




# Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium

George C. diCenzo <sup>1,2</sup>, Michelangelo Tesi<sup>1</sup>, Thomas Pfau<sup>3</sup>, Alessio Mengoni <sup>1</sup>✉ & Marco Fondi <sup>1</sup>✉

The mutualistic association between leguminous plants and endosymbiotic rhizobial bacteria is a paradigmatic example of a symbiosis driven by metabolic exchanges. Here, we report the reconstruction and modelling of a genome-scale metabolic network of *Medicago truncatula* (plant) nodulated by *Sinorhizobium meliloti* (bacterium). The reconstructed nodule tissue contains five spatially distinct developmental zones and encompasses the metabolism of both the plant and the bacterium. Flux balance analysis (FBA) suggests that the metabolic costs associated with symbiotic nitrogen fixation are primarily related to supporting nitrogenase activity, and increasing N<sub>2</sub>-fixation efficiency is associated with diminishing returns in terms of plant growth. Our analyses support that differentiating bacteroids have access to sugars as major carbon sources, ammonium is the main nitrogen export product of N<sub>2</sub>-fixing bacteria, and N<sub>2</sub> fixation depends on proton transfer from the plant cytoplasm to the bacteria through acidification of the peribacteroid space. We expect that our model, called 'Virtual Nodule Environment' (ViNE), will contribute to a better understanding of the functioning of legume nodules, and may guide experimental studies and engineering of symbiotic nitrogen fixation.

<sup>1</sup>Department of Biology, University of Florence, Sesto Fiorentino, Italy. <sup>2</sup>Department of Biology, Queen's University, Kingston, ON, Canada. <sup>3</sup>Life Sciences Research Unit, University of Luxembourg, Belvaux, Luxembourg. ✉email: [alessio.mengoni@unifi.it](mailto:alessio.mengoni@unifi.it); [marco.fondi@unifi.it](mailto:marco.fondi@unifi.it)

**M**acroorganisms are colonized by a staggering diversity of microorganisms, collectively referred to as a ‘holobiont’<sup>1,2</sup>. The intimate association between organisms is often driven by metabolic exchanges: many insects obtain essential nutrients from obligate bacterial symbionts<sup>3</sup>, most plants can obtain phosphorus from arbuscular mycorrhiza in exchange for carbon<sup>4</sup>, and the gut microbiota is thought to contribute to animal nutrition<sup>5,6</sup>. Complex global patterns often emerge during these intimate biological associations<sup>7</sup>, especially when nutritional inter-dependencies are involved<sup>8–10</sup>. The communication between the two metabolic networks of the interacting organisms may give rise to unpredicted phenotypic traits and unexpected emergent properties. Metabolic relationships can span over a large taxonomic range and have profound biological relevance<sup>11–14</sup>. For example, the interactions between bacteria and multicellular organisms have been suggested to be key drivers of evolutionary transitions, leading to eukaryotic diversification and to the occupancy of novel niches<sup>9,15,16</sup>. The study of the association of two biological entities is mainly challenged by the size of the system and by the unpredictability of their metabolic interactions. Theoretical, systems-level models are required to unravel the intimate functioning of metabolic associations and to eventually exploit their potential in biotechnological applications.

Symbiotic nitrogen fixation (SNF) is a paradigmatic example of the importance and the complexity of natural biological associations. SNF is a mutualistic relationship between a group of plant families, including the Fabaceae, and a polyphyletic group of alpha- and beta-proteobacteria known as rhizobia, or a taxa of Actinobacteria (*Frankia* spp.), in which the plants provide a niche and carbon to the bacteria in exchange for fixed nitrogen<sup>17</sup>. SNF involves constant metabolic cross-talk between the plant and the bacteria<sup>18</sup>, and it is a paradigmatic example of bacterial cellular differentiation<sup>19</sup> and sociomicrobiological interactions<sup>20</sup>. The rhizobia intra-cellularly colonize plant cells of a specialized organ known as a root (or stem) nodule. The intra-cellular rhizobia (referred to as bacteroids) are surrounded by a plant-derived membrane, and the term symbiosome is used in reference to the structure consisting of the bacteroid, the plant-derived membrane (i.e., the peribacteroid membrane), and the intervening space (i.e., the peribacteroid space). Nodules with an indeterminate structure, such as those formed by the plant *Medicago truncatula*, are divided into spatially distinct developmental zones<sup>21</sup> with a distal apical meristem and a proximal nitrogen fixation zone.

SNF plays a key role in the global nitrogen cycle and is central to sustainable agricultural practices by reducing the usage of synthetic nitrogen fertilizers whose application results in a multitude of adverse environmental consequences<sup>22–24</sup>. Unfortunately, our ability to maximize the benefit of SNF is limited since rhizobial inoculants are often poorly effective due to low competitiveness<sup>25,26</sup> and because rhizobium symbioses are specific to leguminous plants. Manipulating the rhizobium – legume interaction for biotechnological purposes will require an in-depth understanding of the symbiotic interaction, as well as an ability to predict the consequences of genetic changes and environmental perturbations.

From a metabolic perspective, genome-scale metabolic reconstruction (GENREs) and constraint-based modeling has great potential to fulfill these roles. A GENRE also serves as a comprehensive knowledgebase of an organism’s metabolism, containing hundreds to thousands of metabolic and transport reactions, most of which are linked to the corresponding gene(s) whose gene product(s) catalyzes the reaction<sup>27,28</sup>. With the aid of mathematical approaches such as flux balance analysis (FBA), GENREs can be used to identify emergent system-level properties, to predict active reactions, and to identify essential genes<sup>29</sup>. Compared to simple enrichment analyses that are typical in omics studies, GENRE-based methods allow for the interpretation of data in a connected manner based on network topology

and to infer the effects of changes in remote pathways on the overall cell physiology. When considering interacting entities, for example, this approach can predict the consequence of mutations in one organism on the metabolism of the other. However, multi-organism metabolic reconstructions are still in their infancy, and very few examples of combined models exist compared to single strain GENREs<sup>8,14,30–32</sup>.

Despite the importance of metabolism to SNF<sup>18</sup>, there has been limited use of metabolic modeling in the study of rhizobia and SNF. To date, GENREs of varying quality have been reported for only three rhizobia: *Sinorhizobium meliloti*<sup>33–35</sup>, *Rhizobium etli*<sup>36–38</sup>, and *Bradyrhizobium diazoefficiens*<sup>39</sup>. Currently, *M. truncatula*<sup>40</sup> and *Glycine max*<sup>41</sup> are the only legumes with published GENREs. With the exception of the *G. max* GENRE, these GENREs have been used in preliminary analyses of SNF, providing results generally consistent with expectations. However, all analyses to date suffer from two major limitations. Simulations with the rhizobium models ignore plant metabolism, while simulations with the *M. truncatula* GENRE (based on the genome sequence published in 2011<sup>42</sup>, which has since been updated in 2014<sup>43</sup> and again 2018<sup>44</sup>) involved a very limited draft *S. meliloti* metabolic reconstruction (whose genome sequence was published in 2001<sup>45</sup>). Furthermore, all simulations have focused on the final stage of SNF and have not considered the different steps of the preceding developmental progression where metabolism remains poorly understood<sup>18</sup>.

Here, we report a holistic in silico representation of the integrated metabolism of the holobiont consisting of a *M. truncatula* plant nodulated by *S. meliloti*, which we refer to as a Virtual Nodule Environment (ViNE). Our combined, multi-compartment reconstruction accounts for the metabolic activity of shoot and root tissues together with a nodule consisting of five developmental zones. We report initial characterizations of ViNE using FBA, including zone-specific metabolic properties, trade-offs between nitrogen-fixation and plant growth, and the usage of dicarboxylates as a carbon source by bacteroids. Going forward, we expect ViNE will provide a powerful platform for hypothesis generation aimed at understanding and quantitatively evaluating SNF, as well as guiding attempts at engineering SNF for increased symbiotic efficiency.

## Results

### Validation of iGD1348, an updated *S. meliloti* reconstruction.

Prior to constructing the integrated plant – bacterium metabolic model, an updated metabolic reconstruction of *S. meliloti* Rm1021 was prepared as described in Methods and Supplementary Note 1. Briefly, the highly refined core metabolic network iGD726<sup>34</sup> was combined with the comprehensive accessory metabolism of the genome-scale metabolic network iGD1575<sup>33</sup>. Most of the reactions were manually curated through comparison against the literature, referenced where possible, and mass and charge balanced. The updated model consists of 1348 genes (Supplementary Table 1), and incorporates information from 240 literature sources (Supplementary Data 1) that includes transposon-sequencing (Tn-seq) data<sup>34</sup> and Phenotype Micro-Array data<sup>33,46,47</sup> for wild-type and mutant strains. The reduced size of iGD1348 relative to the older iGD1575 model is a consequence of deleting poorly characterized genes that appeared to be incorrectly associated with reactions, and the removal of pathways that produced dead-end metabolites.

Several tests were performed to validate the quality of the newly prepared *S. meliloti* reconstruction. Flux balance analysis (FBA) was used to simulate growth using glucose or succinate as the sole source of carbon, with or without the inclusion of an NGAM reaction. Inclusion of an NGAM reaction resulted in a specific growth rate reduction of  $\sim 0.043 \text{ h}^{-1}$  and  $0.030 \text{ h}^{-1}$  for

growth with glucose and succinate, respectively. This result confirmed the absence of energy leaks in iGD1348 that would allow for spontaneous energy production.

Using FBA, the ability of *S. meliloti* to catabolize 163 carbon sources to support growth was predicted with the iGD1348 and iGD1575 models (Supplementary Data 2). As previously reported<sup>33</sup>, simulations with the iGD1575 model correctly predicted growth with 67 of the 85 (79%) substrates experimentally shown to support growth of *S. meliloti*. Nicely, simulations with the iGD1348 model correctly predicted growth with 76 of these 85 (89%) substrates, including all 67 that supported growth of the iGD1575 model. This result confirmed that iGD1348 incorporates the majority of the accessory metabolism of *S. meliloti*, and that it is a better representation of total cellular *S. meliloti* metabolism than the previous genome-scale model.

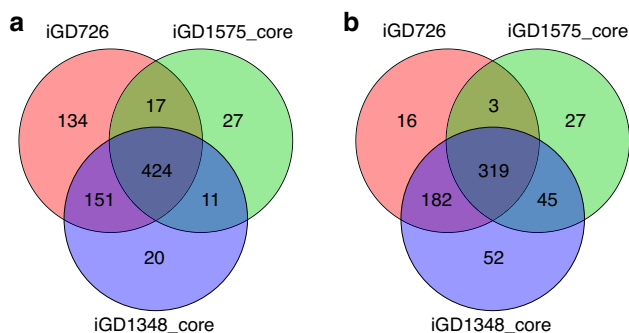
Context-specific core metabolic models, containing a minimal set of reactions for biomass production in a given environment, were extracted from the iGD1348 and iGD1575 metabolic models through the integration of Tn-seq data<sup>34</sup> using Tn-Core<sup>48</sup>. The accuracy of the resulting core metabolic models was determined through comparison with iGD726, a manually prepared core metabolic model of *S. meliloti*<sup>34</sup>. As summarized in Fig. 1, the iGD1348 core model displayed greater overlap with the iGD726 model than did the core model generated from iGD1575 (Fig. 1a). In particular, of the genes essential in the iGD726 model, 96% were essential in the iGD1348 core model, whereas only 62% were essential in the iGD1575 core model (Fig. 1b). This result confirmed that the newly prepared iGD1348 reconstruction better represents the core metabolic network of *S. meliloti* than does the iGD1575 reconstruction. Overall, these tests confirmed that iGD1348 is a high-quality representation of *S. meliloti* metabolism, and that it is an improvement over previous iterations of *S. meliloti* GENRES.

### Construction and validation of a model of a nodulated legume.

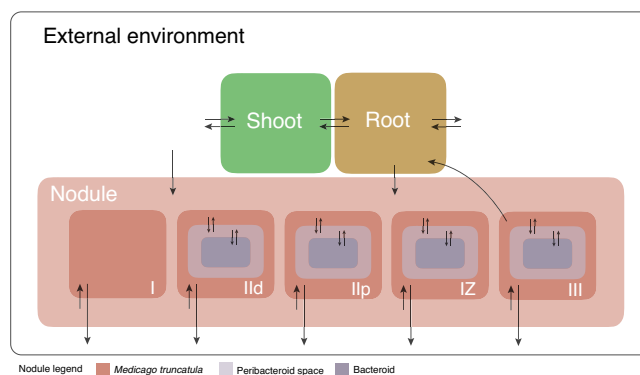
Obtaining high-quality reconstructions of *M. truncatula* and *S. meliloti* metabolism was a prerequisite to generating an in silico genome-scale metabolic network of an entire nodulated legume (referred to as ViNE). For *S. meliloti*, we used the updated model described in the previous section. In the case of *M. truncatula*, we used a recently published reconstruction that was updated to match the most recent version of the *M. truncatula* genome annotation (see Supplementary Note 2).

Integrating the *S. meliloti* and *M. truncatula* metabolic models resulted in a model encompassing shoot, root, and nodule tissues as summarized in Fig. 2 and Table 1. Memote<sup>49</sup> was used to evaluate the overall quality of the integrated reconstruction. This analysis revealed an overall quality score of 72%, which falls well in the range of the scores of the original two models (78% and 65% for the iGD1348 and the *M. truncatula* reconstructions, respectively) and is consistent with ViNE being a high-quality metabolic reconstruction. The detailed reports are provided in Supplementary Data 3.

In total, ViNE includes 746 unique *S. meliloti* genes and 1,327 unique *M. truncatula* genes. The nodule was subdivided into five zones to match the spatially, and transcriptionally, distinct developmental zones that are simultaneously present in indeterminate legume nodules such as those formed by *M. truncatula*<sup>21,50</sup>. Several simulations were performed to evaluate the reliability of the model. Using FBA, the maximal rate of plant (shoot + root) growth of the nodulated system was predicted to be  $\sim 0.044 \text{ g day}^{-1} (\text{g plant dry weight})^{-1}$ , while the  $\text{N}_2$ -fixation rate was predicted to be  $\sim 3 \mu\text{mol h}^{-1} (\text{g plant dry weight})^{-1}$ . Both values are reasonable predictions; *Medicago sativa* plants have an experimentally determined growth rate of  $\sim 0.1 \text{ g day}^{-1} (\text{g plant}$



**Fig. 1** Overlap between the iGD726 model and core metabolic models derived from iGD1575 and iGD1348. Venn diagrams illustrating the overlap in (a) the total gene content, and (b) the essential genes of the following three models: the manually prepared iGD726 core model, a core model derived from iGD1575, and a core model derived from iGD1348. Core models of iGD1575 and iGD1348 were prepared using Tn-Core and published Tn-seq data.



**Fig. 2** Visual depiction of ViNE. A schematic summarizing the overall structure of the *S. meliloti* nodulated *M. truncatula* plant developed in this work. The model contains three plant tissues (shoot, root, nodule) with the nodule subdivided into five developmental zones (I, II, IIp, IZ and III). Arrows indicate transport reactions with the direction representative of the directionality of the transport reactions. The scale of the figure has no meaning.

dry weight)<sup>-1</sup><sup>51</sup> and  $\text{N}_2$ -fixation rates around  $7 \mu\text{mol h}^{-1} (\text{g plant dry weight})^{-1}$  based on acetylene reduction assays<sup>52</sup>. Moreover, the predicted carbon cost of supporting  $\text{N}_2$ -fixing nodules (nodule  $\text{CO}_2$  export per  $\text{N}_2$  fixed) was  $\sim 4.2 \text{ g C g}^{-1} \text{ -N}$ , falling within the normal range of  $3\text{--}5 \text{ g C g}^{-1} \text{ -N}$  based on experimental measurements<sup>53</sup>.

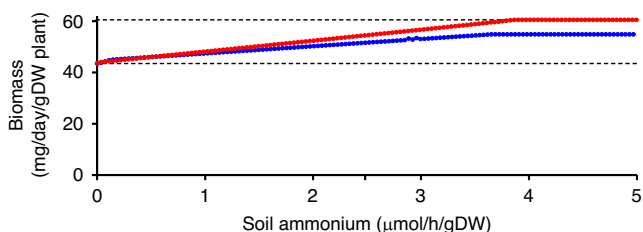
Next, FBA was used to examine the effects of adding exogenous ammonium to the soil on plant growth considering two situations: (i) the  $\text{N}_2$ -fixation efficiency (defined as the rate of  $\text{N}_2$ -fixation per gram nodule) could vary while the rate of nodulation (defined as the percent of total plant biomass being nodule biomass) was constant at 2%, and (ii) the rate of nodulation could vary while the  $\text{N}_2$ -fixation efficiency was constant at  $150 \mu\text{mol h}^{-1} (\text{g nodule dry weight})^{-1}$ . As expected, increasing the availability of exogenous ammonium increased the rate of plant growth, with the effect more pronounced when the rate of nodulation was allowed to decrease with increasing ammonium since the plant no longer had to invest in nodule maintenance (Fig. 3).

We then simulated the effects of individual bacteria gene deletions on plant biomass production (Supplementary Data 4)

**Table 1 Summary properties of ViNE.**

Model feature	Shoot	Root	Zone I	Zone II <sub>d</sub>	Zone II <sub>p</sub>	Zone IZ	Zone III
Genes							
<i>M. truncatula</i>	1295	1292	236	265	243	228	116
<i>S. meliloti</i>	0	0	0	640	629	638	246
Reactions <sup>a</sup>							
<i>M. truncatula</i>	937	944	494	543	559	530	318
<i>S. meliloti</i>	0	0	0	662	654	670	201
Metabolites							
<i>M. truncatula</i>	831	825	490	568	597	581	286
<i>S. meliloti</i>	0	0	0	751	747	764	222

<sup>a</sup>Excludes reactions for transfer of metabolites between tissues or between *M. truncatula* and the peribacteroid space.



**Fig. 3 Effect of exogenous ammonium on *M. truncatula* growth.** The effects of increasing the availability of soil ammonium on the growth rate of nodulated *M. truncatula* was examined. Simulations were run allowing either the  $N_2$ -fixation efficiency to vary while the rate of nodulation remained constant at 2% (blue) or allowing the rate of nodulation to vary while the  $N_2$ -fixation efficiency remained constant at  $150 \mu\text{mol h}^{-1} (\text{g nodule dry weight})^{-1}$  (red). The dashed lines indicate the maximal rate of plant growth with exogenous ammonium (upper) and the maximal rate of plant growth when relying on  $N_2$ -fixation (lower).

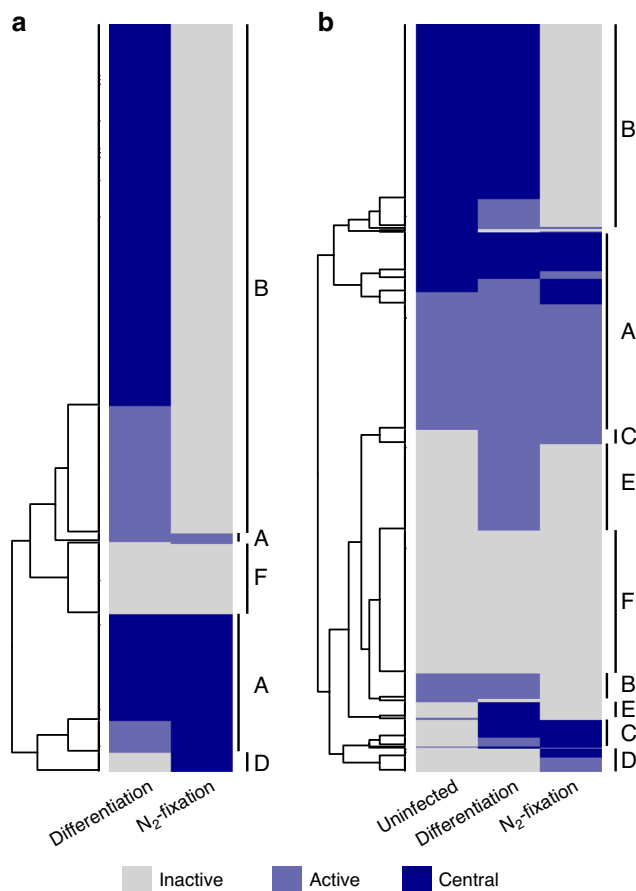
and compared the results to published experimental data. The model was able to accurately predict the phenotypes of many *S. meliloti* mutants. For example, *S. meliloti* genes such as *nifH* (nitrogenase), *dctA* (succinate transport), *ilvI* (branched-chain amino acid biosynthesis), *aatA* (aspartate transaminase), *pgk* (phosphoglycerate kinase), and *nrdJ* (ribonucleotide reductase) were correctly predicted to be essential, while *pyc* (pyruvate carboxylase), *glnA* (glutamine synthetase), *pckA* (phosphoenolpyruvate carboxykinase), and *leuB* (leucine biosynthesis) were correctly predicted to be non-essential<sup>54–61</sup>. Similarly, the removal of plant-encoded nodule sucrose synthase, phosphoenolpyruvate carboxylase, and homocitrate synthase reactions abolished nitrogen fixation, as expected<sup>62–64</sup>. However, it is important to note that the predictions were not perfect. For example, deleting *argG* (arginine biosynthesis) or *carA* (carbamoyl phosphate synthase) did not result in the expected phenotypes, while the incorrect malic enzyme (*tme* instead of *dme*) was predicted to be essential<sup>65–67</sup>. Taken together, these analyses provide support for the general reliability of ViNE as a representation of nodule metabolism.

**Metabolic progression and nutrient exchange during nodule development.** The presence of five nodule zones in ViNE provided an opportunity to examine the metabolic changes associated with the development of an effective nodule. To accomplish this, FBA was used to predict the flux distribution through the integrated metabolic networks of each nodule zone, and to simulate the effects of individually deleting each gene, or removing each reaction, specifically in a single nodule zone.

Additionally, a robustness analysis was performed to evaluate how perturbations in the flux of individual bacteroid reactions influence the predicted rate of plant growth. The outputs of these analyses are provided as Supplementary Data 5 and 6, and they are summarized in Fig. 4 and Supplementary Figs. S1 and S2. For simplicity, here we focus on the reaction-level analyses, and we split the nodule into only three sections: uninfected (zone I), differentiating (zones II<sub>d</sub>, II<sub>p</sub>, and IZ), and nitrogen-fixing (zone III). Highlighting the overall similarity of zones II<sub>d</sub>, II<sub>p</sub>, and IZ, and thus supporting their grouping, the robustness analysis indicated that roughly 90% of the bacteroid reactions that had to carry flux in one of these zones had to carry flux in all three zones to maximize plant growth.

The most notable difference comparing the uninfected and differentiating zones was an increase in the number of active reactions related to energy production, including carbon and nucleotide metabolism. This result suggests that the accommodation of differentiating bacteroids may place additional energy demands on the plant cell, and that few additional metabolic functions are required. In contrast, the transition from the differentiating zone to the  $N_2$ -fixing zone was associated with a marked decrease in the number of active reactions in both the plant and bacterial cells, consistent with published transcriptomic and proteomic datasets<sup>50,68–70</sup>. Highlighting this result, ~ 560 bacteroid reactions had to carry flux in the differentiating zones to optimize plant growth, whereas only 167 bacteroid reactions had to carry flux in the  $N_2$ -fixation zone for maximal plant growth.

The lack of biomass production in the  $N_2$ -fixation zone meant that most biomass biosynthetic pathways were predicted to be inactive and non-essential. However, bacterial pathways related to the production of cofactors for nitrogenase or energy production remained essential; this included FMN, heme, cobalamin, pyridoxine phosphate, and glutathione biosynthesis, as well as the pentose phosphate pathway. Similarly, the TCA cycle, oxidative phosphorylation, and purine biosynthesis in bacteroids were predicted to be essential in the  $N_2$ -fixation zone, presumably to supply the massive amounts of energy required by nitrogenase. Biosynthesis of methionine and SAM were also predicted to be essential. Few other notable bacterial reactions were required in the  $N_2$ -fixation zone (Supplementary Data 5). In the plant compartment, the majority of the active reactions were related to central carbon metabolism for the production of energy and  $C_4$ -dicarboxylates for use by the bacteroids, while other active reactions were involved in the assimilation of ammonium through the formation of glutamine. Consistent with experimental works [reviewed by<sup>18,71</sup>], the FBA results indicated that the plant nodule cells are provided sucrose as a carbon/energy source; in fact, ~ 30% of all carbon fixed by the plant leaves was sent to nodule zone III. The sucrose was then hydrolyzed and metabolized to phosphoenolpyruvate, of which ~ 80% was



**Fig. 4 Nodule zone-specific metabolism.** Heatmaps are presented displaying which reactions are inactive (gray), active but whose removal does not impair plant growth (light blue), or central (dark blue; growth reduction > 10% for the model missing the reaction compared to the full model) in the different nodule zones. In the differentiation zone, reactions are marked as central only if it was central in each of zone Ild, IIp, and IZ. Heatmaps are shown for (a) *S. meliloti* reactions and (b) *M. truncatula* reactions. Reactions were clustered using hierarchical clustering, and the following main clusters were identified: A - constitutively active; B - specific to growing cells; C - specific to infected cells; D - specific to the nitrogen-fixation zone; E - specific to the differentiation zone; F - constitutively inactive.

diverted to oxaloacetate through a cytoplasmic phosphoenolpyruvate carboxykinase reaction for use in the production of C<sub>4</sub>-dicarboxylates.

Next, nutrient exchange between the plant and bacterial partners was examined. While the prevailing evidence suggests C<sub>4</sub>-dicarboxylates (succinate, malate, fumarate) are the primary carbon source for N<sub>2</sub>-fixing bacteroids<sup>18,72–74</sup>, the source of carbon for differentiating bacteroids has not been established. The FBA results suggested that differentiating bacteroids primarily use sugars, likely sucrose, as a carbon source. This is consistent with micrographic evidence suggesting that bacterial mutants unable to use C<sub>4</sub>-dicarboxylates can undergo at least partial differentiation<sup>72,73</sup>. Currently, it is commonly accepted that nitrogen is primarily exported from bacteroids as ammonia<sup>75,76</sup>; however, some studies have suggested that L-alanine could be a major nitrogen export product<sup>77,78</sup>. Our FBA simulations were consistent with ammonia being the primary export product in the *S. meliloti* – *M. truncatula* symbiosis. However, prior to constraining the nodule reaction space, reducing the availability of oxygen to the bacteroids resulted in

**Table 2 Contributions of N<sub>2</sub>-fixation and nodulation to the metabolic costs of SNF.**

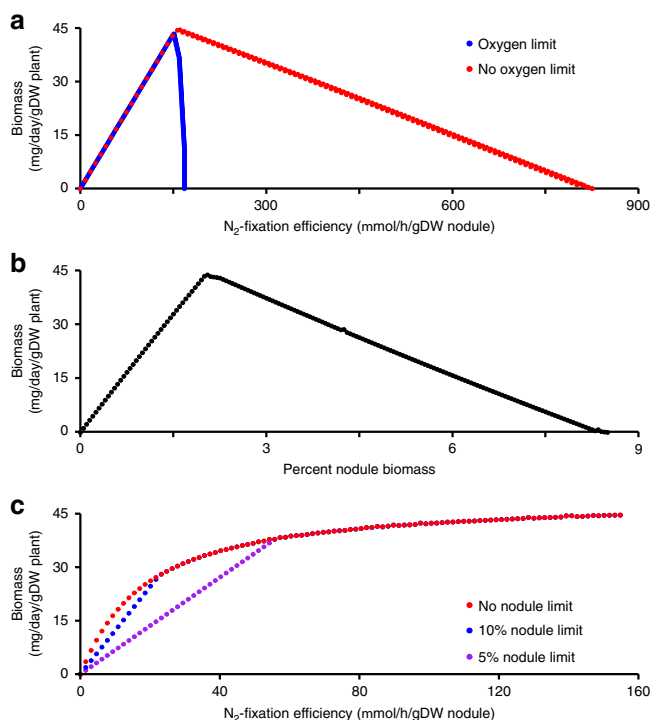
Nitrogen source	Nodulation state	Relative plant growth rate
N <sub>2</sub> -fixation	Nodulated	0.717
N <sub>2</sub> -fixation	Nodulated but without maintenance costs	0.781
N <sub>2</sub> -fixation	Non-nodulated <sup>a</sup>	0.812
Exogenous ammonium	Non-nodulated <sup>a</sup>	1

<sup>a</sup>Non-nodulated indicates that nodule biomass was excluded from the overall biomass reaction and nodule maintenance costs were removed. As N<sub>2</sub>-fixation in ViNE is unlinked to the production or maintenance of nodule biomass (unless stated otherwise), N<sub>2</sub>-fixation can occur without nodulation.

a shift in the nitrogen export product from ammonia to L-alanine. Thus, the detection of L-alanine versus ammonia as an export product could be due, in part, to differences in experimental set-up that may influence oxygen availability to the bacteroid. Also, experimental data suggest that rhizobial biosynthesis of some amino acids is essential for the symbiosis while the biosynthesis of others is not, and that the phenotypes may be symbiosis-specific [reviewed by<sup>79</sup>]. Similarly, our FBA simulations suggested that rhizobial biosynthesis of approximately half of the amino acids was essential for the symbiosis.

**Sociobiology of symbiotic nitrogen fixation.** By containing a representation of an entire nodule, ViNE allowed for an evaluation of the metabolic costs associated with SNF. Using FBA, the maximal plant growth rate of the nodulated system (without exogenous ammonium) was estimated to be ~72% of the maximal growth rate of a nodule-free system supplied with non-limiting amounts of exogenous ammonium (Fig. 3, Table 2). The largest factor contributing to the difference in growth was the direct energetic cost of supporting N<sub>2</sub>-fixation (~67% of the difference; Table 2). The remaining third of the difference was explained by the cost of synthesizing (~11% of the difference) and maintaining (~22% of the difference) the nodule and bacteroid tissue (Table 2).

We next evaluated the relationship between N<sub>2</sub>-fixation efficiency (while keeping the rate of nodulation at 2%) and the rate of plant growth. When the N<sub>2</sub>-fixation efficiency was below the optimum, there was a linear relationship between N<sub>2</sub>-fixation and biomass production (Fig. 5a). However, excessive N<sub>2</sub>-fixation quickly resulted in impaired plant growth, with a 10% excess of N<sub>2</sub>-fixation collapsing the symbiosis (Fig. 5a). We hypothesized that this result was due to insufficient energy to support both the excess N<sub>2</sub>-fixation and the ATP maintenance costs. Consistent with this hypothesis, removing the upper limit on the rate of zone III oxygen uptake resulted in a gradual decrease in plant growth as the rate of N<sub>2</sub>-fixation was increased above the optimal (Fig. 5a). In this case, excessive N<sub>2</sub>-fixation was less detrimental than insufficient N<sub>2</sub>-fixation; the effect of increasing the N<sub>2</sub>-fixation efficiency by 50 μmol h<sup>-1</sup> (g nodule dry weight)<sup>-1</sup> increased or decreased the rate of plant growth by 14.7 or 3.4 mg day<sup>-1</sup> (g plant dry weight)<sup>-1</sup> when below or above the optimum, respectively. We next examined the consequences of varying the rate of nodulation (i.e., the ratio between plant and nodule biomass) while maintaining a constant N<sub>2</sub>-fixation efficiency. The simulations demonstrated linear relationships between the rate of nodulation and plant growth both above and below the optimum (Fig. 5b), with increasing the percent nodulation resulting in a 3-fold greater impact when below the optimum compared to above the optimum. Overall, these



**Fig. 5 Relationships between plant growth and rate of N<sub>2</sub>-fixation or nodulation.** In all panels, the sole source of nitrogen for the plant is through N<sub>2</sub>-fixation. **a** Pareto frontiers showing the relationship between the N<sub>2</sub>-fixation efficiency (with a constant rate of nodulation of 2%) and the rate of plant biomass production using ViNE with default parameters (blue) or no limit on zone III oxygen usage (red). **b** The relationship between the rate of nodulation and the rate of plant biomass production (with a constant N<sub>2</sub>-fixation efficiency of 150 μmol h<sup>-1</sup> (g nodule dry weight)<sup>-1</sup>). **c** The effect of the N<sub>2</sub>-fixation efficiency on the rate of plant growth, with the amount of nodule biomass optimized to maximize plant growth and without a limit on zone III oxygen uptake (see Supplementary Fig. 3 for simulations with an oxygen uptake limit). Nodule biomass was either uncapped (red) or limited to 10% (blue) or 5% (purple) of the overall plant + nodule biomass.

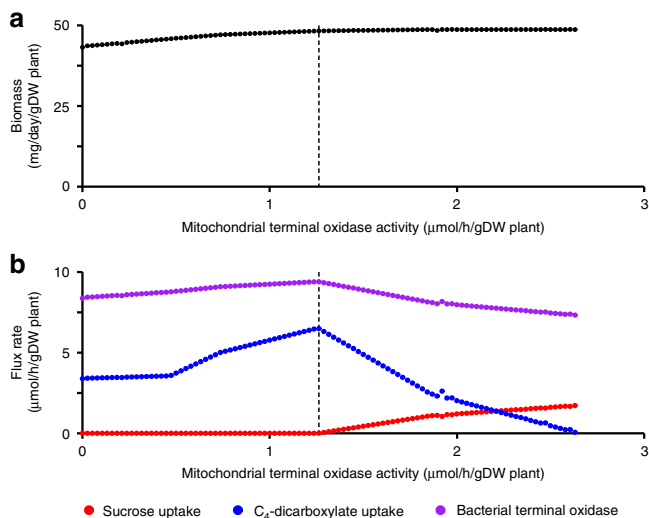
simulations suggest that a slightly too efficient symbiosis is preferable (for plant biomass production) over a slightly inefficient symbiosis, unless the required rate of O<sub>2</sub> usage exceeds the nodule oxygen diffusion limit.

The previous simulations represented simple scenarios where only a single variable differed. In reality, a change in the N<sub>2</sub>-fixation efficiency should be accompanied by a change in the rate of nodulation as a result of legume autoregulation of nodulation<sup>80</sup>. We therefore ran simulations where the N<sub>2</sub>-fixation efficiency was varied and the rate of nodulation was optimized to maximize plant growth. Strikingly, the simulations suggested a pattern of diminishing returns associated with increasing the N<sub>2</sub>-fixation efficiency (Fig. 5c and Supplementary Fig. 3); decreasing the N<sub>2</sub>-fixation efficiency 50% from the maximum tested value resulted in a mere 10% decrease in plant growth. The half-maximal growth rate was achieved with a N<sub>2</sub>-fixation efficiency of just 10% the maximal, although this required that the nodule accounted for almost 13% of the total biomass. If we assume an upper limit of nodulation at 10% or 5% of the total biomass, the benefits of low N<sub>2</sub>-fixation efficiencies are decreased although the pattern of diminishing returns remains (Fig. 5c and Supplementary Fig. 3). In these cases, half-maximal plant growth rate is achieved at 12% or 21%, respectively, of the highest tested N<sub>2</sub>-fixation efficiency. Overall, these simulations support that even a poor symbiosis is likely to provide a noticeable benefit to the plant.

**H<sup>+</sup> and O<sub>2</sub> influence on the carbon source provided to bacteroids.** It is well-established that C<sub>4</sub>-dicarboxylates (malate, succinate, fumarate) are the primary carbon source provided to nitrogen-fixing zone III bacteroids<sup>18,72,73</sup>; however, the reason for this remains unclear. We therefore attempted to uncover a metabolic explanation using ViNE. Surprisingly, preliminary FBA simulations with the ViNE precursor model (i.e., prior to constraining the reaction space) suggested that the N<sub>2</sub>-fixing bacteroids of nodule zone III are provided sucrose, not C<sub>4</sub>-dicarboxylates, as the primary carbon source. Unexpectedly, forcing the use of C<sub>4</sub>-dicarboxylates resulted in the model being unable to fix nitrogen or produce plant biomass. During those simulations, protons of the plant cytosol could be transferred to the peribacteroid space but were not allowed to be used by the N<sub>2</sub>-fixing bacteroids. However, this may not be realistic since the peribacteroid space of N<sub>2</sub>-fixing bacteroids is acidic due to import of protons from the plant cytosol<sup>81–83</sup>. If the analysis was repeated and the zone III bacteroids were provided access to the protons of the peribacteroid space, it became possible for C<sub>4</sub>-dicarboxylates to serve as the primary carbon source and support N<sub>2</sub>-fixation and plant growth. These results suggest that the plant-driven acidification of the peribacteroid space is essential for the metabolic functioning of the bacteroid.

Although the transfer of protons to the periplasm allowed C<sub>4</sub>-dicarboxylates to support N<sub>2</sub>-fixation, the rate of plant biomass production nevertheless remained higher when the N<sub>2</sub>-fixing bacteroids were provided sucrose instead of C<sub>4</sub>-dicarboxylates. To further investigate this difference, ViNE was modified to contain reactions for the transport and metabolism of sucrose by N<sub>2</sub>-fixing bacteroids (see Supplementary Note 3). Consistent with results from the precursor model, FBA simulations suggested the ability of bacteroids to use sucrose (plus C<sub>4</sub>-dicarboxylates) increased the plant growth rate by 6.4% relative to when bacteroids were supplied only C<sub>4</sub>-dicarboxylates. ViNE contains a limit on the rate of oxygen uptake by zone III nodule tissue as the concentration of free oxygen in the N<sub>2</sub>-fixation zone is known to be low<sup>84</sup> to protect the highly oxygen-sensitive nitrogenase enzyme. The expected consequence of this is a constriction in the rate of flux through the electron transport chain, thus, restricting nodule and bacteroid metabolism by limiting ATP production and the removal of reductant. We wondered whether the use of sucrose versus C<sub>4</sub>-dicarboxylates may be modulated by the free oxygen concentration of the nodule. The concentration of free oxygen in the N<sub>2</sub>-fixation zone has been experimentally demonstrated to be <50 nM<sup>84</sup>. Notably, the K<sub>m</sub> values of the mitochondrial and bacterial terminal oxidases towards oxygen are 50–100 nM<sup>85,86</sup> and 7 nM<sup>87</sup>, respectively. These enzyme kinetics suggest that the metabolism of the plant fraction, but not the bacteroid fraction, of the nodule is likely to be oxygen limited<sup>88,89</sup>, a conclusion that is supported by measurements of nodule adenylate pools<sup>90</sup>. Therefore, we ran a series of simulations in which the upper limit of the mitochondrial terminal oxidase reaction of zone III was varied, with no overall limit on the use of oxygen by the nodule. Gradually reducing the flux through the mitochondrial terminal oxidase was associated with a gradual replacement of sucrose with C<sub>4</sub>-dicarboxylates as the carbon source provided to N<sub>2</sub>-fixing bacteroids (Fig. 6). This result is consistent with the hypothesis that the low free oxygen concentration of the N<sub>2</sub>-fixation zone could be a contributing factor as to why bacteroids are provided C<sub>4</sub>-dicarboxylates, and not sugars, as the primary carbon source.

Assuming the nodule (consisting primarily of zone III tissue) accounts for 2% of total plant biomass, and that bacteroid biomass accounts for 25% of nodule biomass, the maximal rate of predicted C<sub>4</sub>-dicarboxylate import by N<sub>2</sub>-fixing bacteroids (1.3 mmol h<sup>-1</sup> [g bacteroid dry weight]<sup>-1</sup>) was similar to



**Fig. 6 Effects of limiting the mitochondrial terminal oxidase of the  $N_2$ -fixing zone.** FBA simulations were run with a modified version of ViNE in which bacteroid metabolism could be supported by sucrose in addition to  $C_4$ -dicarboxylates. No overall limit on oxygen usage of the nodule was set during the simulations, but a limit was set on the activity of the mitochondrial terminal oxidase of the zone III nodule tissue. **a** The effect on plant growth rate of varying the mitochondrial terminal oxidase of the zone III nodule tissue. **b** The effect on specified flux rates of varying the mitochondrial terminal oxidase of the zone III nodule tissue. Red - the sucrose uptake rate of  $N_2$ -fixing bacteroids; blue - the uptake rate of  $C_4$ -dicarboxylates of  $N_2$ -fixing bacteroids; purple - the flux rate of the terminal oxidase of  $N_2$ -fixing bacteroids. The dashed line indicates the mitochondrial terminal oxidase flux rate below which no sucrose is used by the bacteroids.

experimentally determined uptake rates by *S. meliloti* bacteroids ( $1.1$  to  $1.3$   $\text{mmol h}^{-1}$  [ $\text{g bacteroid dry weight}^{-1}$ ])<sup>35,91</sup>. Considered together, these simulations provide evidence that  $C_4$ -dicarboxylates can support optimal plant growth under physiologically relevant conditions.

## Discussion

Models of the integrated metabolism of various holobionts (consisting of a host and its symbiotic microorganisms) would be valuable tools to understand the emergent properties of these systems<sup>92,93</sup>. However, to date there are few examples of constraint-based metabolic modeling being used to study metabolic interactions [e.g.<sup>8,31</sup>], with this approach most commonly used to study the human gut microbiome<sup>94</sup>. Here, we developed a broadly adaptable pipeline for modeling the metabolism of interacting organisms across physiologically distinct tissue (sub) sections. Using metabolic network reconstruction and constraint-based modeling, we studied the metabolism of a legume root nodule and SNF, a well-established model of inter-organismal metabolic exchange and cellular differentiation.

Our model (ViNE) accounts for plant shoot, root, and nodule tissues, with the nodule encompassing the metabolism of both the plant and bacterial partners and subdivided into five developmental zones. This is an advance over previous attempts at modeling SNF<sup>33,35–40</sup>, most of which focused solely on bacterial metabolism while treating the plant as a black box. The increased complexity of ViNE allows for: i) more accurate simulations of the nutrient exchange, ii) analysis of the metabolic differentiation associated with nodule development, iii) examination of unexpected emergent properties of the symbiosis resulting from inter-organism interactions, and iv) the possibility to perturb the

network at the single reaction level. Initial simulations with ViNE supported that this model does a good job at capturing the metabolism of a legume nodule. Nevertheless, as with all models, ViNE predictions were imperfect; of the 38 genes both present in ViNE and listed as being involved in symbiosis according to Additional File 5 of Galardini et al.<sup>95</sup> (excluding genes involved in early nodulation, a process absent in ViNE), deletion of 84% were correctly predicted to have a symbiotic phenotype in ViNE. However, as we often compared simulated phenotypes for *M. truncatula* with experimental data for *M. sativa*, and given that rhizobium mutant phenotypes are often plant specific (e.g.<sup>96–98</sup>), we cannot rule out that some of the inconsistencies are the result of plant-specific phenotypes. Going forward, we intend to continue to manually refine and update ViNE to maximize consistency with experimental observations.

FBA simulations with ViNE revealed a pattern of diminishing returns in terms of plant growth (as a proxy for fitness) as the  $N_2$ -fixation efficiency increased, assuming that the rate of nodulation could also vary (Fig. 5c). This observation has potential implications for engineering SNF for biotechnological applications. It suggests that when developing rhizobial inoculants, maximizing competition for nodule occupancy may have a greater impact than maximizing the rate of  $N_2$ -fixation. This result also supports efforts aimed at engineering  $N_2$ -fixing symbiosis with cereals<sup>99</sup> by highlighting how even a low-efficiency symbiosis has the potential to have a noticeable benefit on crop yield.

At the same time, the pattern of diminishing returns is interesting from an evolutionary perspective<sup>100</sup>. In particular, the evolution of  $N_2$ -fixation efficiency may be influenced by the rhizobium community diversity, assuming that nodule infection increases the fitness of rhizobia<sup>101</sup>. In an environment dominated by a single rhizobium, kin selection may favor the evolution of a poorly efficient symbiosis as it would increase nodule number and thus the size of the niche for colonization by the rhizobia. On the other hand, in a highly diverse environment, evolution of strains capable of entering into a highly efficient symbiosis may be favored, as this would lead to fewer nodules and thus less plant resources being allocated to competing rhizobium strains, thereby limiting the spread of less mutualist (viz. cheater) strains<sup>20,102</sup>.

Of particular interest to us were the metabolic exchanges between the plant and rhizobia, both during  $N_2$ -fixation and during differentiation. The carbon source(s) of differentiating rhizobia remain poorly understood. Results with ViNE suggested that sucrose may be a major carbon source for the differentiating bacteroids. However, *S. meliloti* mutants unable to transport sucrose are not impaired in nodule formation<sup>103</sup>, suggesting that differentiating bacteroids have access to at least one other carbon source. Interestingly, a *S. meliloti pyc* mutant unable to grow with glycolytic carbon sources was not impaired in differentiation<sup>60</sup>. Similarly, *S. meliloti pckA*<sup>58</sup> and *tpi*<sup>104</sup> mutants unable to grow with gluconeogenic carbon sources remained capable of differentiating. Thus, it seems likely that differentiating bacteroids have access to a variety of glycolytic and gluconeogenic carbon substrate, with sugars possibly serving as the main carbon source in wild type nodules. If so, the restriction of carbon flow to  $N_2$ -fixing bacteroids to just  $C_4$ -dicarboxylates may be the result of active remodeling of the peribacteroid membrane during differentiation.

In attempting to identify conditions favoring the use of  $C_4$ -dicarboxylates as a carbon source by  $N_2$ -fixing bacteroids, ViNE also provided insights into the metabolic exchange in the  $N_2$ -fixation zone. The peribacteroid space of  $N_2$ -fixing bacteroids is known to be acidic due to the activity of  $H^+$ -ATPases on the peribacteroid membrane<sup>81–83</sup>. This acidification contributes to the import of  $C_4$ -dicarboxylates and the export of ammonium from/to the plant cytosol and the peribacteroid space<sup>105</sup>, and it

may contribute to the lysis of non-functional symbiosomes<sup>106</sup>. Our FBA simulations suggest that the plant-derived protons of the peribacteroid space may also be actively used by the bacteroid to support its metabolism.

Although it is generally accepted that nodules are low oxygen environments<sup>84</sup>, the site of O<sub>2</sub>-limitation has been debated. Based on the average concentration of free oxygen in the nodule, enzyme kinetics data are consistent with the mitochondria being O<sub>2</sub>-limited and the bacteroids being O<sub>2</sub>-sufficient<sup>84–87</sup>. Measurements of the adenylate pools of the plant and bacterial nodule fractions support this conclusion<sup>90</sup>. However, others have argued that nodule adenylate measurements suggest that bacteroids, not the plant, are O<sub>2</sub>-limited<sup>107</sup>. Similarly, it was suggested that mitochondria cluster near the periphery of the cell near air pockets, resulting in elevated local O<sub>2</sub> concentrations<sup>108,109</sup>. The FBA results presented here predicted that C<sub>4</sub>-dicarboxylates are the optimal carbon source for N<sub>2</sub>-fixing bacteroids only when the plant mitochondria are O<sub>2</sub>-limited while the bacteroids are O<sub>2</sub>-sufficient (Fig. 6). This result supports the hypothesis that mitochondria, and not bacteroids, are O<sub>2</sub>-limited in wild type nodules.

Although potentially powerful, the use of metabolic modeling to study SNF is not without limitations. In particular, the accuracy of predictions is restricted by the quality of the imposed flux constraints, and unfortunately, experimental kinetic data for key enzymes and nutrient exchange reactions in the nodule is generally lacking. FBA also fails to actively incorporate regulatory feedback control during simulations that could influence the metabolic properties of the nodule. Furthermore, the lack of finished legume genomes, an incomplete ability to ensure correct subcellular compartmentalization in eukaryotic cells, and difficulty in experimentally validating the functions of plant genes can limit the quality of models of SNF.

In sum, this work presented a complex metabolic model representing the full metabolism of a rhizobium-nodulated legume, as well as a series of simulations demonstrating the potential for this model to help address genetic, evolutionary, metabolic, and sociobiological questions. Future work will be aimed at continuing to refine and improve the quality of the model, and to using it to generate hypotheses to guide experimental studies and to assist in the interpretation of experimental datasets.

## Methods

**Construction of a virtual nodule environment.** Before starting the reconstruction of the integrated metabolic model, we refined/updated the existing reconstructions for *S. meliloti* and *M. truncatula*. Concerning *S. meliloti*, we started from the existing core metabolic reconstruction iGD726<sup>34</sup>, which is a highly curated model with 726 genes and 681 reactions covering the pathways required to produce all biomass components when provided glucose, succinate, or sucrose as a carbon source, and made the following changes: (i) updated the biomass composition (Supplementary Data 7), (ii) manually added and curated accessory metabolic pathways, (iii) performed an automated expansion based on a slightly modified version of the iGD1575 model reported in<sup>34</sup>, (iv) mass and charge balanced all reactions, and (v) added an ATP hydrolysis reaction to account for non-growth associated maintenance (NGAM) costs. The final *S. meliloti* reconstruction, termed iGD1348, is available in Supplementary Data 1 in SBML, XLS, and MATLAB COBRA format. In terms of the *M. truncatula* reconstruction, a recently published reconstruction<sup>40</sup> was used in this work. Since that version was based on the *M. truncatula* genome annotation version Mt3.5v5<sup>42</sup>, the gene names in the reconstruction were updated to match the most recent annotation (version 5.0;<sup>44</sup> see Supplementary Note 2 for the details of this procedure, and Supplementary Data 8 for the model in SBML format). In addition, we included an energetic cost (0.25 mol of ATP per mol of transported compound) for all single-metabolite diffusion reactions to limit inappropriate transport between compartments. Next, a tissue-specific *M. truncatula* model containing shoot and root tissues was generated using the ‘BuildTissueModel’ function of Pfau et al.<sup>40</sup>, all genes were relabelled according to the different model compartments (e.g., ‘Root\_’), and a few ion transport reactions were added. See Supplementary Notes 1 and 2 for additional details on preparation of the *S. meliloti* and *M. truncatula* GENRES, respectively.

Afterward, the *S. meliloti* and *M. truncatula* reconstructions were integrated following the procedure described in detail in Supplementary Notes 4–6, and all reactions producing dead-end metabolites were removed (Supplementary Data 9). This pipeline generated an integrated, multi-compartment metabolic reconstruction embedding seven overall compartments: shoot tissue, root tissue, the nodule apical meristem (nodule zone I) and four nodule developmental zones representing the combined metabolism of the plant and the bacterium (zone II distal, zone II proximal, interzone II-III, and the nitrogen-fixing zone III). Additionally: (i) the nodule compartments were metabolically connected to the root compartment using appropriate nutrient exchange reactions, (ii) the reaction space of each nodule zone was constrained based on the *M. truncatula* – *S. meliloti* zone-specific RNA-seq data of Roux and coworkers<sup>50</sup>, which was reanalyzed as described in Supplementary Note 7, and (iii) biomass reactions were prepared for each compartment and the biomass of each compartment was combined using appropriate ratios estimated from the literature. Unless stated otherwise, the shoot and root tissue accounted for 98% of biomass (at a 2:1 ratio) and the nodule tissue collectively accounted for 2% of biomass. The final version of the integrated model was named ViNE; it is schematically represented in Fig. 2 and it is provided in Supplementary Data 10 as MATLAB COBRA and SBML formatted files.

**Metabolic modeling procedures.** Model integration, model manipulations, and FBA simulations were performed in MATLAB R2016b (mathworks.com) using the SBMLToolbox version 4.1.0<sup>110</sup>, libSBML version 5.13.0<sup>111</sup>, and scripts from the COBRA Toolbox commit 9b10fa1<sup>112</sup>, the TIGER Toolbox version 1.2.0-beta<sup>113</sup>, FASTCORE version 1.0<sup>114</sup>, and the Tn-Core Toolbox version 2.2<sup>48</sup>. The iLOG CPLEX Studio 12.7.1 solver (ibm.com) was used for nearly all FBA simulations; the exception was for the preparation of iGD1348, during which the Gurobi version 7.0.1 solver (gurobi.com) was used. The switch to CPLEX was prompted by numerical issues that were solved by switching the solver. All custom scripts used in this study are available through a GitHub repository ([github.com/diCenzo-GC/ViNE\\_Reconstruction](https://github.com/diCenzo-GC/ViNE_Reconstruction)).

Each gene found in multiple tissues or nodule zones was distinguished by a unique gene name to facilitate tissue-specific gene deletion analyses. When performing global single or double gene deletion analyses, all versions of the gene were simultaneously deleted followed by the removal of all constrained reactions. In contrast, zone- or tissue-specific gene deletion analyses involved deleting just the gene version specific to the zone or tissue of interest. Flux variability analyses were performed with the requirement that flux through the objective function was at least 99% of the optimal flux. The robustness analyses involved first identifying the approximate flux range for each reaction in which the plant growth rate was non-zero. Then, for each reaction, the flux rate of the reaction was set to various values within the previously identified flux range, and the objective value was maximized. For simulations in which the rate of nodulation could vary, nodule biomass was removed from the objective reaction and instead forced through a nodule biomass sink reaction at the appropriate rate; maintenance costs and oxygen availability were modified accordingly (see Supplementary Note 8, for details). For simulations comparing the effect of providing zone III bacteroids sucrose versus C<sub>4</sub>-dicarboxylates as the carbon source, a modified version of ViNE was prepared as described in Supplementary Note 3.

Except when stated otherwise, the objective function for all simulations was:

$$0.65333 \text{ Shoot biomass} + 0.32666 \text{ Root biomass} + 0.00100 \text{ Nodule Zone I biomass} + 0.00675 \text{ Nodule Zone IId biomass} + 0.00225 \text{ Bacteroid Zone IId biomass} + 0.00675 \text{ Nodule Zone IIp biomass} + 0.00225 \text{ Bacteroid Zone IIp biomass} + 0.00075 \text{ Nodule Zone IZ biomass} + 0.00025 \text{ Bacteroid Zone IZ biomass} \rightarrow 1.00 \text{ Biomass}$$

Except when stated otherwise, the following constraints were placed on ViNE during the simulations: (i) when N<sub>2</sub>-fixation was the sole source of nitrogen, the uptake of nitrate and ammonia by the roots was set to 0; (ii) the rate of oxygen uptake by zone III tissue was 8.9847 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>; (iii) the rate of light uptake by the leaves was 1000 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>; (iv) the rate of CO<sub>2</sub> uptake by the leaves was 1000 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>; (v) maintenance costs (ATP hydrolysis) for shoot and root tissue were ~74 and 37 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>, respectively; (vi) maintenance costs (ATP hydrolysis) for zone I, IId, IIp, IZ, and III plant nodule tissue were ~111, 750, 750, 83, and 833 nmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>, respectively; (vii) maintenance costs (ATP hydrolysis) for zone IId, IIp, IZ, and III bacteroid tissue were ~5.7, 5.7, 0.6, and 6.3 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup>, respectively; and (viii) the only carbon sources available to bacteroids in the zone III nodule tissue was succinate, malate, and fumarate at a maximal rate of 1000 μmol h<sup>-1</sup> (g plant dry weight)<sup>-1</sup> each. No limit on the carbon sources available to the bacteroids of other nodule zones was set.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The authors declare that the data supporting the findings of this study are available within the paper (in Supplementary Data 1–10). No restrictions apply to data availability.



**Code availability**

Code and scripts used to generate the integrated reconstruction are available at [https://github.com/diCenzo-GC/ViNE\\_Reconstruction](https://github.com/diCenzo-GC/ViNE_Reconstruction).

Received: 22 October 2019; Accepted: 28 April 2020;

Published online: 22 May 2020

**References**

- Theis, K. R. et al. Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. *mSystems* **1**, e1002226 (2016).
- Vandenkoomhuyse, P., Quaiser, A., Duhamel, M., Le Van, A. & Dufresne, A. The importance of the microbiome of the plant holobiont. *N. Phytol.* **206**, 1196–1206 (2015).
- McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl Acad. Sci. USA* **106**, 15394–15399 (2009).
- Smith, S. E., Jakobsen, I., Grønlund, M. & Smith, F. A. Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorus acquisition. *Plant Physiol.* **156**, 1050–1057 (2011).
- den Besten, G. et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res* **54**, 2325–2340 (2013).
- Heinken, A. et al. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome* **7**, 75–18 (2019).
- Lu, M. & Hedin, L. O. Global plant-symbiont organization and emergence of biogeochemical cycles resolved by evolution-based trait modelling. *Nat. Ecol. Evol.* **3**, 239–250 (2019).
- Fondi, M. & Di Patti, F. A synthetic ecosystem for the multi-level modelling of heterotroph-phototroph metabolic interactions. *Ecol. Model.* **399**, 13–22 (2019).
- Opatovsky, I. et al. Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment. *BMC Genomics* **19**, 402 (2018).
- Fabbiano, S., Suárez-Zamorano, N. & Trajkovski, M. Host-microbiota mutualism in metabolic diseases. *Front Endocrinol.* **8**, 267 (2017).
- Klein, C. C. et al. Exploration of the core metabolism of symbiotic bacteria. *BMC Genomics* **13**, 438 (2012).
- Bertrand, E. M. et al. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl Acad. Sci. USA* **112**, 9938–9943 (2015).
- Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–684 (2012).
- Ankrah, N. Y. D., Chouaia, B. & Douglas, A. E. The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *mBio* **9**, 141 (2018).
- Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723–735 (2008).
- Moya, A., Peretó, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet* **9**, 218–229 (2008).
- Mus, F. et al. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl Environ. Microbiol.* **82**, 3698–3710 (2016).
- Udvardi, M. & Poole, P. S. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev. Plant Biol.* **64**, 781–805 (2013).
- Gibson, K. E., Kobayashi, H. & Walker, G. C. Molecular determinants of a symbiotic chronic infection. *Annu Rev. Genet* **42**, 413–441 (2008).
- Checucci, A., diCenzo, G. C., Bazzicalupo, M. & Mengoni, A. Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. *Front Microbiol* **8**, 2207 (2017).
- Vasse, J., De Billy, F., Camut, S. & Truchet, G. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J. Bacteriol.* **172**, 4295–4306 (1990).
- Erismann, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).
- Randall, G. W. & Mulla, D. J. Nitrate nitrogen in surface waters as influenced by climatic conditions and agricultural practices. *J. Environ. Qual.* **30**, 337–344 (2001).
- Shcherbak, I., Millar, N. & Robertson, G. P. Global metaanalysis of the nonlinear response of soil nitrous oxide (N<sub>2</sub>O) emissions to fertilizer nitrogen. *Proc. Natl Acad. Sci. USA* **111**, 9199–9204 (2014).
- Triplett, E. W. & Sadowsky, M. J. Genetics of competition for nodulation of legumes. *Annu Rev. Microbiol.* **46**, 399–422 (1992).
- Streeter, J. G. Failure of inoculant rhizobia to overcome the dominance of indigenous strains for nodule formation. *Can. J. Microbiol.* **40**, 513–522 (1994).
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B. Ø. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **7**, 129–143 (2009).
- Haggart, C. R., Bartell, J. A., Saucerman, J. J. & Papin, J. A. Whole-genome metabolic network reconstruction and constraint-based modeling. *Method Enzymol.* **500**, 411–433 (2011).
- Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
- 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).
- Ankrah, N. Y. D., Luan, J. & Douglas, A. E. Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. *J. Bacteriol.* **199**, 17 (2017).
- Rodenburg, S. Y. A. et al. Metabolic model of the*i>Phytophthora infestans* </i>-tomato interaction reveals metabolic switches during host colonization. *mBio* **10**, 96 (2019).
- diCenzo, G. C. et al. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **7**, 12219 (2016).
- diCenzo, G. C. et al. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLOS Genet* **14**, e1007357 (2018).
- 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).
- Resendis-Antonio, O., Reed, J. L., Encarnación, S., Collado-Vides, J. & Palsson, B. Ø. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLOS Comput Biol.* **3**, 1887–1895 (2007).
- Resendis-Antonio, O., Hernández, M., Mora, Y. & Encarnación, S. Functional modules, structural topology, and optimal activity in metabolic networks. *PLOS Comput Biol.* **8**, e1002720 (2012).
- 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).
- 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**, 607–620 (2017).
- Pfau, T. et al. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* **8**, 12504 (2018).
- Moreira, T. B. et al. A genome-scale metabolic model of soybean (*Glycine max*) highlights metabolic fluxes in seedlings. *Plant Physiol.* **180**, 1912–1929 (2019).
- Young, N. D. et al. The *Medicago genome* provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520–524 (2011).
- Tang, H. et al. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* **15**, 312 (2014).
- Pecric, Y. et al. Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants* **4**, 1017–1025 (2018).
- Galibert, F. et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668–672 (2001).
- Biondi, E. G. et al. Metabolic capacity of *Sinorhizobium (Ensifer) meliloti* strains as determined by Phenotype MicroArray analysis. *Appl Environ. Microbiol.* **75**, 5396–5404 (2009).
- diCenzo, G. C., MacLean, A. M., Milunovic, B., Golding, G. B. & Finan, T. M. Examination of prokaryotic multipartite genome evolution through experimental genome reduction. *PLOS Genet* **10**, e1004742 (2014).
- diCenzo, G. C., Mengoni, A. & Fondi, M. Tn-Core: a toolbox for integrating Tn-seq gene essentiality data and constraint-based metabolic modelling. *ACS Synth. Biol.* **8**, 158–169 (2019).
- Lieven, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
- 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).
- Lötscher, M., Klumpp, K. & Phytologist, H. S. N. Growth and maintenance respiration for individual plants in hierarchically structured canopies of *Medicago sativa* and *Helianthus annuus*: the contribution of current and old assimilates. *N. Phytol.* **164**, 305–316 (2004).
- Zamani, M., diCenzo, G. C., Milunovic, B. & Finan, T. M. A putative 3-hydroxyisobutyryl-CoA hydrolase is required for efficient symbiotic nitrogen fixation in *Sinorhizobium meliloti* and *Sinorhizobium fredii* NGR234. *Environ. Microbiol.* **19**, 218–236 (2017).
- Minchin, F. R. & Witty, J. F. Respiratory/carbon costs of symbiotic nitrogen fixation in legumes. In *Plant Respiration: From Cell to Ecosystem* (eds Lambers,

- H. & Ribas-Carbo, M.). 195–205 (Springer Netherlands, 2005). [https://doi.org/10.1007/1-4020-3589-6\\_11](https://doi.org/10.1007/1-4020-3589-6_11).
54. Dixon, R., Kennedy, C., Kondorosi, A., Krishnapillai, V. & Merrick, M. Complementation analysis of *Klebsiella pneumoniae* mutants defective in nitrogen fixation. *Mol. Gen. Genet* **157**, 189–198 (1977).
  55. Engelke, T., Jording, D., Kapp, D. & Puhler, A. Identification and sequence analysis of the *Rhizobium meliloti* *dctA* gene encoding the C<sub>4</sub>-dicarboxylate carrier. *J. Bacteriol.* **171**, 5551–5560 (1989).
  56. de las Nieves Peltzer, M. et al. Auxotrophy accounts for nodulation defect of most *Sinorhizobium meliloti* mutants in the branched-chain amino acid biosynthesis pathway. *Mol. Plant-Microbe Interact.* **21**, 1232–1241 (2008).
  57. Rastogi, V. K. & Watson, R. J. Aspartate aminotransferase activity is required for aspartate catabolism and symbiotic nitrogen fixation in *Rhizobium meliloti*. *J. Bacteriol.* **173**, 2879–2887 (1991).
  58. Finan, T. M., McWhinnie, E., Driscoll, B. T. & Watson, R. J. Complex symbiotic phenotypes result from gluconeogenic mutations in *Sinorhizobium meliloti*. *Mol. Plant-Microbe Interact.* **4**, 386–392 (1991).
  59. Taga, M. E. & Walker, G. C. *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. *Mol. Plant-Microbe Interact.* **23**, 1643–1654 (2010).
  60. Dunn, M. F., Araiza, G. & Finan, T. M. Cloning and characterization of the pyruvate carboxylase from *Sinorhizobium meliloti* Rm1021. *Arch. Microbiol.* **176**, 355–363 (2014).
  61. De Bruijn, F. J. et al. *Rhizobium meliloti* 1021 has three differentially regulated loci involved in glutamine biosynthesis, none of which is essential for symbiotic nitrogen fixation. *J. Bacteriol.* **171**, 1673–1682 (1989).
  62. Gordon, A., Minchin, F., James, C. & Komina, O. Sucrose synthase in legume nodules is essential for nitrogen fixation. *Plant Physiol.* **120**, 867–878 (1999).
  63. Schulze, J. et al. Inhibition of alfalfa root nodule phosphoenolpyruvate carboxylase through an antisense strategy impacts nitrogen fixation and plant growth. *Phytochemistry* **49**, 341–346 (1998).
  64. Hakoyama, T. et al. Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* **462**, 514–517 (2009).
  65. Kerppola, T. K. & Kahn, M. L. Symbiotic phenotypes of auxotrophic mutants of *Rhizobium meliloti* 104A14. *J. Gen. Microbiol.* **134**, 913–919 (1988).
  66. Driscoll, B. & Finan, T. M. NAD<sup>+</sup>-dependent malic enzyme of *Rhizobium meliloti* is required for symbiotic nitrogen fixation. *Mol. Microbiol.* **7**, 865–873 (1993).
  67. Driscoll, B. & Finan, T. M. NADP<sup>+</sup>-dependent malic enzyme of *Rhizobium meliloti*. *J. Bacteriol.* **178**, 2224–2231 (1996).
  68. Capela, D., Filipe, C., Bobik, C., Batut, J. & Bruand, C. *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. *Mol. Plant Microbe Interact.* **19**, 363–372 (2006).
  69. Barnett, M. J., Toman, C. J., Fisher, R. F. & Long, S. R. A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. *Proc. Natl Acad. Sci. USA* **101**, 16636–16641 (2004).
  70. Djordjevic, M. A. *Sinorhizobium meliloti* metabolism in the root nodule: A proteomic perspective. *Proteomics* **4**, 1859–1872 (2004).
  71. Liu, A., Contador, C. A., Fan, K. & Lam, H.-M. Interaction and regulation of carbon, nitrogen, and phosphorus metabolisms in root nodules of legumes. *Front Plant Sci.* **9**, 1860 (2018).
  72. Ronson, C. W., Lyttleton, P. & Robertson, J. G. C<sub>4</sub>-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc. Natl Acad. Sci. USA* **78**, 4284–4288 (1981).
  73. Finan, T. M., Wood, J. M. & Jordan, D. C. Symbiotic properties of C<sub>4</sub>-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J. Bacteriol.* **154**, 1403–1413 (1983).
  74. 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).
  75. Li, Y., Parsons, R., Day, D. A. & Bergersen, F. J. Reassessment of major products of N<sub>2</sub> fixation by bacteroids from soybean root nodules. *Microbiology* **148**, 1959–1966 (2002).
  76. Kumar, S., Bourdès, A. & Poole, P. De novo alanine synthesis by bacteroids of *Mesorhizobium loti* is not required for nitrogen transfer in the determinate nodules of *Lotus corniculatus*. *J. Bacteriol.* **187**, 5493–5495 (2005).
  77. Waters, J. K. et al. Alanine, not ammonia, is excreted from N<sub>2</sub>-fixing soybean nodule bacteroids. *Proc. Natl Acad. Sci. USA* **95**, 12038–12042 (1998).
  78. Allaway, D. et al. Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol. Microbiol.* **36**, 508–515 (2000).
  79. Dunn, M. F. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Crit. Rev. Microbiol.* **41**, 411–451 (2015).
  80. Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y.-H. & Gresshoff, P. M. Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* **108**, 789–795 (2011).
  81. Szafran, M. M. & Haaker, H. Properties of the peribacteroid membrane ATPase of pea root nodules and its effect on the nitrogenase activity. *Plant Physiol.* **108**, 1227–1232 (1995).
  82. Pierre, O., Engler, G., Hopkins, J. & Plant, F. B. Peribacteroid space acidification: a marker of mature bacteroid functioning in *Medicago truncatula* nodules. *Plant Cell Environ.* **36**, 2059–2070 (2013).
  83. Fedorova, E. et al. Localization of H<sup>+</sup>-ATPases in soybean root nodules. *Planta* **209**, 25–32 (1999).
  84. King, B. J. & Layzell, D. B. Effect of increases in oxygen concentration during the argon-induced decline in nitrogenase activity in root nodules of soybean. *Plant Physiol.* **96**, 376–381 (1991).
  85. Rawsthorne, S. & Larue, T. A. Preparation and properties of mitochondria from cowpea nodules. *Plant Physiol.* **81**, 1092–1096 (1986).
  86. Millar, A. H., Day, D. A. & Bergersen, F. J. Microaerobic respiration and oxidative phosphorylation by soybean nodule mitochondria: implications for nitrogen fixation. *Plant Cell Environ.* **18**, 715–726 (1995).
  87. Preisig, O., Zufferey, R., Thöny-Meyer, L., Appleby, C. A. & Hennecke, H. A high-affinity *cbz*-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J. Bacteriol.* **178**, 1532–1538 (1996).
  88. Oresnik, I. J. & Layzell, D. B. Composition and distribution of adenylates in soybean (*Glycine max* L.) nodule tissue. *Plant Physiol.* **104**, 217–225 (1994).
  89. Geddes, B. A. & Oresnik, I. J. The mechanism of symbiotic nitrogen fixation. in *The Mechanistic Benefits of Microbial Symbionts* (ed. Hurst, C. J.) 69–97 (Springer, Cham, 2016). [https://doi.org/10.1007/978-3-319-28068-4\\_4](https://doi.org/10.1007/978-3-319-28068-4_4).
  90. Oresnik, I. J. A biochemical basis for oxygen limitation in legume nodules. (Queen's University, Kingston, Ontario, Canada, 1995). PhD Thesis.
  91. McRae, D. G., Miller, R. W., Berndt, W. B. & Joy, K. Transport of C<sub>4</sub>-dicarboxylates and amino acids by *Rhizobium meliloti* bacteroids. *Mol. Plant Microbe Interact.* **2**, 273–278 (1989).
  92. Wolkenhauer, O. Why model? *Front Physiol.* **5**, 21 (2014).
  93. Bosi, E., Bacci, G., Mengoni, A. & Fondi, M. Perspectives and challenges in microbial communities metabolic modeling. *Front Genet* **8**, 88 (2017).
  94. Bauer, E. & Thiele, I. From network analysis to functional metabolic modeling of the human gut microbiota. *mSystems* **3**, 359 (2018).
  95. Galardini, M. et al. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. *BMC Genomics* **12**, 235 (2011).
  96. diCenzo, G. C., Zamani, M., Cowie, A. & Finan, T. M. Proline auxotrophy in *Sinorhizobium meliloti* results in a plant-specific symbiotic phenotype. *Microbiology* **161**, 2341–2351 (2015).
  97. Wippel, K. & Long, S. R. Symbiotic performance of *Sinorhizobium meliloti* lacking ppGpp depends on the *Medicago* host species. *Mol. Plant Microbe Interact.* **32**, 717–728 (2019).
  98. diCenzo, G. C., Zamani, M., Ludwig, H. N. & Finan, T. M. Heterologous complementation reveals a specialized activity for BacA in the *Medicago-Sinorhizobium meliloti* symbiosis. *Mol. Plant Microbe Interact.* **30**, 312–324 (2017).
  99. Rogers, C. & Oldroyd, G. E. D. Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *J. Exp. Bot.* **65**, 1939–1946 (2014).
  100. Denison, R. F. & Kiers, E. T. Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiol Lett.* **237**, 187–193 (2004).
  101. Denison, R. F. & Kiers, E. T. Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr. Biol.* **21**, R775–R785 (2011).
  102. Checucci, A. et al. Mixed nodule infection in *Sinorhizobium meliloti-Medicago sativa* symbiosis suggest the presence of cheating behavior. *Front Plant Sci.* **7**, 835 (2016).
  103. Jensen, J. B., Peters, N. K. & Bhuvaneshwari, T. V. Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti*. *J. Bacteriol.* **184**, 2978–2986 (2002).
  104. Poysti, N. J. & Oresnik, I. J. Characterization of *Sinorhizobium meliloti* triose phosphate isomerase genes. *J. Bacteriol.* **189**, 3445–3451 (2007).
  105. Udvardi, M. K. & Day, D. A. Metabolite transport across symbiotic membranes of legume nodules. *Annu Rev. Plant Physiol. Plant Mol. Biol.* **48**, 493–523 (1997).
  106. Coba de la Peña, T., Fedorova, E., Pueyo, J. J. & Lucas, M. M. The symbiosome: legume and rhizobia co-evolution toward a nitrogen-fixing organelle? *Front Plant Sci.* **8**, 2229 (2017).
  107. Kuzma, M. M. et al. The site of oxygen limitation in soybean nodules. *Plant Physiol.* **119**, 399–408 (1999).
  108. Thumfort, P. P., Atkins, C. A. & Layzell, D. B. A re-evaluation of the role of the infected cell in the control of O<sub>2</sub> diffusion in legume Nodules. *Plant Physiol.* **105**, 1321–1333 (1994).
  109. Bergersen, F. J. Distribution of O<sub>2</sub> within infected cells of soybean root nodules: a new simulation. *Protoplasma* **183**, 49–61 (1994).
  110. Keating, S. M., Bornstein, B. J., Finney, A. & Hucka, M. SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics* **22**, 1275–1277 (2006).

111. Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. LibSBML: an API library for SBML. *Bioinformatics* **24**, 880–881 (2008).
112. Schellenberger, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).
113. Jensen, P. A., Lutz, K. A. & Papin, J. A. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst. Biol.* **5**, 147 (2011).
114. Vlassis, N., Pacheco, M. P. & Sauter, T. Fast reconstruction of compact context-specific metabolic network models. *PLOS Comput Biol.* **10**, e1003424 (2014).

### Acknowledgements

GCD was supported by a postdoctoral fellowship from the Natural Science and Engineering Research Council of Canada, and funding from Queen's University. AM was supported by a grant from Fondazione Cassa di Risparmio di Firenze (project name: Metatrack). This work was partially supported by a grant from the Italian Ministry of Agriculture (MICRO4Legumes, act n. n.89267, 19-12-2019).

### Author contributions

G.D., A.M., and M.F. conceived the study. G.D., T.P., M.F. prepared the metabolic models. G.D., M.T., M.F. ran simulations with the model. G.D., A.M., and M.F. interpreted the results. G.D., A.M., and M.F. wrote the manuscript. All authors critically revised the manuscript and approved the final version.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-16484-2>.

**Correspondence** and requests for materials should be addressed to A.M. or M.F.

**Peer review information** *Nature Communications* thanks Jörg Schwender and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020