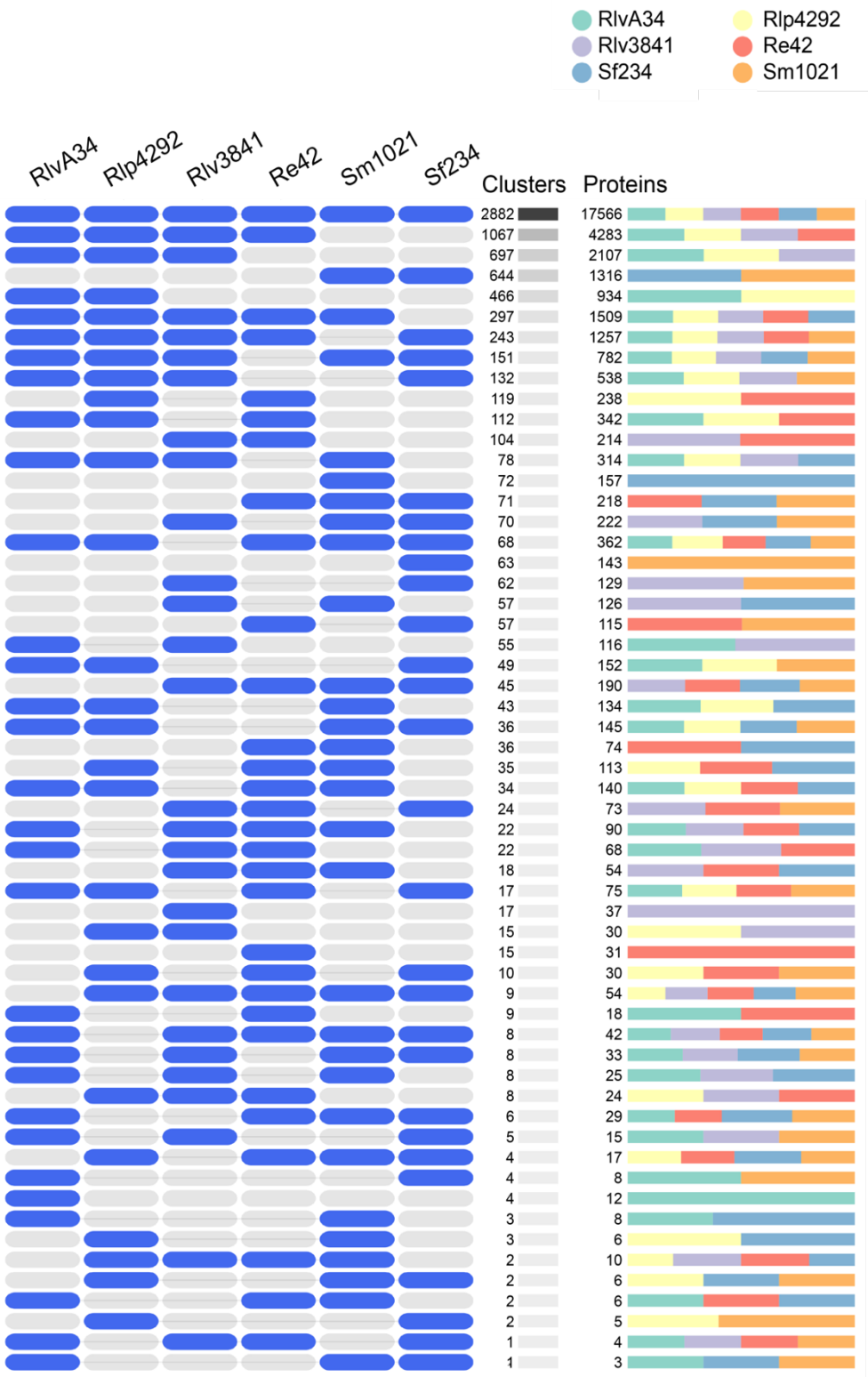



**Fig. S1. Analysis of orthologous gene clusters in the investigated rhizobial species.** The table on the left indicates presence/absence of shared gene clusters for each strain. The number of shared clusters is indicated in the middle column. The table

on the right indicates for each species the number of proteins that are contained in the shared clusters. The bars are scaled to represent for each species the percentage of proteins in the shared clusters rather than the absolute number of proteins. This figure was created with OrthoVenn2 (48).

## Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

| | |
|---|---|
| Title of Paper | Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses |
| Publication Status | Unpublished and unsubmitted work written in a manuscript style |
| Publication Details | C. C. M. Schulte, A. Papachristodoulou, P. S. Poole, Comparative metabolic modelling of bacteroid metabolism in rhizobium-legume symbioses, *in preparation*. |

### Student Confirmation

| | | | |
|---|---|---|---|
| Student Name: | Carolin C. M. Schulte | | |
| Contribution to the Paper | performed model reconstruction and analysis, performed phenotype microarray experiments, wrote manuscript | | |
| Signature | Carolin Schulte | Date | 9th August 2021 |

### Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

| | | | |
|---|---|---|---|
| Supervisor name and title: Professor Philip Poole | | | |
| Supervisor comments<br>Another state-of-the-art study that pushes the boundaries of what we know about the differences in nodule types | | | |
| Signature | Philip Poole | Date | 10th August 2021 |

This completed form should be included in the thesis, at the end of the relevant chapter.

# 6. Discussion and outlook

In summary, this thesis has demonstrated the use of metabolic modelling combined with heterogeneous experimental datasets to elucidate the metabolism of rhizobia in their complex symbiotic relationships with legumes. The results presented in this work particularly advance the current understanding of bacteroid metabolism, which is inherently difficult to investigate using only experimental approaches. The main finding is the control of bacteroid metabolism by plant supply of carbon and oxygen. Both the type of carbon source and the amount of carbon and oxygen provided to bacteroids were shown to play a key role in facilitating symbiotic nitrogen fixation and promoting the secretion of fixed nitrogen to the host plant. In addition, the curated metabolic models for various rhizobial strains that were developed as part of this work provide a framework for the interpretation of experimental results and are anticipated to be a valuable resource for guiding metabolic engineering efforts to improve the efficiency of rhizobium-legume interactions.

In Chapter 3, a highly curated genome-scale model for Rlv3841 was reconstructed and integrated with experimental data to reveal metabolic changes of Rlv3841 as it transitions from the soil into the nodule environment and finally differentiates into bacteroids fixing nitrogen in close interdependence with the plant host. Studies of rhizosphere metabolism have so far often investigated individual catabolic pathways[48–50] and while a whole-genome transcriptome study for Rlv3841 in the rhizosphere of different plants has been performed[35], those results had not been integrated with a comprehensive metabolic network. Our model predicted catabolism of pentose sugars, oligosaccharides, glycolate and aspartate and highlighted the importance of polyphosphate, leucine and AICAR synthesis in the rhizosphere.

Undifferentiated nodule bacteria were predicted to catabolize fructose in addition to dicarboxylates, which would agree with the observation that *Rhizobium trifolii* mutants that cannot transport $C_4$ dicarboxylates undergo at least partial differentiation[51]. The model further predicted that *myo*-inositol catabolism is specifically important for undifferentiated nodule bacteria. It is known that *myo*-inositol catabolism plays a role in competitive nodulation[50] and that *myo*-inositol is abundant in nodules[52]. For these reasons, previous modelling studies included uptake of *myo*-inositol by bacteroids[53,54], although activity of the catabolic enzymes has not been detected in bacteroids[52]. *Myo*-inositol catabolism by undifferentiated nodule bacteria could explain both the competitive disadvantage of mutant strains and the presence of *myo*-inositol in nodules, and also agrees with the suggested catabolism of inositol-derived rhizopines by nodule bacteria as a kin selection strategy[55]. Further experimental studies using reporter fusions and enzyme assays to track expression and activity of the catabolic enzymes will be required to confirm this prediction.

While it is known that dicarboxylates are the main carbon source fuelling bacteroid metabolism, model predictions further suggested uptake of xylose as well as glycolate. Validating the suggested uptake of a sugar feeding into the pentose phosphate pathway will require isolation of mutant strains in various glycolytic/gluconeogenic enzymes and investigation of their symbiotic phenotypes. Should the model predictions be confirmed experimentally, it would be interesting to use the model to investigate possible benefits of this phenomenon. Uptake of glycolate, which was hypothesised based on metabolome data and discrepancies of model predictions with experimental evidence, could explain induction of an incomplete glyoxylate cycle in bacteroids. While malate synthase is highly upregulated[32], isocitrate lyase activity

seems to be absent from bacteroids[56], and the source of glyoxylate for the malate synthase reaction has therefore not been known so far. Another interesting finding was the predicted synthesis of histidine by bacteroids in contrast to the general downregulation of amino acid synthesis, which is due to their growth-arrested state as well as amino acid provision by the plant host[5,57,58]. Transcriptome and proteome data indicate upregulation of histidine synthesis for various rhizobial strains and host plants[22,59]. While this can partly be explained by the enrichment of Nif and Fix proteins with histidine, transport assays with bacteroids should be performed to determine if histidine uptake rates could be limiting. Mutant strains that are auxotrophic for histidine have so far shown strain- and host-specific phenotypes ranging from an inability to induce nodule formation to wild-type levels of fixation[60,61]. The stage at which mutants were impaired in symbiosis formation has not always been determined, warranting further studies with detailed phenotypic characterisation. For Rlv3841, most genes involved in histidine biosynthesis were essential for growth in complete media in an insertion sequencing experiment[36]. However, a second copy of *hisF* (RL0819) was specifically essential for nodule bacteria and bacteroids, and a second copy of *hisH* (RL0820) was required for root colonization and also essential in nodule bacteria and bacteroids. The HisFH complex catalyses formation of both imidazole-glycerol-3-phosphate for histidine synthesis and AICAR. Its essentiality may thus be related to the importance of AICAR synthesis during the infection process as described in Chapter 3. A second copy of *hisD* (pRL80041), which catalyses the final reaction in the histidine biosynthetic pathway, was specifically essential for nodule bacteria, which could indicate a requirement for increased histidine biosynthesis during growth in infection threads.

Chapter 4 contained an in-depth discussion of central features of bacteroid metabolism and how metabolic flux patterns are controlled by the plant host. While it has previously been shown that rhizobia catabolize dicarboxylates[51,62,63], secrete amino acids[64] and synthesize carbon polymers[31,65,66], this work demonstrated the interplay between these observations for the first time. A reduced model of bacteroid metabolism was reconstructed based on experimental datasets. Elementary conversion mode enumeration was then used to demonstrate that carbon polymer synthesis and amino acid secretion by bacteroids decrease oxygen demand per metabolized carbon. However, storage polymer synthesis also increased the oxygen demand per fixed nitrogen and the carbon cost of nitrogen supply to the plant, indicating that plant control of carbon and oxygen supply in response to symbiotic performance may limit diversion of carbon into storage polymers. The use of elementary conversion modes facilitated the investigation of general network properties without constraining the uptake or synthesis of any compound and circumvented the need to define an objective function, which has been done in most metabolic modelling studies of bacteroids so far[22,44,45,54]. Elementary conversion modes in combination with a robustness analysis for the effect of oxygen uptake on metabolic flux patterns indicated that flux through the decarboxylating tricarboxylic acid (TCA) cycle arm is restricted at low oxygen levels. The resulting limited capacity for synthesis of 2-oxoglutarate would restrict the assimilation of ammonia by bacteroids and thus force ammonia secretion to the plant instead. The predicted downregulation of the decarboxylating TCA cycle arm was further validated by metabolic flux analysis of *Azorhizobium caulinodans* bacteroids. Both modelling and experimental results showed that catabolism of dicarboxylates rather than the

abundant sucrose by bacteroids induces a higher oxygen demand and a higher NADH/$NAD^+$ ratio. According to the model, the high oxygen demand would contribute to limiting 2-oxoglutarate synthesis and thus forcing ammonia secretion, however it may also play a role in protecting the oxygen-sensitive nitrogenase enzyme[67]. High availability of reduced electron carriers such as NADH is important for electron supply to nitrogenase[31]. Model predictions and experimental data thus elucidated the fundamental role that plant-controlled carbon and oxygen availability play in enabling carbon and redox balance during nitrogen fixation under microoxic conditions while forcing ammonia secretion to enable a productive symbiotic relationship with the plant host. The concept of plant control over the metabolism of its symbiont as demonstrated in this work is consistent with experimental evidence of sanctioning, where plants reduce the supply of carbon and oxygen to bacteroids that fix little or no nitrogen[68,69]. Finally, metabolic models for different *Rhizobium* and *Sinorhizobium* strains were comparatively analysed in Chapter 5. While newly generated experimental data will inevitably change the content of metabolic reconstructions over time, additional artefactual differences between models are often introduced during the reconstruction process, for example due to the choice of reaction databases and reconstruction tools, the intended application of the reconstruction and the extent of manual curation that is performed. Meaningful inter-species comparisons therefore require careful model reconciliation to identify true differences in metabolic capabilities[70,71]. Models for rhizobial strains that have been reconstructed so far often focus on bacteroids and are therefore limited in scope. In addition, FBA with an objective function forcing the secretion of alanine in addition to ammonia as well as storage polymer production has generally been used[22,44,53]. As demonstrated in Chapter 4, amino acid secretion and

carbon polymer synthesis are highly dependent on carbon and oxygen uptake rates, which are difficult to determine experimentally. Based on the model for Rlv3841 presented in Chapter 3 and a published model for *S. meliloti* 1021[9], genome-scale metabolic models for six rhizobial strains were therefore reconstructed and reconciled. Similar to the studies presented in Chapters 3 and 4, nutrient uptake rates for bacteroid models were left largely unconstrained and gene expression data were used instead to determine nutrient uptake and metabolic flux distributions in the different host-symbiont pairings. Comparing models for reversibly and terminally differentiated bacteroids of closely related strains indicated a higher physiological complexity of reversibly differentiated bacteroids, which is in agreement with their similarity to free-living rhizobia and in contrast to the major physiological changes observed in terminally differentiated bacteroids[72]. Consistent with experimental measurements of symbiotic efficiency, the bacteroid models for reversibly differentiated bacteroids predicted lower metabolic activity and symbiotic efficiency compared to terminally differentiated bacteroids[73]. A better understanding of the mechanisms inducing terminal differentiation may therefore enable the engineering of novel symbioses with improved efficiency[74]. The models were further able to identify malic enzyme as the main pathway for pyruvate synthesis in *S. meliloti*[75], whereas activity of both malic enzyme and phosphoenolpyruvate carboxykinase/pyruvate kinase was predicted for Rlv3841[76], and host-dependent differences in pyruvate synthesis were correctly predicted for *S. fredii*[77]. While the good agreement of model predictions with experimental evidence indicates successful creation of symbiosis-specific models, comparison of flux samples for pathways outside of central carbon metabolism was difficult because most differences identified through machine learning were due to

variations in predicted metabolic activity. For example, lower nitrogenase activity directly resulted in a lower demand for cofactors such as iron-sulfur clusters. Identification of additional distinguishing features of the investigated symbioses may thus require specification of further information on strain- and host plant-specific metabolic properties, which could be obtained from the literature or inferred from omics datasets, similar to the concept of metabolic tasks that has been applied to models of human metabolism[21,78]. Alternatively, an approach similar to the elementary conversion mode analysis described in Chapter 4 could be applied to investigate properties of the metabolic networks irrespective of predicted metabolic activity.

The models of bacteroid metabolism described in this work provide a framework for the application of metabolic engineering strategies to identify metabolic bottlenecks[24,79,80] and subsequently optimise the efficiency of nitrogen fixation in rhizobium-legume symbioses through targeted engineering of rhizobia and/or target crops. However, it is important to note that unlike traditional metabolic engineering projects, the interaction with the plant host makes rational engineering of bacteroid metabolism more challenging compared to controlled cultivations in bioreactors. To this end, integration of plant and bacterial models will be required to advance the investigation of metabolite exchanges, as recently demonstrated for the symbiosis between *S. meliloti* and *Medicago truncatula*[9]. Similarly, implementation of dynamic modelling frameworks could be beneficial to elucidate the response of rhizobia to changes in nitrogen demand, which varies diurnally[81] and is impacted by changing environmental conditions, such as rising temperatures and $CO_2$ levels caused by global warming[82,83]. Dynamic modelling approaches have been used to investigate temporal resource allocation in cyanobacteria[84,85], and similar approaches could be

adapted for bacteroid metabolism. Building on the curated metabolic models presented in this work, additional information can be readily incorporated to account for macromolecular expression, enzyme kinetics, thermodynamics or proteome constraints[86–89]. In combination with model-driven improvements to photosynthetic efficiency[90], the engineering of highly efficient plant-microbe symbioses enabled by metabolic modelling will be essential for addressing the global challenge of increasing food demand in a sustainable way.

# 7. Bibliography


1. Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).

2. Choudhury, A. T. M. A. & Kennedy, I. R. Nitrogen fertilizer losses from rice soils and control of environmental pollution problems. *Commun. Soil Sci. Plant Anal.* **36**, 1625–1639 (2005).

3. Fowler, D. *et al.* The global nitrogen cycle in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130164 (2013).

4. Poole, P., Ramachandran, V. & Terpolilli, J. Rhizobia: from saprophytes to endosymbionts. *Nat. Rev. Microbiol.* **16**, 291–303 (2018).

5. Ledermann, R., Schulte, C. C. M. & Poole, P. S. How rhizobia adapt to the nodule environment. *J. Bacteriol.* **203**, e00539–20 (2021).

6. Prell, J. & Poole, P. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol.* **14**, 161–168 (2006).

7. Jansma, J. & El Aidy, S. Understanding the host-microbe interactions using metabolic modeling. *Microbiome* **9**, 16 (2021).

8. Heinken, A., Sahoo, S., Fleming, R. M. T. & Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **4**, 28–40 (2013).

9. DiCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A. & Fondi, M. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* **11**, 1–11 (2020).

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

11. Nielsen, J. Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).

12. Cuevas, D. A. *et al.* From DNA to FBA: how to build your own genome-scale metabolic model. *Front. Microbiol.* **7**, 907 (2016).

13. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* **28**, 245–248 (2010).

14. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci.* **97**, 5528–5533 (2000).

15. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.* **46**, 931–936 (2018).

16. Wang, H. *et al.* RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Comput. Biol.* **14**, e1006541 (2018).

17. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).

18. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).

19. Lu, H. *et al.* A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).

20. Mora Salguero, D. A. *et al.* Development of a *Chlamydomonas reinhardtii* metabolic network dynamic model to describe distinct phenotypes occurring at different $CO_2$ levels. *PeerJ* **6**, e5528 (2018).

21. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425 (2013).

22. Resendis-Antonio, O. *et al.* Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. *BMC Syst. Biol.* **5**, 120 (2011).

23. Heinken, A., Basile, A. & Thiele, I. Advances in constraint-based modelling of microbial communities. *Curr. Opin. Syst. Biol.* https://doi.org/10.1016/j.coisb.2021.05.007 (2021).

24. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).

25. Denison, R. F. A Darwinian perspective on improving nitrogen-fixation efficiency of legume crops and forages. in *Crop Physiology* (Second Edition) (eds. Sadras, V. O. & Calderini, D. F.) 207–222 (Academic Press, 2015).

26. Pankievicz, V. C. S., Irving, T. B., Maia, L. G. S. & Ané, J. M. Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biology* **17**, 1–17 (2019).

27. Mus, F. *et al.* Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).

28. Ryu, M. H. *et al.* Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* **5**, 314–330 (2020).

29. Mendoza-Suárez, M. A. *et al.* Optimizing *Rhizobium*-legume symbioses by simultaneous measurement of rhizobial competitiveness and $N_2$ fixation in nodules. *Proc. Natl. Acad. Sci.* **117**, 9822–9831 (2020).

30. Checcucci, A., DiCenzo, G. C., Bazzicalupo, M. & Mengoni, A. Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. *Front. Microbiol.* **8**, 2207 (2017).

31. Terpolilli, J. J. *et al.* Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. *J. Bacteriol.* **198**, 2864–2875 (2016).

32. Karunakaran, R. *et al.* Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**, 4002–4014 (2009).

33. Wheatley, R. M. *et al.* Role of $O_2$ in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate. *J. Bacteriol.* **199**, e00572–16 (2017).

34. DiCenzo, G. C. *et al.* Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genet.* **14**, e1007357 (2018).

35. Ramachandran, V. K., East, A. K., Karunakaran, R., Downie, J. A. & Poole, P. S. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol.* **12**,

R106 (2011).

36. Wheatley, R. M. *et al.* Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc. Natl. Acad. Sci.* **117**, 23823–23834 (2020).

37. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**, 817–837 (2014).

38. Peng, J. *et al.* RNA-Seq and microarrays analyses reveal global differential transcriptomes of *Mesorhizobium huakuii* 7653R between bacteroids and free-living cells. *PLoS One* **9**, e93626 (2014).

39. Schulte, C. C. M. *et al.* Metabolic control of nitrogen fixation in rhizobium-legume symbioses. *Sci. Adv.* **7**, eabh2433 (2021).

40. Blazier, A. S. & Papin, J. A. Reconciling high-throughput gene essentiality data with metabolic network reconstructions. *PLOS Comput. Biol.* **15**, e1006507 (2019).

41. Blazier, A. S. & Papin, J. A. Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* **3**, 299 (2012).

42. Ramon, C., Gollub, M. G. & Stelling, J. Integrating *–omics* data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.* **62**, 563–574 (2018).

43. Kim, M. K. & Lun, D. S. Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput. Struct. Biotechnol. J.* **11**, 59–65 (2014).

44. Contador, C. A., Lo, S.-K., Chan, S. H. J. & Lam, H.-M. Metabolic analyses of

nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* **5**, e00516–19 (2020).

45. Yang, Y., Hu, X.-P. & Ma, B.-G. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states. *Mol. BioSyst* **13**, 607 (2017).

46. Pfau, T. *et al.* The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* **8**, 12504 (2018).

47. Schulte, C. C. M., Ramachandran, V. K., Papachristodoulou, A. & Poole, P. S. Genome-scale metabolic modelling of lifestyle changes in *Rhizobium leguminosarum*. Preprint at https://doi.org/10.1101/2021.07.28.454262 (2021).

48. Vanderlinde, E. M., Hynes, M. F. & Yost, C. K. Homoserine catabolism by *Rhizobium leguminosarum* bv. *viciae* 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environ. Microbiol.* **16**, 205–217 (2014).

49. Yost, C. K., Rath, A. M., Noel, T. C. & Hynes, M. F. Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **152**, 2061–2074 (2006).

50. Fry, J., Wood, M. & Poole, P. S. Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Mol. Plant-Microbe Interact.* **14**, 1016–1025 (2001).

51. Ronson, C. W., Lyttleton, P. & Robertson, J. G. $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc. Natl. Acad. Sci.* **78**, 4284–4288 (1981).

52. Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. *Myo*-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **140**, 2787–2795 (1994).

53. Zhao, H., Li, M., Fang, K., Chen, W. & Wang, J. *In silico* insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PLoS One* **7**, e31287 (2012).

54. DiCenzo, G. C. *et al.* Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **7**, 1–10 (2016).

55. Oono, R., Denison, R. F. & Kiers, E. T. Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol.* **183**, 967–979 (2009).

56. Johnson, G. V, Evans, H. J. & Ching, T. Enzymes of the glyoxylate cycle in rhizobia and nodules of legumes. *Plant Physiol.* **41**, 1330–1336 (1966).

57. Prell, J. *et al.* Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci.* **106**, 12477–12482 (2009).

58. Lodwig, E. M. *et al.* Amino-acid cycling drives nitrogen fixation in the legume–rhizobium symbiosis. *Nature* **422**, 722–726 (2003).

59. Green, R. T., East, A. K., Karunakaran, R., Downie, J. A. & Poole, P. S. Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. *Microb. Genomics* **5**, (2019).

60. Yadav, Vashishat, Kuykendall & Hashem. Biochemical and symbiotic properties

of histidine-requiring mutants of *Rhizobium leguminosarum* biovar *trifolii*. *Lett. Appl. Microbiol.* **26**, 22–26 (1998).

61. Dunn, M. F. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Crit. Rev. Microbiol.* **41**, 411–451 (2015).

62. Mitsch, M. J., DiCenzo, G. C., Cowie, A. & Finan, T. M. Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* **84**, e01561–17 (2018).

63. Finan, T. M., Wood, J. M. & Jordan, D. C. Symbiotic properties of $C_4$-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J. Bacteriol.* **154**, 1403–1413 (1983).

64. Allaway, D. *et al.* Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Molecular Microbiology* **36**, 508–515 (2000).

65. Lodwig, E. M. *et al.* Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Mol. Plant-Microbe Interact.* **18**, 67–74 (2005).

66. Marroquí, S. *et al.* Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. *J. Bacteriol.* **183**, 854–864 (2001).

67. Gallon, J. R. R. The oxygen sensitivity of nitrogenase: a problem for biochemists and microorganisms. *Trends Biochem. Sci.* **6**, 19–23 (1981).

68. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003).

69. Westhoek, A. *et al.* Conditional sanctioning in a legume–*Rhizobium* mutualism. *Proc. Natl. Acad. Sci.* **118**, e2025760118 (2021).

70. Oberhardt, M. A., Puchałka, J., Martins dos Santos, V. A. P. & Papin, J. A. Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Comput. Biol.* **7**, e1001116 (2011).

71. Blais, E. M. *et al.* Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat. Commun.* **8**, 14250 (2017).

72. Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proc. Natl. Acad. Sci.* **103**, 5230–5235 (2006).

73. Oono, R. & Denison, R. F. Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* **154**, 1541–1548 (2010).

74. Alunni, B. & Gourion, B. Terminal bacteroid differentiation in the legume-rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytol.* **211**, 411–417 (2016).

75. Driscoll, B. T. & Finan, T. M. $NAD^+$-dependent malic enzyme of *Rhizobium meliloti* is required for symbiotic nitrogen fixation. *Mol. Microbiol.* **7**, 865–873 (1993).

76. Mulley, G. *et al.* Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. *J. Bacteriol.* **192**, 4944–4953 (2010).

77. Zhang, Y., Aono, T., Poole, P. & Finan, T. M. $NAD(P)^+$-malic enzyme mutants of *Sinorhizobium* sp. strain NGR234, but not *Azorhizobium caulinodans* ORS571, maintain symbiotic $N_2$ fixation capabilities. *Appl. Environ. Microbiol.*

**78**, 2803–2812 (2012).

78. Richelle, A., Chiang, A. W. T., Kuo, C.-C. & Lewis, N. E. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLOS Comput. Biol.* **15**, e1006867 (2019).

79. Mol, V. *et al.* Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism. *Metab. Eng.* **65**, 123–134 (2021).

80. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).

81. Cárdenas-Navarro, R., Adamowicz, S. & Robin, P. Diurnal nitrate uptake in young tomato (*Lycopersicon esculentum* Mill.) plants: test of a feedback-based model. *J. Exp. Bot.* **49**, 721–730 (1998).

82. Noyce, G. L., Kirwan, M. L., Rich, R. L. & Megonigal, J. P. Asynchronous nitrogen supply and demand produce nonlinear plant allocation responses to warming and elevated $CO_2$. *Proc. Natl. Acad. Sci.* **116**, 21623–21628 (2019).

83. Vlassak, K. M., Vanderleyden, J. & Graham, P. H. Factors influencing nodule occupancy by inoculant rhizobia. *Crit. Rev. Plant Sci.* **16**, 163–229 (1997).

84. Rügen, M., Bockmayr, A. & Steuer, R. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Sci. Rep.* **5**, 15247 (2015).

85. Ofaim, S., Sulheim, S., Almaas, E., Sher, D. & Segrè, D. Dynamic allocation of carbon storage and nutrient-dependent exudation in a revised genome-scale

model of *Prochlorococcus*. *Front. Genet.* **12**, 586293 (2021).

86. Sánchez, B. J. *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **13**, 935 (2017).

87. Yang, X. *et al.* Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metab. Eng.* **67**, 133–144 (2021).

88. Salvy, P. & Hatzimanikatis, V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat. Commun.* **11**, 30 (2020).

89. Lerman, J. A. *et al. In silico* method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* **3**, 929 (2012).

90. Wang, Y., Chan, K. X. & Long, S. P. Toward a dynamic photosynthesis model to guide yield improvement in $C_4$ crops. *Plant J.* https://doi.org/10.1111/tpj.15365 (2021).