Molecular BioSystems

PAPER

View Article Online
View Journal | View Issue



Cite this: *Mol. BioSyst.*, 2017, 13, 607

# Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states†

Yi Yang, Xiao-Pan Hu and Bin-Guang Ma*

Received 28th July 2016,
Accepted 27th January 2017

DOI: 10.1039/c6mb00553e

rsc.li/molecular-biosystems

*Bradyrhizobium diazoefficiens* is a rhizobium able to convert atmospheric nitrogen into ammonium by establishing mutualistic symbiosis with soybean. It has been recognized as an important parent strain for microbial agents and is widely applied in agricultural and environmental fields. In order to study the metabolic properties of symbiotic nitrogen fixation and the differences between a free-living cell and a symbiotic bacteroid, a genome-scale metabolic network of *B. diazoefficiens* USDA110 was constructed and analyzed. The metabolic network, *i*YY1101, contains 1031 reactions, 661 metabolites, and 1101 genes in total. Metabolic models reflecting free-living and symbiotic states were determined by defining the corresponding objective functions and substrate input sets, and were further constrained by high-throughput transcriptomic and proteomic data. Constraint-based flux analysis was used to compare the metabolic capacities and the effects on the metabolic targets of genes and reactions between the two physiological states. The results showed that a free-living rhizobium possesses a steady state flux distribution for sustaining a complex supply of biomass precursors while a symbiotic bacteroid maintains a relatively condensed one adapted to nitrogen-fixation. Our metabolic models may serve as a promising platform for better understanding the symbiotic nitrogen fixation of this species.



## Introduction

Rhizobia usually refer to a group of Gram-negative bacteria that can establish symbiotic relationship with legumes and reduce atmospheric nitrogen into ammonium to provide nitrogen nutrition for the host plants. Some rhizobia can perform symbiotic nitrogen fixation (SNF) in the form of bacteroids. During the SNF process, ammonium produced by bacteroids is secreted into the host plant in exchange for a carbon and energy source,[1] which maintains a stable symbiotic relationship. The legume–rhizobium system has the highest efficiency and largest share[2] among all kinds of biological nitrogen fixation systems and thus is an important model in the research field of mutualism[3] that has been studied for a long time with practical application in improvements for agriculture and the environment.

Induction, establishment and maintenance of the symbiotic relationship between rhizobia and legumes are associated with complex events including signal transduction, response, gene expression and regulation. At the metabolic level, this relationship is manifested as the adjustment and combination of the metabolic networks on both sides, eventually forming a coordinated and stabilized micro-environment of nodule cells. Development of high-throughput technologies has provided large-scale and systematic raw materials for symbiotic nitrogen

*Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China. E-mail: mbg@mail.hzau.edu.cn*

† Electronic supplementary information (ESI) available: Table S1. Basic properties of *i*YY1101. Table S2. Essential genes predicted in the FL state. Table S3. Essential genes predicted in the SNF state. Table S4. Essential genes associated with symbiotic nitrogen fixation detected by simulated single-gene-deletion. Table S5. Essential gene pairs predicted in the FL state. Table S6. Essential gene pairs predicted in the SNF state. Fig. S1. Phenotype phase plane for three kinds of N outputs, $NH_3$, L-alanine and L-aspartate, and the trade-offs between them. Fig. S2. The proportions of gene pairs with different lethalities in different metabolic models. Fig. S3. Simulated double-gene-knockout to the FL and SNF models. Fig. S4. Simulated single-reaction-knockout to the FL and SNF models. Supplementary method. A method for determining the biomass objective function for *B. diazoefficiens* USDA110. Supplementary file S1. Important information during the reconstruction of the metabolic network of *B. diazoefficiens* USDA110, including reaction sources, genome annotation, details of reactions and metabolites and the gaps unable to be filled. Supplementary file S2. Detailed results of the phenotypic microarray experiment and the corresponding *in silico* predictions. Supplementary file S3. The reconstructed genome-scale metabolic network model of *B. diazoefficiens* USDA110, *i*YY1101. Supplementary file S4. The constraint-based metabolic network model of *B. diazoefficiens* USDA110 specific for the FL state. Supplementary file S5. The constraint-based metabolic network model of *B. diazoefficiens* USDA110 specific for the SNF state. Supplementary file S6. Detailed results of simulation including essentiality analysis for genes and reactions and metabolic control analysis. See DOI: 10.1039/c6mb00553e

fixation research.[4–6] Integration of high-throughput data, physiological experiments and *in silico* simulation by the systems biology approach is helpful in gaining new insights into SNF from an overall perspective, and this is exactly what constraint-based metabolic modeling is competent for. Many metabolic networks have been manually constructed and widely used in fields such as improving gene annotation,[7] predicting essentiality of genes[8] and assisting in metabolic engineering.[9,10] As for constraint-based metabolic reconstruction for symbiotic nitrogen fixation, three genome-scale models, *i*OR363 (*Rhizobium etli* CFN42),[11] *i*OR450 (*Rhizobium etli* CFN42)[12] and *i*HZ565 (*Sinorhizobium meliloti* 1021),[13] have been published.

Besides *Rhizobium* and *Sinorhizobium*, *Bradyrhizobium* is another important group of rhizobia.[14] Since being discovered, the *Glycine max* (soybean) – *B. diazoefficiens* symbiotic system has always been an important model for SNF research.[15,16] *B. diazoefficiens* USDA110 (formerly known as *B. japonicum* USDA110),[17] as the type strain of *Bradyrhizobium* with efficient SNF ability,[18–20] is an important parent strain for agricultural microbial agents.[21,22] Documentation of the metabolic network of *B. diazoefficiens* USDA110 is greatly needed. After its whole genome was published in 2002,[23] the corresponding transcriptome and proteome data sets were also established in recent years;[24–27] by accumulating knowledge from physiological and biochemical research, a solid foundation has been laid for the construction and analysis of the genome-scale metabolic network of *B. diazoefficiens* USDA110.

All of the three published genome-scale metabolic models of rhizobia were reconstructed to simulate the bacteroid whose physiological function was highly specialized and could nearly be considered as an organelle dedicated to nitrogen fixation.[28] However, the physiological behaviors and features of free-living (FL) rhizobia are different from those of bacteroids due to a more complicated habitat in terms of a wider range of available nutrient resources,[1] and the differences were intriguing to be investigated from the perspective of metabolism. Moreover, for *B. diazoefficiens*, there are many puzzles related to the metabolism of its bacteroid; for example, the nitrogen output form of the bacteroids has been debated for two decades.[29–34] A reasonably curated metabolic model of *B. diazoefficiens* can provide a biochemically structured platform to analyze these puzzles. In this work, the genome-scale metabolic network for *B. diazoefficiens* USDA110 was reconstructed by the combination of genome annotation, biochemical data mining and constraint-based metabolic simulation, and refined by high-throughput experimental phenotypic data and gene expression data. Based on the metabolic model, the differences between FL cells and SNF bacteroids were compared and the form of nitrogen output flux in the bacteroids was predicted and analysed.

# Materials and methods

### Strains and growth conditions

*B. diazoefficiens* USDA110 strain (ACCC15034) was obtained from the Agricultural Culture Collection of China (ACCC). The strain was grown aerobically at 28 °C[17] in an optimized CS7 medium:[35] 15 g of glucose, 0.5 g of $KNO_3$, 0.36 g of $KH_2PO_4$, 1.4 g of $K_2HPO_4$, 0.25 g of $MgSO_4{\cdot}7H_2O$, 0.02 g of $CaCl_2{\cdot}2H_2O$, 0.2 g of NaCl, 6.6 mg of $FeCl_3$, 15 mg of EDTA, 0.16 mg of $ZnSO_4{\cdot}7H_2O$, 0.2 mg of $Na_2MoO_4$, 0.25 mg of $H_3BO_3$, 0.2 mg of $MnSO_4{\cdot}4H_2O$, 0.02 mg of $CuSO_4{\cdot}5H_2O$, 1 μg of $CoCl_2{\cdot}6H_2O$, 1 mg of thiamine-HCl, 1 mg of nicotinamide, 2 mg of calcium pantothenate, 1 μg of biotin, 1 L of distilled water, and pH 7.0.

### Phenotypic microarray

Phenotypic microarray (PM) analysis was performed by using a service from Biolog.[36] In total, 190 carbon, 95 nitrogen, 59 phosphor and 35 sulfur nutrient sources were tested. The experimental process was implemented as the PM Procedure for Gram-negative Bacteria. The strain was grown at 28 °C on an SM agar plate. A single colony was cultured to an optical density at 600 nm ($OD_{600}$) of 0.5 in CS7 medium, harvested by centrifugation and resuspended to an $OD_{600}$ of 0.2 in CS7 medium lacking the nutrient source to be tested (*e.g.* for carbon source, lacking glucose). Each well of the four kinds of Biolog GN2 microplates (PM 1–4) was inoculated with 50 μL of bacterial suspension and 150 μL of CS7 medium (also without the nutrient source to be tested). Then these plates were incubated at 28 °C for 168 h in a microplate reader and tested once per hour. Each nutrient substrate was tested in duplicate. The data were analyzed using an OmniLog PM package according to the Biolog protocols.

### Metabolic network reconstruction

We adopted the published protocols[37] for the genome-scale metabolic network reconstruction of *B. diazoefficiens* USDA110. It consisted of five major steps: draft reconstruction, draft refinement, objective function (OF) determination, conversion of draft into a computable model and model refinement. The reconstruction process is schematically shown in Fig. 1.

The information about gene-enzyme-reaction-pathway relationships for *B. diazoefficiens* USDA110 was searched in related databases such as KEGG and UniprotKB, and the corresponding reactions made up a reaction set (the 1st dataset). The automated metabolic model of *B. diazoefficiens* USDA110 was reconstructed using Model SEED RAST (the 2nd dataset). The amino acid sequences of the enzyme set of models *i*OR450 and *i*HZ565 were used as the reference, and putative orthologs were selected by using BLASTP in *B. diazoefficiens* USDA110 with the criteria for identifying candidates as follows: identity $> 30\%$, *E*-value $< 1 \times 10^{-12}$ and matched length $> 60\%$; then, the corresponding reactions associated with these candidate enzymes were identified (the 3rd dataset). Then these three datasets were combined and redundancy was removed, resulting in the draft metabolic network. The sources of reactions are curated in Supplementary file S1, ESI.†

Then, the draft was refined by integrating the information from the related databases, predicting tools and the relevant literature. Databases such as BRENDA, BiGG, MetaCyc and MetaNetX were used to unify the information about the metabolites, enzymes and reactions filtered from different data sources.

 

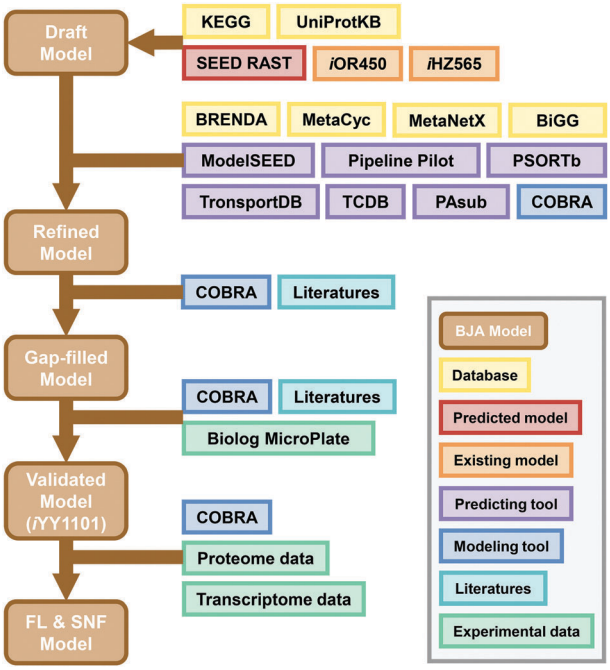


Fig. 1 Schematic representation of the procedure used to reconstruct the genome-scale metabolic model of *B. diazoefficiens* USDA110. BJA Model, the metabolic model of *B. diazoefficiens* USDA110. The color scheme is annotated on the lower-right corner in the figure.

Cofactors participating in reactions were determined by searching databases such as KEGG, UniprotKB and BRENDA. Subcellular localization of the reactions, namely, enzyme compartment was predicted by PSortb and PASub. Gene-protein-reaction (GPR) associations were determined by searching KEGG and UniprotKB. The Model SEED was used to determine reaction thermodynamics in a related *in vivo* physiological state, and then the reaction directionalities were inferred. The ionization state of each metabolite was calculated by Pipeline Pilot in the related *in vivo* physiological state, and then the chemical equations were balanced based on the charged formula of metabolites. Transport reactions were determined using the information in KEGG, UniprotKB, TransportDB and TCDB. For the reactions whose information is different in different data sources and hard to be unified and integrated, they were determined by integrating knowledge from the literature and other metabolic models.

Dozens of gaps existed in the refined model, which should be filled iteratively. Constraint-based Reconstruction and Analysis (COBRA) Toolbox 2.0[38] was used to detect the dead-end metabolites and related gaps. And the gap-filling candidate reactions were identified using an optimization-based algorithm (growthExpMatch) integrated in the COBRA Toolbox. Then the candidate reactions were verified by integrating knowledge from the literature and other metabolic models. It is hard to fill all of the gaps in metabolic networks; if the gaps have no effect on the flux distribution reflecting specific physiological conditions and no evidence exists to fill them, they were retained in the model. The 84 reactions added through gap-filling and the 6 dead-end metabolites are presented in Supplementary file S1, ESI.†

Biolog phenotypic microarray experiments were used to test the ability of *B. diazoefficiens* USDA110 to use nutrients. And the *in silico* optimal growths were simulated using the COBRA Toolbox through FBA under the corresponding substrate input. Then the contradictions between these two kinds of results were used to further refine the gap-filled model. Other experimental results of nutrient source utilization for *B. diazoefficiens*[17,39–41] were used as reference to resolve these contradictions. As for the nutrient source utilization which was negative in simulation and positive in PM and other experimental results, related transport, exchange and metabolic reactions were added; thus the gap-filled model was refined and validated. The reactions validated through the PM experiment are presented in Supplementary file S1, ESI.†

An integrated proteomics and transcriptomics reference dataset[24,27] which reflected the wild type free-living rhizobium and the symbiotic bacteroid was used to impose constraints associated with enzyme expression levels on the model so as to reduce the solution space of the flux distribution of the validated model and make it closer to the real physiological state. The original datasets were filtered according to the standards in the related literature.[24,27] Then the omics data were mapped onto the validated model using a linear programming algorithm (GIMME[42]) integrated in the COBRA Toolbox. Condition-specific models (FL and SNF models) were reconstructed and used in the subsequent analysis.

### Flux balance analysis

Flux balance analysis (FBA) was used to optimize specific objective functions (OFs) and the flux distributions of a metabolic network under specified conditions.[43] The whole metabolic system is represented by a stoichiometric matrix $\boldsymbol{S}$ in which $\boldsymbol{S}_{ij}$ represents the stoichiometric coefficient of metabolite $i$ in reaction $j$. Flux throughout all the reactions is represented by a vector $\boldsymbol{v}$. Specific OFs ($\boldsymbol{f}$) such as biomass or ATP consumption are specified to represent the objectives of the related physiological states. OFs were optimized (max/min $\boldsymbol{f}^T\boldsymbol{v}$) subject to the steady-state constraint ($\boldsymbol{S}\cdot\boldsymbol{v} = 0$), capacity constraints ($\boldsymbol{v}_{lb} \leq \boldsymbol{v} \leq \boldsymbol{v}_{ub}$) and thermodynamic constraints (directions of fluxes), and flux distributions were determined.

FBA was implemented using the COBRA Toolbox 2.0 in the MATLAB environment. The GNU linear programming kit (GLPK) solver was employed to optimize OFs in FBA except for gap-filling where the Gurobi 6.5.0 solver was used instead to speed up calculation. SBMLToolbox and libSBML were used to process SBML files. Procedures involving FBA include gap-filling, simulation of the nutrient source utilization in validating model reconstruction and essentiality analysis for reactions and genes.

### Flux variability analysis

By means of FBA, a specific optimal flux distribution supporting the associated metabolic objective can be obtained. However, for the *in vivo* metabolic system, flux distribution reflecting a specific physiological state is a space rather than a single point. In this research, flux variability analysis (FVA) was used to calculate

the flux distribution space in comparisons between different metabolic models.

FVA is a bi-level optimization procedure derived from FBA.[44] As in FBA, OF was optimized. Then maximum and minimum values of each reaction flux were optimized (max/min $v_i$) subject to related constraints ($\boldsymbol{S}\cdot\boldsymbol{v} = 0$, $\boldsymbol{v}_{lb} \leq \boldsymbol{v} \leq \boldsymbol{v}_{ub}$ and $\boldsymbol{f}^T\boldsymbol{v} = Z_{obj}/Z_{obj}'$, where $Z_{obj}/Z_{obj}'$ denotes the optimal or suboptimal value of an OF). $\boldsymbol{v}_{min} \leq \boldsymbol{v} \leq \boldsymbol{v}_{max}$ is the flux distribution space.

Like FBA, FVA was performed using the COBRA Toolbox and the LP problems were solved using GLPK and Gurobi. Procedures involving FVA include: (i) simulation and comparison of the four different product output forms in defining OF of the SNF state; (ii) simulation and comparison of the SNF state by using *i*OR450, *i*HZ565 and *i*YY1101 SNF models; (iii) simulation and comparison of FL and SNF states by using *i*YY1101 FL and SNF models.

### Simulated gene/reaction knockout

Simulated gene/reaction knockout was used to investigate the essentiality of a given gene/reaction in a particular physiological state. It was conducted with the nutrient input of the CS7 medium (without metal ions except Fe) for the free-living state and with an uptake rate of 5 mmol $gDW^{-1}$ $h^{-1}$ L-malate and succinate for the symbiotic state. First, FVA was used to calculate the flux spaces of the FL/SNF models and the range of potential flux values of each reaction was determined. Then, each of the reactions/genes with non-zero-fluxes was set to zero flux to simulate gene/reaction knockout; the impact of each reaction/gene deletion on the OF was simulated by FBA. The ratio of the OF optimal value of the mutant strain to that of the corresponding wild-type strain ($O_{mt}/O_{wt}$) was used as a measure of reaction/gene essentiality. And reactions/genes were divided into three sets: essential, deletion of which blocks OF flux ($O_{mt}/O_{wt} < 0.01$), semi-essential, deletion of which affects but does not block OF flux ($0.01 < O_{mt}/O_{wt} < 0.99$), and non-essential, deletion of which has little effect on OF flux ($0.99 < O_{mt}/O_{wt}$).

Simulated gene pair knockout was conducted under the same conditions as the simulated gene/reaction knockout. Total gene pairs include all of the possible combinations of two genes except for three cases: (i) at least one corresponding gene is essential; (ii) consisting of the same gene, such as gene_A + gene_A; (iii) redundant, such as gene_A + gene_B *vs.* gene_B + gene_A. So the total number of gene pairs can be derived *via*

$$N_{GPs} = \frac{(N_G - N_{EG})^2 - (N_G - N_{EG})}{2}$$

where $N_{GPs}$, $N_G$ and $N_{EG}$ represent the total number of gene pairs, the total number of genes, and the number of essential genes, respectively. The classification criteria for the essentiality of gene pairs were the same as those in the simulated gene/reaction knockout.

### Metabolic control analysis

Metabolic Control Analysis (MCA)[45] was used to analyze the sensitivities of specific fluxes to the changes of other fluxes. The Flux Control Coefficient (FCC) was used to quantify the sensitivity. It is defined as

$$C_{v_i}^{J} = \frac{\partial J}{\partial v_i}\frac{v_i}{J} = \frac{\partial \ln J}{\partial \ln v_i}$$

where $J$ is the specific flux influenced by others and $v_i$ is the flux influencing the specific one. The COBRA Toolbox was used to calculate the FCC of OF influenced by the perturbations of other fluxes in the SNF model. First, the flux space of the SNF model was calculated by FVA. The fluxes were divided into four types: zero, for which $\boldsymbol{v}_{max} = \boldsymbol{v}_{min} = 0$; infinite, for which $\boldsymbol{v}_{max} = \boldsymbol{v}_{ub}$ or $\boldsymbol{v}_{min} = \boldsymbol{v}_{lb}$; changeable, for which $2(\boldsymbol{v}_{max} - \boldsymbol{v}_{min})/(|\boldsymbol{v}_{max}| + |\boldsymbol{v}_{min}|) > 0.001$; fixed, for which $2(\boldsymbol{v}_{max} - \boldsymbol{v}_{min})/(|\boldsymbol{v}_{max}| + |\boldsymbol{v}_{min}|) < 0.001$. The changes of the OF flux influenced by small changes (0.001 was used) of the fluxes of other reactions was calculated by FBA. Then, the FCC of OF influenced by each reaction in the infinite, changeable and fixed types was calculated according to the definition.

## Results

### Metabolic network construction

In our model, macromolecular components in biomass such as DNA, RNA and protein were replaced by their small molecular precursors, nucleotides and amino acids, and thus the corresponding reactions involving replication, repair, transcription, translation, folding, sorting and degradation should not be included in the draft. There are a series of metabolic pathways in KEGG such as xenobiotics biodegradation and human diseases which are not related to normal physiological conditions of this bacterium. Therefore, the reactions in these metabolic pathways should be deleted unless they are also present in other metabolic pathways. The reactions selected from KEGG and UniprotKB databases were combined with the reactions determined by homologous alignment to the *i*OR450 and *i*HZ565 model, as well as the reactions in the predicted model generated using the Model SEED RAST, resulting in the draft model.

Metabolic network construction is an iterative process (Fig. 1), and for each step in the reconstruction, with the addition of new information, the accuracy and rationality of the constructed network continue to increase. During the unification and integration of the original datasets, mutual validation was performed at the same time and the data were filtered. The standard Gibbs free energy changes of reactions included in the draft model were estimated, and thus the reversibility of reactions that are related to the directions of pathways and the connectivity among them could be inferred. Chemical equations were balanced based on the precisely calculated ionization states of metabolites, which ensures the accuracy of the stoichiometry matrix. Cofactors of the reactions were determined, which effectively reduced the redundancy of the reaction set and correspondingly reduced the Type III-extreme pathways (internal endless loop) in the simulation result. Metabolic dead ends were identified and the corresponding gaps were filled, which ensured that the models were functional for related physiological states. In order to correctly reflect phenotypic traits, the model predictions (*e.g.*, optimal values of OF)

 

should be consistent with known physiological properties and correspondingly, processes of simulation and model refinement were also iterative, involving adjustment in reactions, pathways and network structure. In this work, the PM experiment results and the previously published proteome and transcriptome of this species[24,27] were used to refine the network structure to increase the rationality of the model.

### Model refinement by high-throughput substrate utilization experiments

Phenotypic microarray (PM) technology is a kind of high-throughput experiment that can detect the ability of a specific microbe to use a series of substrates.[46] It can be used to assess a metabolic model qualitatively by comparing PM results with model predictions. The utilization for each substrate tested in the PM experiment was simulated by FBA. Because the PM experiment can only be implemented on free-living rhizobia, each simulation was performed under the conditions of FL and the C/N/P/S source was replaced with the nutrient source to be tested. If the simulation and experimental results are inconsistent, changes would be made on the structure of the metabolic network by adding or removing reactions supported by the literature. After refinement, the *in silico* predictions of the *i*YY1101 model could fit the metabolic phenotypes with an overall accuracy of 80.5% (Table 1, details are presented in Supplementary file S2, ESI†). The reactions validated through the PM experiment are presented in Supplementary file S1, ESI.†

### Metabolic network model of *B. diazoefficiens* USDA110

As shown in Fig. 1, the genome-scale metabolic network of *B. diazoefficiens* USDA110 was reconstructed after refinement of the draft model, gap-filling and validation by the PM experiment. The genome-scale metabolic model, *i*YY1101, includes two compartments (cytosol and extracellular), 1031 reactions, 662 metabolites, and 1101 genes, covering 13.24% of the 8317 protein coding genes identified from the whole genome. The basic features of *i*YY1101 are shown in Table S1 (ESI†). Among all the 1031 reactions, there are 821 metabolic reactions (reactions not belonging to transport, exchange and OF) distributed in 55 KEGG metabolic pathways (Fig. 2A). The detailed metabolic network model is available in Supplementary file S3, ESI.†

Table 1 Comparisons between the *in silico* predictions and PM experimental results for different nutrient sources after the refinement of the model

| | | Agreement | | Disagreement | | |
|---|---|---|---|---|---|---|
| Source | NC | E–G C–G | E–NG C–NG | E–G C–NG | E–NG C–G | Agreement rate (%) |
| Carbon | 76 | 31 | 34 | 5 | 6 | 85.5 |
| Nitrogen | 48 | 19 | 17 | 8 | 4 | 75.0 |
| Phosphorus | 18 | 7 | 9 | 2 | 0 | 88.9 |
| Sulfur | 7 | 3 | 2 | 0 | 2 | 71.4 |
| Total | 149 | 60 | 62 | 15 | 12 | 81.9 |

E, experimental; C, computational; G, growth; NG, no growth; NC, no. of comparisons. Agreement rate = agreement/NC.

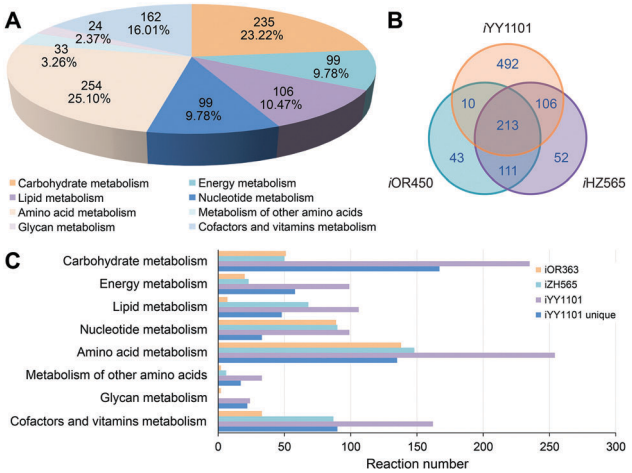


Fig. 2 Classification of the metabolic reactions of the *i*YY1101 model and a comparison to the existing metabolic network models of rhizobia. (A) Distribution of the 821 metabolic reactions of *i*YY1101 classified according to the KEGG pathway database. (B) Venn diagram of reaction overlaps shared between *B. diazoefficiens* model *i*YY1101, *R. etli* model *i*OR450 and *S. meliloti* model *i*HZ565. Only metabolic reactions were considered, *i.e.*, transport reactions, exchange reactions and OFs were excluded. (C) Distribution of metabolic reactions in the three symbiotic nitrogen fixation models, *i*YY1101, *i*OR363, *i*HZ565, and the 492 unique metabolic reactions of *i*YY1101 classified according to the KEGG pathway database.

### Comparison to the existing metabolic network models of rhizobia

The *i*YY1101 model is the first published metabolic model for *Bradyrhizobium*. Compared with the other three symbiotic nitrogen fixation metabolic network models, the scale of *i*YY1101 is significantly expanded (Fig. 2B and Table 2), especially for the transport and exchange reactions. This is because the other three studies focused only on simulating the symbiotic nitrogen fixation state, while our study aimed at simulating both of the FL and SNF states. The significant differences in transport and exchange reactions can reflect the ability of free-living rhizobia to assimilate a wide range of rhizosphere nutrient sources,[47] which is consistent with the large number of unique reactions in *i*YY1101, especially for the metabolism of carbohydrate, energy, other amino acids and glycan (Fig. 2B and C). For free-living *B. diazoefficiens*, carbohydrate metabolism reflects a broad spectrum of available carbon sources; energy metabolism is related to carbon fixation, methane metabolism and more sophisticated nitrogen and sulfur metabolism; glycan metabolism involves peptidoglycan and lipopolysaccharide biosynthesis; peptidoglycan contains D-amino acids that are classified into the metabolism of other amino acids.

### Condition-specific models (FL and SNF model)

Given the fact that *i*YY1101 was constructed based on whole-genome annotation, it covers the metabolic potential of all possible physiological states. In order to precisely characterize the specific physiological states (FL rhizobium and SNF bacteroid), enough constraints should be imposed on *i*YY1101. Therefore, the integrated proteomics and transcriptomics datasets[24,27] were used for further narrowing the flux space to reflect the

Table 2 Basic properties of the four symbiotic nitrogen fixation models

| Model | Organism | Reaction no. | Metabolite no. | Gene no. | Coverage of annotated ORF (%) | Transport reaction no. | Exchange reaction no. |
|---|---|---|---|---|---|---|---|
| *i*OR363 | *Rhizobium etli* CFN42 | 387 | 371 | 363 | 6.02 | 11 | 12 |
| *i*OR450 | *Rhizobium etli* CFN42 | 405 | 377 | 450 | 7.46 | 11 | 12 |
| *i*HZ565 | *Sinorhizobium meliloti* 1020 | 527 | 522 | 565 | 9.09 | 19 | 23 |
| *i*YY1101 | *Bradyrhizobium diazoefficiens* USDA110 | 1031 | 661 | 1101 | 13.24 | 106 | 99 |

Table 3 Basic properties of the *i*YY1101, FL and SNF models

| Model | Reaction no. | Metabolite no. | Gene no. | Pathway no. | Exchange reaction no. | OF no. |
|---|---|---|---|---|---|---|
| *i*YY1101 | 1031 | 661 | 1101 | 55 | 99 | 5 |
| FL | 683 | 499 | 779 | 51 | 32 | 1 |
| SNF | 442 | 307 | 586 | 48 | 28 | 1 |

related phenotypes. The original datasets were filtered according to the standards in the related literature, and, 5284 and 3540 genes were recognized as expressed for FL and SNF states, respectively. After integration of the gene expression information, condition-specific models were reconstructed (Table 3). The detailed condition-specific models are available in Supplementary file S4 and S5, ESI.†

Previous work[48] has quantitatively detected the nitrogen fixation rates of the bacteroid of *B. diazoefficiens* USDA110 in different oxygen inputs and found that the nitrogen fixation rate increased with increased respiration. In order to test how gene expression constraints benefit the condition-specific models, $O_2$ and L-malate were provided to the *i*YY1101 and SNF models at the experimentally determined rate, respectively; the nitrogen fixation rates were simulated by setting the nitrogen fixation reaction as the OFs in the two models, respectively. As shown in Fig. 3A, the predictions of the SNF model are more consistent with the experimental results than *i*YY1101 (the sum of absolute residuals: 0.19 *vs.* 1.08). By adding gene expression constraints, a series of reactions, which are associated with unexpressed genes (enzymes) and thus unable to carry fluxes, were filtered out, resulting in a more realistic metabolic model and a better prediction ability (Fig. 3B, $R^2$ = 0.93). Therefore, the integration of the proteomics and transcriptomics data was effective and indeed benefited the accuracy of the SNF model. Meanwhile, the high correlations between our model predictions and the published experimental data (Fig. 3) further validated the reliability of our metabolic models.

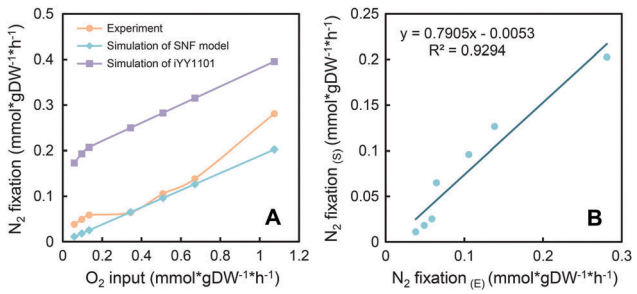


Fig. 3 Agreement between the experimental and *in silico* nitrogen fixation rates of *B. diazoefficiens* before and after the addition of gene expression constraints. (A) The nitrogen fixation rate plotted as a function of the oxygen input. The SNF model predictions show better agreement with the experiment than *i*YY1101. (B) The correlation of the nitrogen fixation rates between the experimental data and the SNF model predictions. $N_2$ fixation$_{(E)}$, experimental nitrogen fixation rate; $N_2$ fixation$_{(S)}$, predicted nitrogen fixation rate by SNF model simulation.

## Defining objective functions and substrate input sets

Metabolic network models are useful to simulate the physiological states of specific biological systems. The OF and substrate input set (SIS) are two key elements reflecting the physiological states: the OF captures the biochemical goal of the metabolic system and the SIS represents the environmental conditions. Optimizing the flux of OF imposes strong constraints that reflect the specific physiological states on metabolic networks. Simulation of the corresponding phenotype can be achieved during this process. Therefore, determining the objective function is a key step in metabolic network reconstruction.

All of the three previously published metabolic reconstructions for rhizobia simulated the phenotype of bacteroid in a symbiotic state, and thus their OFs reflected the corresponding physiological characteristics such as nitrogen fixation, poly-β-hydroxybutyrate (PHB) biosynthesis and metabolite exchange between the bacteroid and the host.

Like other free-living microorganisms, a free-living rhizobium also needs to maintain biomass synthesis and thus the biomass reaction is a suitable choice as the OF of the FL model. To determine the biomass reaction, the biomass constituents and their fractional contributions as well as the energy consumption during biomass synthesis which includes growth-associated ATP maintenance (GAM) and non-growth-associated ATP maintenance (NGAM) need to be known.[37] The chemical composition of the *B. diazoefficiens* USDA110 cell is very special: (i) PHB, which is the main energy storage substance in a free-living *B. diazoefficiens* USDA110 cell, accounts for 16% of the cell dry weight;[49] (ii) *B. diazoefficiens* USDA110 synthesizes and secretes capsular polysaccharides (CPS) and extracellular polysaccharides (EPS), which also account for 16% of the cell dry weight.[50] The details of biomass were acquired by literature mining (Supplementary method, ESI†).

The physiological properties of bacteroids are also quite distinctive: cell growth is almost stagnant; carbon and nitrogen assimilation and protein synthesis are depressed; DNA replication is extremely sluggish; synthesis and secretion of CPS and EPS are completely stopped; the nitrogen-fixing system operates efficiently; PHB is massively accumulated (accounting for 50% of the cell dry weight);[51] and the proportion of leghemoglobin

 

(which contains heme) in the total cytoplasmic protein content of the nodule cell can increase up to 20–30%.[52] Heme is synthesized in bacteroids and then transported into the cytoplasm of the nodule cell where it combines with the apoprotein and generates mature leghemoglobin.[53,54] In summary, except for PHB and heme, all other components of the biomass reaction could be ignored in the OF reflecting the SNF state.

Nitrogen output flux is the most significant material exchange for symbiotic nitrogen fixation. However, the nitrogen output form is not that clear. In several studies, ammonium was considered as the principal export product of nitrogen fixation,[29,30] while others believed that amino acids such as alanine and aspartate are secreted.[31–34] So, we set up four kinds of nitrogen outputs: $NH_3$, L-alanine, $NH_3$ + L-alanine and $NH_3$ + L-alanine + L-aspartate, and defined the corresponding OFs as $OF_N$, $OF_A$, $OF_{NA}$ and $OF_{NAA}$, respectively. Then the optimal nitrogen fixation rates were simulated by FVA under these four conditions ($SNF_N$, $SNF_A$, $SNF_{NA}$ and $SNF_{NAA}$,) and the flux distributions were curated and analyzed. The flux distributions under the four modeling conditions were quite similar: (i) a big core reaction set was shared by the four SNF models (Fig. 4A); (ii) the shared reaction set (103 reactions) covered

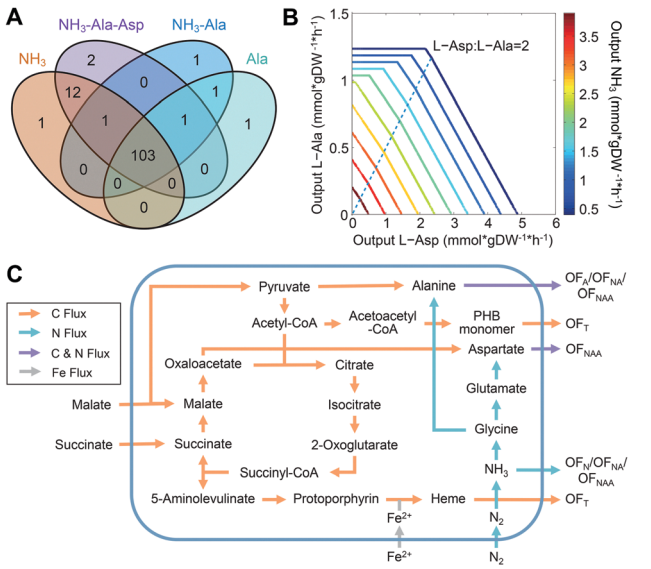


Fig. 4 Analysis of the nitrogen outputs in defining the OF of the SNF models. (A) Overlap of non-zero-fluxes among the four SNF models with different OFs ($OF_N$, $OF_A$, $OF_{NA}$ and $OF_{NAA}$). The name of the set represents the nitrogen output form. (B) Phenotype phase plane for the three kinds of nitrogen outputs, $NH_3$, L-alanine and L-aspartate, and the trade-offs between them. The phase plane is shown in 2-D. The output fluxes of L-alanine and L-aspartate were varied incrementally across the viable range. The magnitude of $NH_3$ output was expressed by the color scale. The dashed line represents that the SNF system has maximum nitrogen output when the output ratio of L-aspartate and L-alanine was 2 : 1. More details are shown in Fig. S1 (ESI†). (C) Schematic diagram of material transformation flux shared by the four SNF models. $OF_N$, $OF_A$, $OF_{NA}$ and $OF_{NAA}$ represent four different OFs with the four kinds of nitrogen output forms: $NH_3$, L-alanine, $NH_3$ + L-alanine and $NH_3$ + L-alanine + L-aspartate, respectively. $OF_T$ represents all of the four OFs. Fluxes of different elements are represented by arrows of different colors. The OF is composed of the output product of nitrogen fixation, the PHB monomer and heme: nitrogen-fixation-product [c] + Heme [c] + PHB monomer [c] → nitrogen-fixation-product [e] + Heme [e].

the core metabolism of all relevant components of the four OFs (Fig. 4C). So, although there are differences in the details, all of these four conditions are possible SNF states, just as in the previous research findings.[31] However, the differences in nitrogen fixation efficiency and utilization efficiency of the energy source can be reflected by the nitrogen output flux. Table 4 showed that the $SNF_{NAA}$ state had the largest nitrogen output flux, so we chose it to represent the SNF state. Then the quantitative relationship between $NH_3$, L-alanine and L-aspartate outputs was simulated by FBA. Fig. 4B and Fig. S1 (ESI†) revealed the full flux space for nitrogen output and the trade-offs between these three output forms. It showed that when the output ratio of L-aspartate and L-alanine was 2 : 1, $NH_3$ (as well as the total nitrogen output) showed the maximum output flux, therefore, this ratio was used in the OF of the SNF model.

Similar to the OF, the SIS is another key element that represents the nutritional conditions of a specific physiological state. Nutrient exchange between microorganisms and the environment is very complex and relevant studies are often inadequate, so it's hard to simulate the substrate input of a metabolic model precisely. Nevertheless, our study focused on the differences between the FL and SNF states, hence, as long as it can be ensured that the model can reflect the main features of the corresponding physiological state, the SIS can be simplified appropriately. *B. diazoefficiens* has evolved a large and complex genome (9.1Mb)[23] containing a wide range of catabolic systems to access a great range of nutrients present at low concentrations in soil and the plant rhizosphere.[55] Therefore, it's hard to determine a standard state describing the free-living cell. However, CS7 medium is an important kind of synthetic medium for culturing rhizobia and was used in the PM experiment of this study, so we chose substrates in CS7 medium as the SIS of the FL state. Dicarboxylic acids, particularly malate and succinate, are considered as the primary carbon sources for bacteroids,[56] so they were chosen as the carbon input of the SNF state. During simulation, we found that the flux of the OF for the SNF state can be sustained by a very simple input: $O_2$, $N_2$, $PO_4^{3+}$, $SO_4^{2+}$, $Fe^{2+}$ and a carbon source, so this simplest substrate input set was adopted as the SIS of the SNF state.

### Metabolic properties of the bacteroid reflected by modeling

Using the above defined OF and SIS for the SNF state, the metabolic flux distribution over the network was computed by FVA. The reactions with non-zero fluxes were selected and compared with the expression profile of enzymes; meanwhile, they were also compared with the reactions of non-zero fluxes in the two previously published models, *i*OR450 and *i*HZ565. First, a set of reactions related to dicarboxylate metabolism were detected, which was consistent with the observation that C4 dicarboxylates are the principal carbon source for the *B. diazoefficiens* bacteroids. In the SNF model, the metabolism of C4 dicarboxylates was considerably overlapped with the TCA cycle that completely functions without any gap. The existence of a complete TCA cycle was consistent with *i*HZ565[13] and *i*OR450,[12] and was also in agreement with the proteomic data of *B. diazoefficiens*.[27] Oxidative phosphorylation is linked with the TCA cycle, and is important

Table 4 Quantitative nitrogen output of the SNF models with four kinds of nitrogen output

| Physiological state | $SNF_N$ | $SNF_A$ | $SNF_{NA}$ | $SNF_{NAA}$ |
|---|---|---|---|---|
| Nitrogen output form | $NH_3$ | Ala | $NH_3$ + Ala | $NH_3$ + Ala + Asp |
| Optimal value of OF (mmol $gDW^{-1}$ $h^{-1}$) | 4.31 | 3.23 | 2.60 | 1.75 |
| Nitrogen output flux (mmol $gDW^{-1}$ $h^{-1}$) | 4.31 | 3.23 | 5.21 | 5.25 |

for nitrogen fixation for reducing oxidative damage to nitrogenase by consuming $O_2$ and generating ATP molecules needed for nitrogen fixation. The SNF model contains a complete electron transport chain. All of the four types of complexes for oxidative phosphorylation are embedded in the cell membrane, so it is difficult to extract and detect them using proteomic methods. In spite of the lack of evidence in proteomic data, our result was supported by the same simulation result of *i*OR450 and *i*HZ565 models. As for PHB metabolism, unlike *S. meliloti* which does not accumulate PHB at the bacteroid stage,[13,57] the flux distribution of the SNF model shows the complete PHB biosynthesis pathway that converts considerable amounts of its carbon source into the storage polymer, and this result was in agreement with the proteomic data[25] and with the simulation result of the *i*OR450 model. For porphyrin and chlorophyll metabolism, the synthesis of heme is active, which agrees with the simulation of *i*HZ565.[13] In summary, although there are differences between the rhizobia species, the important characteristics related to symbiotic nitrogen fixation in our model can be confirmed by experimental data and simulation results of other SNF models.

### Metabolic differences between the FL and SNF states

The flux distributions in the FL and SNF models were simulated by FVA. The reactions with non-zero-flux were picked out and they made up the non-zero-fluxes-space. The non-zero-fluxes-space of the FL model is about three times greater than the SNF model (Fig. 5), which reflects that the scale of the metabolic system is significantly affected by the metabolic goals. The flux distributions of FL and SNF models were visualized on the global KEGG metabolic map and the overlap between the two models was treated as the conservative core metabolic flux set (Fig. 5). This set contains 108 reactions accounting for 22.9% of the FL fluxes and 90.8% of the SNF fluxes. Pathway enrichment analysis

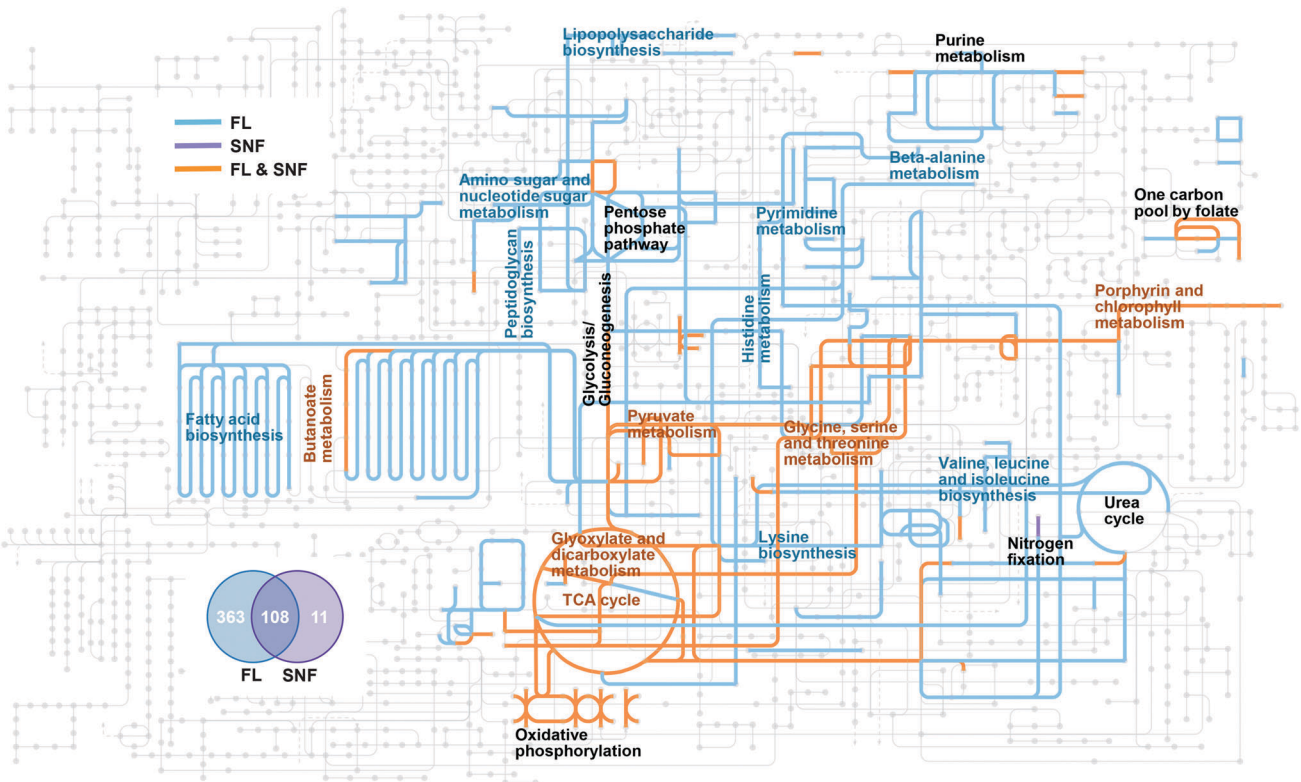


Fig. 5 Overlap of fluxes in the FL and SNF models and the corresponding flux distribution on the global KEGG metabolic map01100. Grey nodes represent metabolites and the lines connecting the nodes represent reactions. The Venn sub-diagram shows the overlap between the flux sets of the FL and SNF models. Three kinds of fluxes in the Venn sub-diagram are visualized on the KEGG map01100 with lines of different colors. Transport reactions, exchange reactions and OFs are not shown. Grey lines represent the reactions with zero flux or the reactions that do not exist in the models. Two kinds of pathways with enriched fluxes in Table 5 are annotated by fonts of different colors: brown, pathways with enriched FL & SNF fluxes; blue-gray, pathways with enriched FL unique fluxes. Other important metabolic pathways with non-zero fluxes are annotated by black fonts.

 

Table 5 Pathways with enriched fluxes in the FL and SNF states simulated by FVA (number of fluxes in a pathway >7, fisher's exact test, $p$-value < 0.05)

| Flux | Pathway ID | Pathway name | Reaction no. | $p$-Value |
|---|---|---|---|---|
| FL unique | bja00061 | Fatty acid biosynthesis | 27 | 0.000159 |
| | bja00240 | Pyrimidine metabolism | 27 | 0.000159 |
| | bja00290 | Valine, leucine and isoleucine biosynthesis | 19 | 0.014037 |
| | bja00540 | Lipopolysaccharide biosynthesis | 13 | 0.015679 |
| | bja00300 | Lysine biosynthesis | 12 | 0.021662 |
| | bja00550 | Peptidoglycan biosynthesis | 11 | 0.029909 |
| | bja00520 | Amino sugar and nucleotide sugar metabolism | 20 | 0.036833 |
| | bja00340 | Histidine metabolism | 10 | 0.041272 |
| | bja00410 | Beta-alanine metabolism | 10 | 0.041272 |
| FL & SNF | bja00020 | Citrate cycle (TCA cycle) | 9 | $2.89 \times 10^{-5}$ |
| | bja00630 | Glyoxylate and dicarboxylate metabolism | 12 | $9.32 \times 10^{-5}$ |
| | bja00650 | Butanoate metabolism | 9 | 0.00217 |
| | bja00860 | Porphyrin and chlorophyll metabolism | 8 | 0.00282 |
| | bja00620 | Pyruvate metabolism | 9 | 0.00424 |
| | bja00260 | Glycine, serine and threonine metabolism | 10 | 0.00900 |

FL unique, fluxes only in FL state; FL & SNF, fluxes shared by FL and SNF states.

was performed to analyze the distribution of FL unique fluxes and the core metabolic fluxes (Table 5 and Fig. 5). The differences between the distributions and enrichments reflect the characteristics of the related physiological states: the large proportion of the unique metabolic fluxes of the FL model is associated with a large number of biomass precursors involved in the anabolism of free-living cells, while the more simplified and centralized metabolic system of the SNF model reflects the stagnation of vegetative growth and the support for nitrogen fixation.

The conservative core fluxes were enriched in the TCA cycle, pyruvate metabolism, glyoxylate and dicarboxylate metabolism, glycine, serine and threonine metabolism, butanoate metabolism and porphyrin and chlorophyll metabolism (number of fluxes in pathway > 7, fisher's exact test, $p$-value < 0.05, Table 5). Among them, TCA metabolism and pyruvate metabolism are major components of the core pathways for the intermediate metabolites; therefore, most of the involved reactions are housekeeping. In addition, glyoxylate and dicarboxylate metabolism involves conversion and utilization of carbon sources; glycine, serine and threonine metabolism is a crossroad that connects many important pathways such as EMP, TCA, purine metabolism, sulfur metabolism, glyoxylate and dicarboxylate metabolism, and especially the metabolism of several other amino acids; butanoate metabolism is related to PHB biosynthesis; porphyrin and chlorophyll metabolism is connected with biosynthesis of heme; all of them are shared by the two models.

FL unique fluxes are mainly enriched in the metabolic pathways of carbohydrates (amino sugar and nucleotide sugar), fatty acid, nucleotides (pyrimidine), amino acids (valine, leucine and isoleucine, lysine, histidine) and cell wall components (peptidoglycan and lipopolysaccharide) (Table 5), and all of these pathways have connections with the synthesis of precursors of biomass and transformation of intermediate metabolites.

## Differences in essential genes

Gene deletion is one of the main strategies to determine gene function, but its experimental process is quite resource-consuming. Furthermore, to some extent it's hard to genetically manipulate the SNF system because of its considerable complexity. So the *in silico* gene deletion that is high-throughput and easy to conduct has unique advantages. A simulated single gene deletion study by FBA was conducted under the FL and SNF conditions. As shown in Fig. 6, the FL model has 110 essential genes, accounting for 18.52% of all the 594 FL genes, while the SNF model has 38, accounting for 16.38%. The lists of essential genes can be found in Tables S2–S4 (ESI†). Details of the results are available in Supplementary file S6, ESI.†

In many cases, an enzyme or its subunits may be encoded by different genes, and the impact of one gene's deletion can be offset by another, so none of them is essential although the corresponding reaction is. This problem can be resolved by *in silico* double or multiple gene deletion. According to double-gene-deletion simulation, both of the FL and SNF models have 47 essential genes pairs, accounting for 0.04% and 0.25% of the 116 886 gene pairs of the FL model and 18 721 of the SNF model, respectively (Fig. S2, ESI†). For example, the reaction

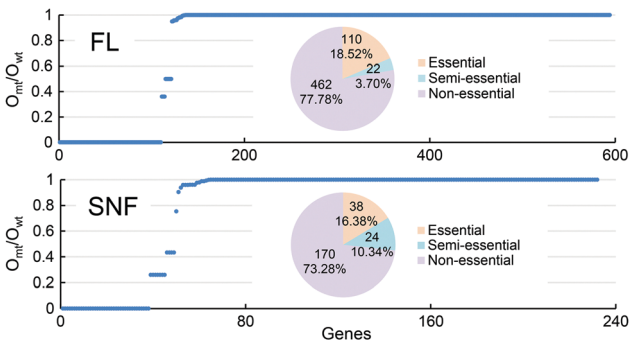


Fig. 6 Simulated single-gene-knockout to the FL and SNF models. The horizontal axis represents genes in metabolic models; only genes proved to be related with fluxes by FVA were considered; the vertical axis represents the ratio of the optimal objective function value for a single gene-deletion strain to that for the wild-type strain. $O_{mt}/O_{wt}$, the ratio of the optimal objective function value for a single gene-deletion strain to that for the wild-type strain. Pie charts show the proportions of three kinds of genes: essential, semi-essential and non-essential.

which reduces acetoacetate to 3-hydroxybutanoate is the final step of biosynthesis of the PHB monomer and is essential for SNF. It is catalyzed by (R)-3-hydroxybutanoate: $NAD^+$ oxidoreductase which is encoded either by gene blr1488 or by blr7029. And both of them were not detected as essential genes during simulated single gene deletion, but recognized as an essential gene pair in the double-gene-deletion simulation. The list of essential gene pairs can be found in Tables S5 and S6 (ESI†), and they have a reference value for the future physiology and molecular biology research on symbiotic nitrogen fixation. The details on the results of simulated double-gene-knockout are shown in Fig. S3 and Supplementary file S6, ESI.†

### Differences in essential reactions

Essentiality analysis for reactions can reflect the extent to which the deletion of a specific reaction influences certain phenotypes of a specific physiological state. As shown in Fig. S4 (ESI†), the FL model has 190 essential reactions (40.34% of all the FL reactions), while the SNF model only has 23 (19.33% of all the SNF reactions). It can be noticed that the proportion of essential reactions in the FL model (40.34%) is obviously larger than that in the SNF model (19.33%), which is different from the situation of essential genes where the proportion of essential genes in the FL model (18.52%) is quite similar to that in the SNF model (16.38%), indicating the fact that essential reactions and essential genes reflect different levels of essentiality: essential reactions may be overlooked by simulated single gene deletion due to gene redundancy. Details of the simulation results can be obtained from Supplementary file S6, ESI.†

The distribution of three kinds of reactions (essential, semi-essential and non-essential) in the metabolic pathways (transport reactions, exchange reactions and OF were excluded) is shown in Fig. 7. By definition, the deletion of an essential reaction will result in the complete blockage of the synthetic pathway for the related precursors of OF components, so it is most important to analyze the essential ones. Pathway enrichment analysis was performed to analyze the distribution of essential reactions of the FL and SNF models, and the enriched pathways (number of fluxes in pathway $> 6$, fisher's exact test, $p$-value $< 0.05$) are shown in Table 6. It was found that 12 essential reactions in the SNF model are related to nitrogen fixation, heme biosynthesis and PHB precursor synthesis, and 7 of them are enriched in the porphyrin and chlorophyll metabolism. In the FL model, 178 essential reactions are associated with the metabolism of many precursors of the biomass components and enriched in the metabolic pathways of carbohydrates (amino sugar and nucleotide sugar), lipids (fatty acid and glycerophospholipid), amino acids (valine, leucine, isoleucine, histidine, lysine) and cell wall components (peptidoglycan and lipopolysaccharide). All of these results show the metabolic differences between the two physiological states.

### Differences in the topological structure of the metabolic system

By comparing Tables 5 and 6, we came to an interesting conclusion that for the FL model, essential reactions were mainly associated with the metabolism of precursors of the biomass

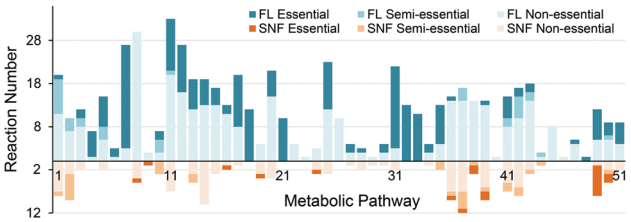


Fig. 7 Distribution of reactions with different essentialities in metabolic pathways. Horizontal axis displays 51 metabolic pathways. Pathway name: (1) glycolysis/gluconeogenesis; (2) citrate cycle (TCA cycle); (3) pentose phosphate pathway; (4) pentose and glucuronate interconversions; (5) fructose and mannose metabolism; (6) galactose metabolism; (7) fatty acid biosynthesis; (8) fatty acid degradation; (9) synthesis and degradation of ketone bodies; (10) oxidative phosphorylation; (11) purine metabolism; (12) pyrimidine metabolism; (13) alanine, aspartate and glutamate metabolism; (14) glycine, serine and threonine metabolism; (15) cysteine and methionine metabolism; (16) valine, leucine and isoleucine degradation; (17) valine, leucine and isoleucine biosynthesis; (18) lysine biosynthesis; (19) lysine degradation; (20) arginine and proline metabolism; (21) histidine metabolism; (22) tyrosine metabolism; (23) phenylalanine metabolism; (24) tryptophan metabolism; (25) phenylalanine, tyrosine and tryptophan biosynthesis; (26) beta-alanine metabolism; (27) D-glutamine and D-glutamate metabolism; (28) D-alanine metabolism; (29) glutathione metabolism; (30) starch and sucrose metabolism; (31) amino sugar and nucleotide sugar metabolism; (32) lipopolysaccharide biosynthesis; (33) peptidoglycan biosynthesis; (34) glycerolipid metabolism; (35) glycerophospholipid metabolism; (36) pyruvate metabolism; (37) glyoxylate and dicarboxylate metabolism; (38) propanoate metabolism; (39) butanoate metabolism; (40) C5-branched dibasic acid metabolism; (41) one carbon pool by folate; (42) methane metabolism; (43) carbon fixation in photosynthetic organisms; (44) carbon fixation pathways in prokaryotes; (45) vitamin B6 metabolism; (46) nicotinate and nicotinamide metabolism; (47) pantothenate and CoA biosynthesis; (48) folate biosynthesis; (49) porphyrin and chlorophyll metabolism; (50) nitrogen metabolism; (51) sulfur metabolism.

components rather than the conservative core metabolism. To further investigate this phenomenon, we quantitatively explored the overlap between the essential reaction set and the unique flux sets (the FL flux set and the SNF flux set in Fig. 5) or the conservative core flux set (the FL & SNF flux set in Fig. 5). As shown in Fig. 8, the proportions of essential reactions to the core flux set are 12.96% and 14.81%, while to the unique flux sets are 48.48% and 63.64% (please note that the number of SNF unique fluxes is relatively low). This showed that essential reactions are indeed mainly associated with the unique flux sets.

It is well known that in house-keeping core metabolic pathways such as EMP, TCA and glycine, serine and threonine metabolism, many key metabolites are hubs of the network with multiple generating and consuming pathways: if one pathway is blocked, it can be replaced by another. This implies that the differences between the core flux set and the unique flux set may be due to the redundancy of paths between metabolites in different parts of the metabolic network. The redundancy of the paths of the network can be reflected by the distribution of degree and average shortest path lengths between nodes, so these two topological indicators were analyzed and compared for the core sub-network (core flux set) and the unique sub-network (unique flux set). The result showed that for the FL model, the degree of nodes in the core sub-network is indeed higher than

 

Table 6 Pathways in the FL and SNF models with enriched essential reactions simulated by single reaction deletion (number of fluxes in pathway >6, fisher's exact test, $p$-value < 0.05)

| Model | Pathway ID | Pathway name | Reaction no. | $p$-Value |
|---|---|---|---|---|
| FL | bja00061 | Fatty acid biosynthesis | 24 | $1.91 \times 10^{-8}$ |
| | bja00520 | Amino sugar and nucleotide sugar metabolism | 19 | $1.92 \times 10^{-6}$ |
| | bja00540 | Lipopolysaccharide biosynthesis | 13 | $2.09 \times 10^{-6}$ |
| | bja00300 | Lysine biosynthesis | 12 | $5.83 \times 10^{-6}$ |
| | bja00550 | Peptidoglycan biosynthesis | 11 | $1.62 \times 10^{-5}$ |
| | bja00340 | Histidine metabolism | 10 | $4.47 \times 10^{-5}$ |
| | bja00564 | Glycerophospholipid metabolism | 9 | 0.0184 |
| | bja00290 | Valine, leucine and isoleucine biosynthesis | 12 | 0.0300 |
| SNF | bja00860 | Porphyrin and chlorophyll metabolism | 7 | $2.03 \times 10^{-6}$ |

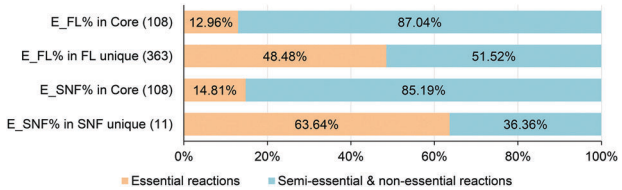


Fig. 8 Essentiality of reactions in the conservative core metabolism system and a unique metabolism system. Core, fluxes shared by the FL and SNF models determined by FVA; FL unique, fluxes only existing in the FL model determined by FVA; SNF unique, fluxes only existing in the SNF model determined by FVA; E_FL, essential reactions in the FL model; E_SNF, essential reactions in the SNF model. The total numbers of fluxes are shown in parentheses.

the nodes in the unique sub-network (Fig. 9A, $p$-value = $2.78 \times 10^{-6}$, two-sided Wilcoxon rank sum test, median: 6 *vs.* 4, $n$: 101 *vs.* 361, $\alpha = 0.05$) and the average shortest path length is indeed lower for nodes in the core sub-network (Fig. 9C, $p$-value = $2.69 \times 10^{-15}$, two-sided Wilcoxon rank sum test, median: 4.40 *vs.* 5.71, $n$: 101 *vs.* 361, $\alpha = 0.05$), thus the structure and function of the conservative core sub-network will not be seriously affected by single reaction deletion and the essential reactions are enriched in the unique sub-network. For the SNF model, the differences are not significant (Fig. 9B and D; for the degree, $p$-value = 0.389, two-sided Wilcoxon rank sum test, median: 4 *vs.* 4, $n$: 97 *vs.* 18, $\alpha = 0.05$; for the average shortest path length, $p$-value = 0.154, two-sided Wilcoxon rank sum test, median: 4.31 *vs.* 3.74, $n$: 97 *vs.* 18, $\alpha = 0.05$). Compared with the FL model, the unique sub-network of the SNF model is relatively small (only have 18 nodes, Fig. 9B); as a result, its topological properties might not be reflected as remarkably as the larger FL model.

### Analysis of the regulation in the SNF system by metabolic control analysis

To understand the regulation of the SNF system is one of the major goals of this research, so MCA was performed. MCA quantifies the response of system variables to the changes of other system variables and the response is reflected by control coefficients. In this research, FCC was used to depict the flux sensitivity of the OF of the SNF model to the changes of other fluxes. As in the four types of fluxes in FVA results, all of the FCCs for zero, infinite and changeable fluxes are 0, so only the FCC of the fixed fluxes is shown in Supplementary file S6, ESI.† Perturbations, which may significantly affect (FCC > 0.3) symbiotic nitrogen fixation, are highlighted in Table 7. The results show that 19 reactions may have relatively stronger constraints on symbiotic nitrogen fixation, such as the biosynthesis of ATP, reducing power, PHB and heme as well as the related transport of important substrates. Given the pivotal regulative role of these reactions for the target output of the SNF system, it may be instructive to investigate the functions and regulation mechanisms of the 93 genes related to these key reactions for improving the efficiency of symbiotic nitrogen fixation.

## Discussion

Constructing an OF that appropriately mimics the corresponding physiological state is critical for constraint-based reconstruction and simulation. Previous studies have shown contradictions about the output forms of nitrogen-fixing production. The most controversial issue on this subject was the nitrogen output form: ammonium or alanine?[30,32] Li *et al.* suspected that the difference is caused by the effect of different extraction methods,[30] while White *et al.* believed that the output form of ammonia/alanine depends on the substrate supply (especially ammonia) to alanine dehydrogenase,[58] which had been confirmed on other symbiotic nitrogen fixation systems.[31,59] In order to determine a reasonable OF to reflect the nitrogen output, we set up four kinds of SNF OFs and got four corresponding SNF models by simulation with FBA and FVA. Simulation results showed that there was no significant difference between the flux distributions of the four SNF models, especially in the core metabolic network shared by them (Fig. 4A and C). Based on the results of our simulation and analysis, combined with the previous experiments, we suggest that conversions between the four nitrogen output forms might just involve minor adjustments in the structure of the metabolic network if a relatively stable nitrogen output flux is maintained. This reflects the robustness and flexibility of the SNF metabolic network.

According to the results of analysis for the topological structure of the metabolic network, the FL network can be divided into two parts: the highly redundant, conservative, robust core and the lowly redundant, diversified input and output, and this is exactly the typical structural features of a bowtie structure that is an important topological feature of complex networks with the conservative core called a giant strong component (GSC).[60]

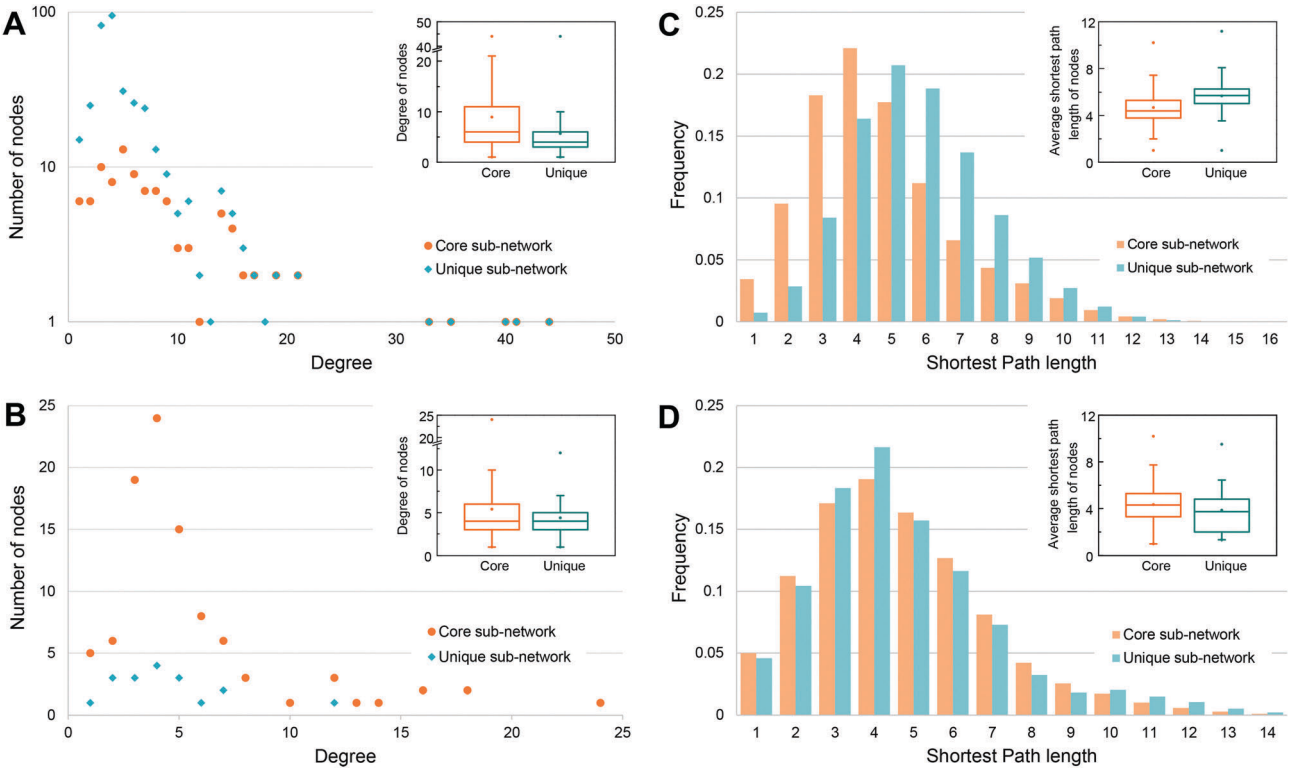


Fig. 9 Differences in the topological structures of the FL and SNF metabolic systems. The distribution of degree of nodes (A and B) and average shortest path length of nodes (C and D) of the core sub-network and unique sub-network in the FL model (A and C) and the SNF model (B and D) were used to analyze the differences in the topological structure of the FL and SNF metabolic systems. Given the large range of degree of nodes in the FL model, the vertical axis of (A) is shown logarithmically. In the sub-boxplot, the box shows the first quartile, the median, and the third quartile; the dot inside the box represents the mean value and the dots below and above the whisker represent the minimum and maximum values of each data set. Core, nodes in the core sub-network; unique, nodes in the unique sub-network.

Table 7 Reactions that may have relatively important influence on symbiotic nitrogen fixation in *B. diazoefficiens* USDA110 detected by metabolic control analysis (FCC > 0.3)

| Reaction ID | FCC (%) | Reaction description | Subsystem | EC number | KEGG ID |
|---|---|---|---|---|---|
| NIT | 100.0 | Reduced ferredoxin: dinitrogen oxidoreductase | Nitrogen fixation | 1.18.6.1 | R05185 |
| FDXNRy | 45.9 | Ferredoxin: $NADP^+$ oxidoreductase | Nitrogen fixation | 1.18.1.2 | |
| COBOU | 63.8 | Cytochrome oxidase bo3 | Oxidative phosphorylation | 1.9.3.1 | |
| ATPS4r | 59.3 | ATP synthase | Oxidative phosphorylation | 3.6.3.14 | |
| ACACT1r | 100.0 | Acetyl-CoA: acetyl-CoA C-acetyltransferase | PHB monomer synthesis | 2.3.1.9 | R00238 |
| BDH | 100.0 | (R)-3-Hydroxybutanoate: $NAD^+$ oxidoreductase | PHB monomer synthesis | 1.1.1.30 | R01361 |
| COAT1 | 100.0 | Acetoacetyl-CoA: acetate CoA-transferase | PHB monomer synthesis | 2.8.3.8 | R01359 |
| CPPPGO | 100.0 | Coproporphyrinogen: oxygen oxidoreductase | Porphyrin and chlorophyll metabolism | 1.3.3.3 | R03220 |
| FCLT | 100.0 | Protoheme ferrolyase | Porphyrin and chlorophyll metabolism | 4.99.1.1 | R00310 |
| HMBS | 100.0 | Porphobilinogen ammonialyase | Porphyrin and chlorophyll metabolism | 2.5.1.61 | R00084 |
| PPBNGS | 100.0 | 5-Aminolevulinate hydrolyase | Porphyrin and chlorophyll metabolism | 4.2.1.24 | R00036 |
| PPPGO | 100.0 | Protoporphyrinogen-IX: oxygen oxidoreductase | Porphyrin and chlorophyll metabolism | 1.3.3.4 | R03222 |
| UPP3S | 100.0 | Hydroxymethylbilane hydrolyase | Porphyrin and chlorophyll metabolism | 4.2.1.75 | R03165 |
| UPPDC1 | 100.0 | Uroporphyrinogen-III carboxylyase | Porphyrin and chlorophyll metabolism | 4.1.1.37 | R03197 |
| FE2t | 100.0 | $Fe^{2+}$ transport | Substrate transport | | |
| N2t | 100.0 | Nitrogen transport | Substrate transport | | |
| O2t | 63.9 | Oxygen transport | Substrate transport | | |
| SUCCt2_2 | 37.4 | Succinate transport | Substrate transport | | |
| MALt2 | 31.0 | L-Malate transport | Substrate transport | | |

As for the bowtie structure, the number of alternative conversion pathways between the nodes of the GSC is greater than the average of the entire network.[61] A series of hubs in the GSC (key metabolites) constitute the universal interfaces through which various input and output processes are connected with the versatile metabolic core. The corresponding metabolites can be easily converted into others.[62] In order to survive and reproduce, free-living rhizobia must synthesize highly complex biomass and meanwhile ensure the simplicity of the overall topological structure for the low cost of the metabolic system,

 

so the bowtie structure of the FL metabolic network can be well positioned to resolve this contradiction. In contrast, due to the relatively simple metabolic target of nitrogen fixation, a relatively condensed metabolic network is enough for the SNF state, so the input and output of the SNF network are rare and the bowtie feature is not significantly manifested.

Differences in the size and bowtie structure between the FL and SNF models reflect environmental selective pressure on the metabolic system. During the process of establishing a symbiotic relationship, the host plants played an important role of domesticator.[63] Rewards and punishments between the members of the symbiotic system may be the most important strategy in the processes of forming and maintaining a stable symbiotic relationship.[64,65] The corresponding strategies have also been effectively supported by experiment and simulation evidences in the symbiotic nitrogen fixation systems.[66–69] Consequently, the scale and structural features of the SNF metabolic network depend largely on the interaction between rhizobia and host plants.

## Conclusions

In this work, the first genome-scale metabolic network for *Bradyrhizobium*, *i*YY1101, was reconstructed. Based on it, models of a free-living rhizobium and a symbiotic bacteroid were presented, and the properties of these two physiological states and the differences between them were simulated and compared at the metabolic level for the first time. We hope that our models can serve as a promising platform for better understanding symbiotic nitrogen fixation and that the results of the simulated knockouts of genes/reactions and the metabolic control analysis can provide new clues for improving the efficiency of symbiotic nitrogen fixation.

## Conflicts of interest

The authors declare that they have no conflict of interest.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant 31570844 and 31100602), the National Basic Research Program of China (973 project, Grant 2013CB127103), Project J1103510 supported by NSFC, Project 2010QC016, 2011PY070, 2013JC009 and 2662016PY094 supported by the Fundamental Research Funds for the Central Universities, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. We thank Dasong Chen and Youguo Li for assistance with the physiological background of symbiotic nitrogen fixation and sample preparation for the PM experiment.

## Notes and references

1 J. Prell and P. Poole, *Trends Microbiol.*, 2006, **14**, 161–168.
2 C. Masson-Boivin, E. Giraud, X. Perret and J. Batut, *Trends Microbiol.*, 2009, **17**, 458–466.
3 E. Akay and E. L. Simms, *Am. Nat.*, 2011, **178**, 1–14.
4 Y. Tatsukami, M. Nambu, H. Morisaka, K. Kuroda and M. Ueda, *BMC Microbiol.*, 2013, **13**, 180–188.
5 N. W. Oehrle, A. D. Sarma, J. K. Waters and D. W. Emerich, *Phytochemistry*, 2008, **69**, 2426–2438.
6 M. Libault, A. Farmer, L. Brechenmacher, J. Drnevich, R. J. Langley, D. D. Bilgin, O. Radwan, D. J. Neece, S. J. Clough, G. D. May and G. Stacey, *Plant Physiol.*, 2010, **152**, 541–552.
7 H. J. Pel, J. H. de Winde, D. B. Archer, P. S. Dyer, G. Hofmann, P. J. Schaap, G. Turner, R. P. de Vries, R. Albang, K. Albermann, M. R. Andersen, J. D. Bendtsen, J. A. E. Benen, M. van den Berg, S. Breestraat, M. X. Caddick, R. Contreras, M. Cornell, P. M. Coutinho, E. G. J. Danchin, J. M. Debets, P. Dekker, P. W. M. van Dijck, A. van Dijk, L. Dijkhuizen, A. J. M. Driessen, C. d'Enfert, S. Geysens, C. Goosen, G. S. P. Groot, P. W. J. de Groot, T. Guillemette, B. Henrissat, M. Herweijer, J. P. T. W. van den Hombergh, C. A. M. J. J. van den Hondel, R. T. J. M. van der Heijden, R. M. van der Kaaij, F. M. Klis, H. J. Kools, C. P. Kubicek, P. A. van Kuyk, J. Lauber, X. Lu, M. J. E. C. van der Maarel, R. Meulenberg, H. Menke, M. A. Mortimer, J. Nielsen, S. G. Oliver, M. Olsthoorn, K. Pal, N. N. M. E. van Peij, A. F. J. Ram, U. Rinas, J. A. Roubos, C. M. J. Sagt, M. Schmoll, J. Sun, D. Ussery, J. Varga, W. Vervecken, P. J. J. van de Vondervoort, H. Wedler, H. A. B. Wösten, A. P. Zeng, A. J. J. van Ooyen, J. Visser and H. Stam, *Nat. Biotechnol.*, 2007, **25**, 221–231.
8 N. C. Duarte, M. J. Herrgard and B. Ø. Palsson, *Genome Res.*, 2004, **14**, 1298–1309.
9 H. H. Cheng, L. M. Whang, C. A. Lin, I. C. Liu and C. W. Wu, *Bioresour. Technol.*, 2013, **141**, 233–239.
10 S. Schatschneider, M. Persicke, S. A. Watt, G. Hublik, A. Pühler, K. Niehaus and F. Vorhölter, *J. Biotechnol.*, 2013, **167**, 123–134.
11 O. R. Antonio, J. L. Reed, S. Encarnacion, J. C. Vides and B. Ø. Palsson, *PLoS Comput. Biol.*, 2007, **3**, 1887–1895.
12 O. R. Antonio, M. Hernandez, E. Salazar, S. Contreras, G. M. Batallar, Y. Mora and S. Encarnacion, *BMC Syst. Biol.*, 2011, **5**, 120–134.
13 H. S. Zhao, M. Li, K. C. Fang, W. F. Chen and J. Wang, *PLoS One*, 2012, **7**, e31287.
14 A. Willems, *Plant Soil*, 2006, **287**, 3–14.
15 W. K. Gillette and G. H. Elkan, *J. Bacteriol.*, 1996, **178**, 2757–2766.
16 M. Hahn and H. Hennecke, *Appl. Environ. Microbiol.*, 1987, **53**, 2247–2252.
17 J. R. M. Delamuta, R. A. Ribeiro, E. Ormeño-Orrillo, I. S. Melo, E. Martínez-Romero and M. Hungria, *Int. J. Syst. Evol. Microbiol.*, 2013, **63**, 3342–3351.
18 B. E. Caldwell, *Agron. J.*, 1969, **61**, 813–815.
19 D. W. Israel, *Agron. J.*, 1981, **73**, 509–516.

20 K. R. Schubert, K. T. Jennings and H. J. Evans, *Plant Physiol.*, 1978, **61**, 398–401.
21 W. J. Hunter and L. D. Kuykendall, *Appl. Environ. Microbiol.*, 1990, **56**, 2399–2403.
22 L. D. Kuykendall, F. M. Hashem and W. J. Hunter, *Plant Soil*, 1996, **186**, 121–125.
23 T. Kaneko, Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. S. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada and S. Tabata, *DNA Res.*, 2002, **9**, 189–197.
24 G. Pessi, C. H. Ahrens, H. Rehrauer, A. Lindemann, F. Hauser, H. S. Fischer and H. Hennecke, *Mol. Plant-Microbe Interact.*, 2007, **20**, 1353–1363.
25 A. D. Sarma and D. W. Emerich, *Proteomics*, 2005, **5**, 4170–4184.
26 A. D. Sarma and D. W. Emerich, *Proteomics*, 2006, **6**, 3008–3028.
27 N. Delmotte, C. H. Ahrens, C. Knief, E. Qeli, M. Koch, H. M. Fischer, J. A. Vorholt, H. Hennecke and G. Pessi, *Proteomics*, 2010, **10**, 1391–1400.
28 D. W. Emerich and H. B. Krishnan, *Biochem. J.*, 2014, **460**, 1–11.
29 S. D. Tyerman, L. F. Whitehead and D. A. Day, *Nature*, 1995, **378**, 629–632.
30 Y. Li, R. Parsons, D. A. Day and F. J. Bergersen, *Microbiology*, 2002, **148**, 1959–1966.
31 D. Allaway, E. M. Lodwig, L. A. Crompton, M. Wood, R. Parsons, T. R. Wheeler and P. S. Poole, *Mol. Microbiol.*, 2000, **36**, 508–515.
32 J. K. Waters, B. L. Hughes II, L. C. Purcell, K. O. Gerhardt, T. P. Mawhinney and D. W. Emerich, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12038–12042.
33 L. F. Whitehead, S. Young and D. A. Day, *Soil Biol. Biochem.*, 1998, **30**, 1583–1589.
34 M. Udvardi and P. S. Poole, *Annu. Rev. Plant Biol.*, 2013, **64**, 781–805.
35 C. M. Brown and M. J. Dilworth, *J. Gen. Microbiol.*, 1975, **86**, 39–48.
36 B. R. Bochner, P. Gadzinski and E. Panomitros, *Genome Res.*, 2001, **11**, 1246–1255.
37 I. Thiele and B. Ø. Palsson, *Nat. Protoc.*, 2010, **5**, 93–121.
38 J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke and B. Ø. Palsson, *Nat. Protoc.*, 2011, **6**, 1290–1307.
39 B. Dreyfus, J. L. Garcia and M. Gillis, *Int. J. Syst. Bacteriol.*, 1988, **38**, 89–98.
40 L. M. Xu, C. Ge, Z. Cui, J. Li and H. Fan, *Int. J. Syst. Bacteriol.*, 1995, **45**, 706–711.
41 P. G. Ansari, D. L. N. Rao and K. K. Pal, *Ann. Microbiol.*, 2014, **64**, 1553–1565.
42 S. A. Becker and B. Ø. Palsson, *PLoS Comput. Biol.*, 2008, **4**, e1000082.
43 J. D. Orth, I. Thiele and B. Ø. Palsson, *Nat. Biotechnol.*, 2010, **28**, 245–248.
44 R. Mahadevan and C. H. Schilling, *Metab. Eng.*, 2003, **5**, 264–276.
45 D. A. Fell, *Biochem. J.*, 1992, **286**, 313–330.
46 B. R. Bochner, P. Gadzinski and E. Panomitros, *Genome Res.*, 2001, **11**, 1246–1255.
47 T. Fuhrer, E. Fischer and U. Sauer, *J. Bacteriol.*, 2005, **187**, 1581–1590.
48 Y. Li, L. S. Green, R. Holtzapffel and D. A. Day, Fraser, *Microbiology*, 2001, **147**, 663–670.
49 S. A. Kim and L. Copeland, *Appl. Environ. Microbiol.*, 1996, **62**, 4186–4190.
50 A. Fabra, J. Angelini, A. Donolo, M. Permigiani and S. Castro, *Anton. Leeuw. Int. J. G.*, 1998, **73**, 223–228.
51 P. P. Wong and H. J. Evans, *Plant Physiol.*, 1971, **47**, 750–755.
52 D. P. S. Verma and A. K. Bal, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 3843–3847.
53 K. D. Nadler, *Plant Physiol.*, 1977, **60**, 433–436.
54 O. Brain and R. Mark, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8390–8393.
55 B. Boussau, E. O. Karlberg, A. C. Frank, B. A. Legault and S. G. E. Andersson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9722–9727.
56 E. Lodwig and P. Poole, *Crit. Rev. Plant Sci.*, 2003, **22**, 37–78.
57 C. X. Wang, M. Saldanha, X. Y. Sheng, K. J. Shelswell, K. T. Walsh, B. W. S. Sobral and T. C. Charles, *Microbiology*, 2007, **153**, 388–398.
58 J. White, J. Prell, E. K. James and P. Poole, *Plant Physiol.*, 2007, **144**, 604–614.
59 S. Kumar, A. Bourdes and P. S. Poole, *J. Bacteriol.*, 2005, **187**, 5493–5495.
60 M. Csete and J. Doyle, *Trends Biotechnol.*, 2004, **22**, 446–450.
61 H. W. Ma and A. P. Zeng, *Bioinformatics*, 2003, **19**, 1423–1430.
62 H. Kitano, *Nat. Rev. Genet.*, 2004, **5**, 826–837.
63 M. Koch, N. Delmotte, H. Rehrauer, J. A. Vorholt, G. Pessi and H. Hennecke, *Mol. Plant-Microbe Interact.*, 2010, **6**, 784–790.
64 D. W. Yu, *Biol. J. Linn. Soc.*, 2001, **72**, 529–546.
65 E. T. Kiers, M. Duhamel, Y. Beesetty, J. A. Mensah, O. Franken, E. Verbruggen, C. R. Fellbaum, G. A. Kowalchuk, M. M. Hart, A. Bago, T. M. Palmer, S. A. West, P. Vandenkoornhuyse, J. Jansa and H. Bücking, *Science*, 2011, **33**, 880–882.
66 R. F. Denison, *Am. Nat.*, 2000, **156**, 567–576.
67 S. A. West, E. T. Kiers, E. L. Simms and R. F. Denison, *Proc. R. Soc. London, Ser. B*, 2002, **269**, 685–694.
68 S. A. West, E. T. Kiers, I. Pen and R. F. Denison, *J. Evol. Biol.*, 2002, **15**, 830–837.
69 E. T. Kiers, R. A. Rousseau, S. A. West and R. F. Denison, *Nature*, 2003, **425**, 78–81.

 