Journal of Experimental Botany, Vol. 76, No. 4 pp. 1164–1178, 2025
https://doi.org/10.1093/jxb/erae446 Advance Access Publication 2 November 2024

RESEARCH PAPER

Host-dependent specialized metabolism of nitrogen export in actinorhizal nodules induced by diazotrophic *Actinomycetota Frankia* cluster-2

Fede Berckx[1,2,‡], Thanh Van Nguyen[1,‡], Rolf Hilker[3], Daniel Wibberg[4,5], Kai Battenberg[6,†], Jörn Kalinowski[4], Alison Berry[6], and Katharina Pawlowski[1,*]

[1] Department of Ecology, Environment and Plant Science, Stockholm University, 106 91 Stockholm, Sweden
[2] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden
[3] German Center for Infection Research, Institute for Medical Microbiology, Justus Liebig University Giessen, D-35392 Giessen, Germany
[4] Center for Biotechnology (CeBiTec), Bielefeld University, D-33615 Bielefeld, Germany
[5] Institute of Bio- and Geosciences IBG-5, Computational Metagenomics, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
[6] Department of Plant Sciences, University of California Davis, Davis, CA 95616, USA
[†] Present address: Stillinger Herbarium, University of Idaho, Moscow, ID 83844-3026, USA

[‡] These authors contributed equally to this work.
* Correspondence: katharina.pawlowski@su.se

Received 7 June 2024; Editorial decision 18 October 2024; Accepted 29 October 2024

Editor: Oswaldo Valdéz López, Universidad Nacional Autonoma de Mexico, Mexico

Abstract

***Frankia* cluster-2 strains are diazotrophs that engage in root nodule symbiosis with actinorhizal plants of the *Cucurbitales* and the *Rosales*. Previous studies have shown that an assimilated nitrogen source, presumably arginine, is exported to the host in nodules of *Datisca glomerata* (*Cucurbitales*), while a different metabolite is exported in the nodules of *Ceanothus thyrsiflorus* (*Rosales*). To investigate if an assimilated nitrogen form is commonly exported to the host by cluster-2 strains, and which metabolite would be exported in *Ceanothus*, we analysed gene expression levels, metabolite profiles, and enzyme activities in nodules. We conclude that the export of assimilated nitrogen in symbiosis seems to be a common feature for *Frankia* cluster-2 strains, but the source of nitrogen is host dependent. The export of assimilated ammonium to the host suggests that 2-oxoglutarate is drawn from the tricarboxylic acid (TCA) cycle at a high rate. This specialized metabolism obviates the need for the reductive branch of the TCA cycle. We found that several genes encoding enzymes of central carbon and nitrogen metabolism were lacking in *Frankia* cluster-2 genomes: the glyoxylate shunt and succinate semialdehyde dehydrogenase. This led to a linearization of the TCA cycle, and we hypothesized that this could explain the low saprotrophic potential of *Frankia* cluster-2.**

Abbreviations: *argB*, acetylglutamate kinase; *argC*, *N*-acetyl-γ-glutamyl-phosphate reductase; *argD*, acetylornithine/succinyldiaminopimelate aminotransferase; *argE/argJ*, bifunctional gene acetylornithine deacetylase; *argF*, ornithine carbamoyltransferase; *argG*, argininosuccinate synthase; *argH*, argininosuccinate lyase; BAP, basic propionate; *citA/citA4*, citrate synthase; GABA, γ-aminobutyrate; *gdh*, glutamate dehydrogenase; *glnA1/glnA2*, glutamine synthetase; *glnII*, glutamine synthetase; *gltA/gltA2*, citrate synthase; *gltB*, glutamate synthase, large subunit; *gltD*, glutamate synthase, small subunit; GS/GOGAT, glutamine synthetase/glutamate synthase; *icd*, isocitrate dehydrogenase; *mdh*, malate dehydrogenase; 2-OG, 2-oxoglutarate; *pepck*, phosphoenolpyruvate carboxykinase; SSA, succinic semialdehyde; SSA-DH, succinic semialdehyde dehydrogenase; TCA, tricarboxylic acid.

Keywords: Actinorhizal symbiosis, *Frankia*, GS synthetase, nitrogenase, root nodules, succinic semialdehyde dehydrogenase, TCA cycle.

Introduction

Root nodules are formed as the result of a symbiotic relationship between a nitrogen- (N) fixing soil bacterium and its host plant. All root nodule-forming plants belong to a single clade (Soltis *et al.*, 1995). While legumes (*Fabaceae*, *Fabales*) and *Parasponia* (*Cannabaceae*, *Rosales*) engage with rhizobia, actinorhizal symbiosis is established between a diverse group of plants within the *Rosales*, *Fagales*, and *Cucurbitales*, and their endosymbiont *Frankia*. Recent phylogenomic studies support the hypothesis that the common ancestor to all nodulating plants was symbiotic, but the symbiosis was subsequently lost in the majority of the lineages (Griesmann *et al.*, 2018; van Velzen *et al.*, 2018). Unlike rhizobia, *Frankia* strains do not depend on their host plant for the oxygen protection of the nitrogenase enzyme. Therefore, the polyphyly of the oxygen protection system for nitrogenase in nodules indicates that the original symbiont would have been *Frankia* (van Velzen *et al.*, 2019).

The bacterial genus *Frankia* can be split into four phylogenetically distinct clades, which are referred to as cluster-1 to -4 (Normand *et al.*, 1996; Nguyen *et al.*, 2016). The first three clusters represent symbiotic strains, and the phylogeny roughly coincides with host specificity. Cluster-4 encompasses strains that do not engage in symbiosis (Normand *et al.*, 1996). With the sole exception of *Frankia coriariae* (Gtari *et al.*, 2015; Nouioui *et al.*, 2017; Gueddou *et al.*, 2019), and despite numerous efforts, cluster-2 strains cannot be cultivated *in vitro*. This implies that most analyses, such as gene expression studies via quantitative reverse transcription–PCR (RT–qPCR), must be conducted on nodules. Studies on all *Frankia* strains are further complicated by the fact that stable transformation has not been reproduced since 2019 (Gifford *et al.*, 2019; Pesce *et al.*, 2019) despite numerous attempts (Kucho et al., 2022).

Regardless of the challenges involved in the investigation of *Frankia* cluster-2, this cluster is particularly interesting. *Frankia* cluster-2 represents the earliest divergent symbiotic clade (Sen *et al.*, 2014; Gtari *et al.*, 2015; Persson *et al.*, 2015; Berckx *et al.*, 2022, 2024). Given that the original root nodule symbiont was *Frankia* (van Velzen *et al.*, 2019), cluster-2 represents the closest relative to the first N-fixing endobacterium. In addition, while host plants of cluster-1 are solely found within actinorhizal *Fagales* and most of the host plants of *Frankia* cluster-3 are found within the *Rosales*, cluster-2 *Frankia* has a broad host range. They engage in symbiosis with all actinorhizal *Cucurbitales*, namely the families *Datiscaceae* and *Coriariaceae*. Cluster-2 also engages with actinorhizal *Rosales* of the family *Rosaceae* and with *Ceanothus* spp. (*Rhamnaceae*).

Aside from the pure taxonomic diversity of actinorhizal host plants, actinorhizal symbiosis shows metabolic diversity, for instance in terms of the exchange of nutrients. During symbiosis, the host plant is provided with fixed N in exchange for photosynthates. In legumes and most actinorhizal plants, such as *Alnus glutinosa* (*Betulaceae*, *Fagales*), the fixed N is exported to the cytosol of infected host cells as ammonium (NH_4^+) (Guan *et al.*, 1996; Alloisio *et al.*, 2010; Udvardi and Poole, 2013; Hay *et al.*, 2020). Ammonium is assimilated in the plant cytosol, and transported to the xylem in different forms depending on the plant species. In actinorhizal nodules of *A. glutinosa*, it was shown that ammonium is assimilated via the plant glutamine synthetase/glutamate synthase (GS/GOGAT) cycle. The assimilation by *Frankia* only occurs at low levels (Guan *et al.*, 1996; Alloisio *et al.*, 2010). In *Datisca glomerata*, several studies indicate that an assimilated form of N, presumably arginine, is exported from the endosymbiont to the cytosol of the infected cell and then transported to the surrounding uninfected cells (Berry *et al.*, 2004, 2011; Salgado *et al.*, 2018). In the cytosol of the uninfected cells, arginine is broken back down to ammonium and re-assimilated via the GS/GOGAT cycle (Berry *et al.*, 2004). A similar accumulation of arginine could be found in nodules of *Coriaria myrtifolia* (*Coriariaceae*, *Cucurbitales*) (Wheeler and Bond, 1970). In conclusion, the export of assimilated N, specifically arginine, from the bacterial symbiont to the plant seems to be a common feature for actinorhizal *Cucurbitales*, which are all host plants of *Frankia* cluster-2. However, no evidence for the export of arginine by *Frankia* cluster-2 has been found in nodules of *Ceanothus thyrsiflorus* (*Rhamnaceae*, *Rosales*) based on plant gene expression levels (Salgado *et al.*, 2018). It has been previously shown that asparagine accumulates in nodules of *Ceanothus velutinus* (Wheeler and Bond, 1970). So while arginine can be excluded as the form of exported N by *Frankia* cluster-2 in *Ceanothus* spp., the export of assimilated N cannot be excluded as such. We wanted to identify the exported N metabolite to find out whether the specialized metabolism of exporting an assimilated form of N was unique to host plants from the *Cucurbitales* or to *Frankia* cluster-2 symbioses in general.

Carbon and N metabolism are directly linked by the tricarboxylic acid (TCA) cycle. This cycle can be closed via several distinct reactions. In the classic TCA cycle, 2-oxoglutarate (2-OG) is the substrate to be converted to succinyl-CoA by the activity of 2-OG dehydrogenase. Succinyl-CoA synthase then converts succinyl-CoA to succinate. An alternative TCA cycle has been described for some bacterial species (Tian *et al.*, 2005), including rhizobia (Green *et al.*, 2000), which is advantageous under reducing conditions as required during N fixation. Here, succinic semialdehyde (SSA) is produced from 2-OG by

2-OG decarboxylase. SSA is then converted to succinate via SSA dehydrogenase (SSA-DH). A third alternative pathway to produce succinate from 2-OG is the γ-aminobutyrate (GABA) shunt (Xiong *et al.*, 2014), which again requires the activity of SSA-DH. Lastly, the TCA cycle can be closed by the glyoxylate shunt, where the production of 2-OG would be avoided altogether (Zhang and Bryant, 2015). We wanted to identify which of the reactions take place to close the TCA cycle in *Frankia* cluster-2. This has not been done before and is important in light of which N metabolite is exported to the host plant.

If not used for maintaining the TCA cycle, 2-OG can be pulled out and used for the assimilation of ammonium in the GS/GOGAT cycle, which leads to the production of glutamate. Glutamate can then be used as a substrate in the biosynthesis of glutamine, asparagine, or arginine. In all root nodule symbioses examined thus far, TCA cycle intermediates are supplied to the microsymbiont by the host plant (Jeong *et al.*, 2004; Udvardi and Poole, 2013). The export of an assimilated form of N, such as arginine, by *Frankia* cluster-2 would require more carbon skeleton input from the host plant than in systems where ammonium is exported directly, such as in *Frankia* cluster-1 symbiosis (Guan *et al.*, 1996). Most, but not all, of these carbon skeletons would be returned by the endosymbiont during the export of assimilated N. Thus, the export of assimilated N by the endosymbiont is not energy efficient for the symbiosis as a whole, as it requires more complex transport processes.

Our study aims to compare features of N and carbon metabolism in nodules of *Frankia* cluster-2 host plants from two different orders: *D. glomerata* representing *Cucurbitales*, and *C. thyrsiflorus* representing *Rosales*. The former was nodulated by *Candidatus* Frankia californiensis Dg2, while the latter was nodulated by the closely related strain *Candidatus* F. californiensis Cv1 (Nguyen *et al.*, 2016, 2019; Normand *et al.*, 2017). As a comparative system, nodules of *A. glutinosa* induced by *Frankia alni* ACN14a (Normand *et al.*, 2007) were included in some of the analyses. For this symbiosis, it is known that ammonium is exported from the bacterium to the host plant (Guan *et al.*, 1996). Our study presents analyses of gene expression levels and enzyme activities related to carbon and N metabolism, as well as protein modelling, which were performed to elucidate the metabolite exchange between host and symbiont.

Materials and methods

Biological material

Ceanothus thyrsiflorus and *Datisca glomerata* plants were grown as previously described by Salgado *et al.* (2018). In brief, *C. thyrsiflorus* plant cuttings were obtained from a local nursery, Cornflower Farms (Elk Grove, CA, USA). The inoculum used, *Candidatus* Frankia californiensis Cv1, originated from nodules of *C. velutinus* plants collected in Sagehen Experimental Forest (Truckee, CA, USA) (Nguyen *et al.*, 2019). Crushed nodules of *C. thyrsiflorus* propagated in the greenhouse were used to inoculate the new plants. The *D. glomerata* plants were germinated from seeds and grown on N-poor soil (S-Jord, Hasselfors Garden AB, Hasselfors, Sweden). Eight weeks after germination, they were transferred to a 1:1 (v/v) soil/sand (1.2–2 mm Quartz; Rådasand AB, Lidköping, Sweden) mixture and inoculated with *Candidatus* F. californiensis Dg2 (Nguyen *et al.*, 2016) from crushed mature *D. glomerata* nodules. *Alnus glutinosa* seeds were obtained from Svenska Skogsplantor (Bålsta, Sweden). Plants were germinated from seeds, which were vernalized for 2 weeks, on N-poor soil. Six-week-old plants were inoculated with *Frankia alni* ACN14a (Normand *et al.*, 2007), which had been grown in basic propionate (BAP) medium without N (Benoist *et al.*, 1992) at 28 °C for 4 weeks. Cells were spun down at 3045 *g* for 10 min and washed twice with sterile milliQ H_2O before use. All plants were grown at a light/dark rhythm of 16 h light/8 h dark, with 26 °C during the light and 19 °C during the dark phase. The greenhouse was kept at 60% relative humidity. For all plants, the nodulation status was confirmed before inoculation to ensure that no plant was already infected. All plants were watered with deionized water twice per week, with one-quarter strength Hoagland's medium with 1 mM KNO_3 once per week before nodulation, and with one-quarter strength Hoagland's medium without N once per week after nodulation (Hoagland and Arnon, 1938). *Frankia coriariae* BMG5.1 (Gtari *et al.*, 2015) was obtained from Leibniz-Institut—Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) and maintained in a modified BAP medium (DSMZ medium 1589, https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium1589.pdf), with the pH adjusted to 9 as recommended by the original publication (Gtari *et al.*, 2015).

RNA isolation, cDNA synthesis, and quantitative reverse transcription–PCR analysis

Mature nodules were collected, flash-frozen in liquid N, and stored at –80 °C until RNA isolation. Nodules were ground in liquid N with a mortar and pestle. RNA was isolated using the Spectrum Plant Total RNA kit (Sigma-Aldrich, Stockholm, Sweden) according to the manufacturer's instructions, except for the addition of Polyclar AT (Serva, Germany) and an additional sonication step as described by Nguyen *et al.* (2019). Genomic DNA was removed using the On-Column DNase I digestion kit (Sigma-Aldrich, Sweden). cDNA was synthesized using the TATAA GrandScript cDNA Synthesis Kit according to the instructions of the manufacturer (TATAA Biocenter, Sweden), except for the reverse transcription-negative control in which the enzyme was omitted. RT–qPCRs were performed using Maxima SYBR Green/ROX (ThermoFisher Scientific) as previously described by Nguyen *et al.* (2019), as follows: initial denaturation at 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C. The melt curve program was 15 s at 95 °C, 15 s at 60 °C, and 15 s at 95 °C. All reactions were performed with four technical replicates and three biological replicates. Primers were designed for sequences conserved between Dg2 and Cv1, while separate primers were designed for ACN14a. Primer efficiency between 95% and 105% was confirmed in all samples. All primer sequences used in this study are given in Supplementary Table S1. The housekeeping gene *infC*, encoding the translation initiation factor IF-3 (Alloisio *et al.*, 2010), and the nitrogenase gene *nifD*, after verifying, were used for normalizing the relative expression data, expressed as ΔCt in calculations and figures. Statistical analysis, which includes Student's *t*-test when comparing Cv1 against Dg2, or one-way ANOVA when comparing ACN14a, Cv1, and Dg2 followed by Tukey post-hoc analysis, and the visualization of the data was performed through RStudio using the Tidyverse package (Wickham *et al.*, 2019; RStudio Team, 2022).

Amino acid profiling

Samples were taken from *C. thyrsiflorus* plants to profile the total amino acid content of nodules, inoculated roots, and uninoculated roots. Samples were treated as described by Persson *et al.* (2016) and extracted as

described by Hacham *et al.* (2002). Profiling analysis was performed at the Molecular Structural Facility, University of California, Davis (CA, USA), on a Hitachi L-8900 Amino Acid Analyzer (Ibaraki, Japan). For statistical analysis, one-way ANOVA was performed on three biological replicates, followed by a Tukey post-hoc analysis. Data were analysed and visualized in RStudio (Wickham *et al.*, 2019; RStudio Team, 2022).

Protein extraction, enzyme assays, and protein modelling and phylogeny

Liquid cultures of *F. alni* ACN14a and *F. coriariae* BMG5.1 were collected via centrifugation (3045 *g* for 10 min) and washed three times in sterile milliQ water. Both culture pellets and flash-frozen nodule samples were ground in liquid N with a mortar and pestle. A 500 mg aliquot of each sample was transferred to 2 ml tubes, and 500 µl of protein isolation buffer was added [100 mM MES, 15% (v/v) ethyleneglycol, 2% (v/v) 2-mercaptoethanol, 1 mM phenylmethylsulfonylfluoride (PMSF)], following an adapted protocol from Berry *et al.* (2004). Bacterial proteins were isolated using the ultrasonic homogenizer Sonoplus HD 2070 (Bandelin Electronic, Berlin, Germany) at 30%, pulsing twice with 60 s each time on ice. Samples were centrifuged briefly to pellet cellular debris, and total soluble protein concentration was determined using BSA as standard (Bradford, 1976), and diluted to the same concentration.

Enzyme activity measurements for the transferase activity of glutamine synthetase (GS) were performed as adapted from Berry *et al.* (2004) and Romanov *et al.* (1998). A 5 µl aliquot of nodule extract, which corresponded to 100 ng of soluble protein, was added to 100 µl of the reaction mixture (20 mM Tris-acetate, 8.75 mM hydroxylamine, 1 mM EDTA, 2.25 mM $MnCl_2$, 17.5 mM NaH_2AsO_4, 2.75 mM ADP, 35 mM glutamine, pH 6.4) and incubated at 30 °C. The reaction was stopped by adding 100 µl of ferric reagent (3.2% w/v $FeCl_3$, 4% w/v trichloroacetic acid, 0.5 N HCl), after 1, 15, 30, 60, and 120 min, respectively. The negative control either did not contain the substrate glutamine, or contained nodule extract that had been boiled for 10 min to denature the proteins completely. To distinguish between heat-labile plant GS activity or bacterial GSII activity, and heat-stable bacterial GSI activity, samples were incubated at 40, 50, or 60 °C for 10 min as the treatments in the literature vary (Edmands *et al.*, 1987; Behrmann *et al.*, 1990). The production of γ-glutamyl hydroxamate was measured in a Hidex Sense microplate reader at 530 nm. The absorbance of the reaction mixture was corrected against the absorbance of the negative controls. One unit of the enzyme was defined as the amount of enzyme that catalyses the formation of 1 µmol γ-glutamyl hydroxamate (Bender *et al.*, 1977). The assay was performed on two technical replicates of two biological replicates.

The enzyme activity assay for SSA-DH was adapted after Tian *et al.* (2005). A 5 µl aliquot of nodule extract, which corresponded to 100 ng of soluble protein, was added to 100 µl of reaction buffer [50 mM HEPES (pH 8.0), 1 mM 2-mercaptoethanol, 8 mM $NADP^+$ or NAD^+]. The absorbance of the extracts at 340 nm was monitored in a Hidex Sense microplate reader for 30 min at 27 °C to allow background activity to cease. Then SSA (Santa Cruz Biotechnology, Germany) was added to a final concentration of 4 mM and the absorbance was monitored for an additional 50 min. In the negative control, no SSA was added. The measurements were corrected against the background absorbance of the buffer with added SSA. The SSA-DH assay was performed on three biological replicates and two technical replicates. The assay was conducted on two independent occasions. All enzyme activity assay data were analysed using one-way ANOVA, followed by Tukey post-hoc analysis, and visualized in RStudio (Wickham *et al.*, 2019; RStudio Team, 2022).

Protein modelling was done through the SWISS-Model portal (https://swissmodel.expasy.org/ visited during September 2021 and March 2022). Visualization of models was done with UCSF Chimera version 1.14 (Goddard *et al.*, 2007; Huang *et al.*, 2014). For phylogenetic analysis, amino acid sequences of *Frankia*, the sequences of the solved crystal structures of 2-OG decarboxylase from *Mycobacterium smegmatis* (reference A0R2B1), and the 2-OG dehydrogenase of *Staphylococcus epidermis* (reference Q5HPC6) were used to run a pBLAST on the NCBI database (March 2022). Sequences were taken from different actinobacterial groups. Sequences were aligned using Clustal Omega on Seaview 5.0 (Gouy *et al.*, 2021), and the N-terminus was trimmed to the predicted first active domain, as sequences varied in length. Re-aligned sequences were then uploaded to the CIPRES science gateway (Miller *et al.*, 2010). The model of substitution was predicted through ModelTest-NG, and a phylogenetic tree was built using RAxML-HPC2 with 100 bootstraps.

Results and discussion

Expression levels of nitrogenase genes allowed for direct comparison of three types of nodules

While the analysis at the transcriptional level has its limitations due to post-transcriptional regulation, it has the advantage of separating plant from bacterial transcription in nodules. Most pathways contain a rate-limiting step catalysed by a key enzyme (Rognstad, 1979). The expression level of the corresponding gene can be the main indicator of the overall activity of the pathway.

Nodules of *A. glutinosa* induced by cluster-1 *F. alni* ACN14a, *C. thyrsiflorus* induced by cluster-2 *Candidatus* Frankia californiensis Cv1, and *D.glomerata* induced by cluster-2 *Candidatus* F. californiensis Dg2 show various anatomical differences (Pawlowski and Demchenko, 2012). It was unclear whether the contribution of N-fixing *Frankia* mRNA in total nodule mRNA was similar enough to allow a direct comparison of bacterial gene expression levels. The expression levels of the structural nitrogenase genes *nifDHK* were compared and analysed against the bacterial housekeeping gene *infC*, encoding the translation initiation factor IF-3 (Alloisio *et al.*, 2010; Nguyen *et al.*, 2019). No significant difference could be found for any of the three genes ($P>0.05$; Supplementary Fig. S1). We concluded that a direct comparison of the different nodule types was appropriate. The expression of *nifD* showed the least variation between biological replicates in all treatments. It was used, together with *infC*, to normalize all further gene expression data.

Frankia cluster-2 assimilates ammonium at similar levels but glutamate has different fates depending on the host plant

To determine if an assimilated N source is exported by *Frankia* cluster-2 to its host plant, in our case *C. thyrsiflorus* and *D. glomerata*, ammonium assimilation activities were investigated by looking at the expression levels of the bacterial genes encoding enzymes of the GS/GOGAT pathway (Fig. 1A). For comparison, the expression levels of *Frankia* genes of the GS/GOGAT pathway were also examined in cluster-1 *Frankia* in nodules of *A. glutinosa*, where ammonia and ammonium are known to be exported to the host (Guan *et al.*, 1996; Alloisio *et al.*, 2010).

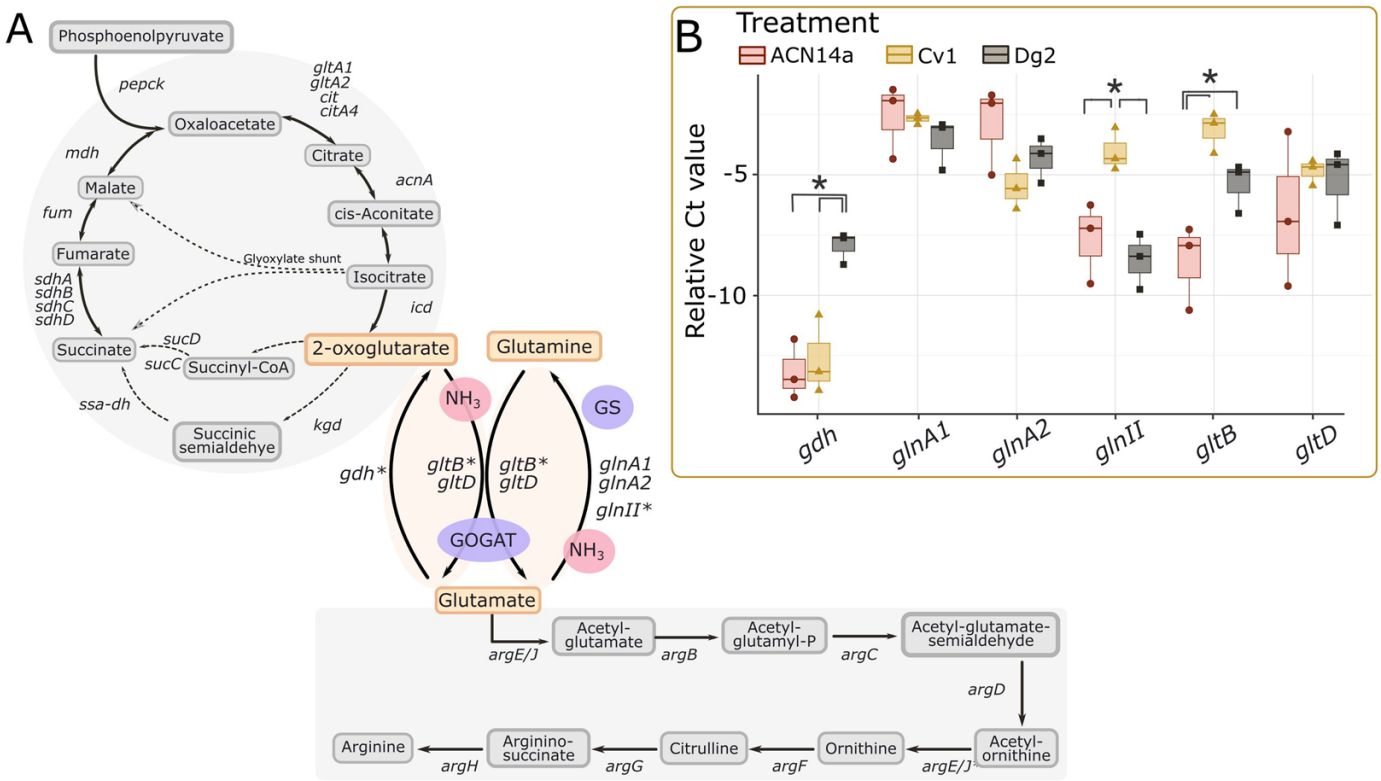

Fig. 1. Relative expression levels (ΔCt value) of genes involved in the GS/GOGAT pathway. (A) Illustration of the GS/GOGAT pathway connected to the TCA cycle and the arginine biosynthesis pathway, given in grey. (B) The gene expression data. The Ct value is normalized against the gene *infC*, encoding the translation initiation factor IF-3 (Alloisio *et al.*, 2010), and the nitrogenase subunit (MoFe protein) gene *nifD*. An asterisk indicates a significant difference (P<0.5), based on one-way ANOVA followed by Tukey post-hoc analysis, of gene expression measured in four technical replicates of three biological replicates of nodules from *Alnus glutinosa* induced by *Frankia alni* ACN14a (red, left), *Ceanothus thyrsiflorus* induced by *Candidatus* Frankia californiensis Cv1 (yellow, centre), and *Datisca glomerata* induced by *Candidatus* Frankia californiensis Dg2 (black, right). Individual data points of the biological repeats are presented. Abbreviations: *gdh*, glutamate dehydrogenase; *glnA1/glnA2*, glutamine synthetase I subunits; *glnII*, glutamine synthetase II; *gltB*, glutamate synthase, large subunit; *gltD*, glutamate synthase, small subunit.

The gene *gltB*, coding for one of the subunits of glutamate synthase, was found to be significantly more highly expressed by *Frankia* in nodules of *C. thyrsiflorus* or *D. glomerata*, respectively, compared with nodules of *A. glutinosa* (Fig. 1B). In addition, the glutamine synthetase gene, *glnII*, was significantly more highly expressed in nodules of *C. thyrsiflorus* compared with nodules of both *D. glomerata* and *A. glutinosa*. On the other hand, the gene *gdh*, encoding glutamate synthase, was expressed at significantly higher levels in nodules of *D. glomerata* compared with *C. thyrsiflorus* and *A. glutinosa* (Fig. 1B). No significant differences could be observed between the expression levels of the genes *glnA1*, *glnA2*, and *gltD* across all three nodule types.

Based on these data, we propose that 2-OG is pulled out of the TCA cycle at similar levels in *Frankia* cluster-2 in nodules, presumably to assimilate ammonium. This does not occur in cluster-1 *F. alni* ACN14a in nodules of *A. glutinosa*, which is in line with previous studies (Guan *et al.*, 1996; Alloisio *et al.*, 2010). However, in *Frankia* cluster-2, the fate of glutamate differs between nodules of *C. thyrsiflorus* and *D. glomerata*: in *C. thyrsiflorus*, more glutamate is converted into glutamine. In nodules of *D. glomerata*, on the other hand, glutamate can support the biosynthesis of arginine, which has been indicated to be the main N export product to the host plant (Berry *et al.*, 2004, 2011; Salgado *et al.*, 2018).

We therefore looked at the expression levels of the arginine biosynthesis genes (Fig. 2) in the two different nodule types induced by *Frankia* cluster-2. This would allow us to see if arginine is synthesized at similar levels in nodules of *C. thyrsiflorus* compared with *D. glomerata*, where it is known to be exported to the host plant. We found that the gene *argE/argJ*, encoding the bifunctional enzyme acetylornithine deacetylase, which catalyses the first step in the biosynthesis pathway (Fig. 2A), was expressed at significantly higher levels in *Frankia* in nodules of *D. glomerata* compared with those of *C. thyrsiflorus* (Fig. 2B). The genes *argB* and *argD* were expressed at significantly higher levels in nodules of *C. thyrsiflorus*. The remaining biosynthesis genes, namely *argC*, *argF*, *argG*, and *argH*, were expressed at similar levels in both nodule types. Xu *et al.* (2020) have demonstrated that the first step, catalysed by ArgE/ArgJ, is rate limiting for arginine biosynthesis. Therefore, our results indicate that more glutamate is

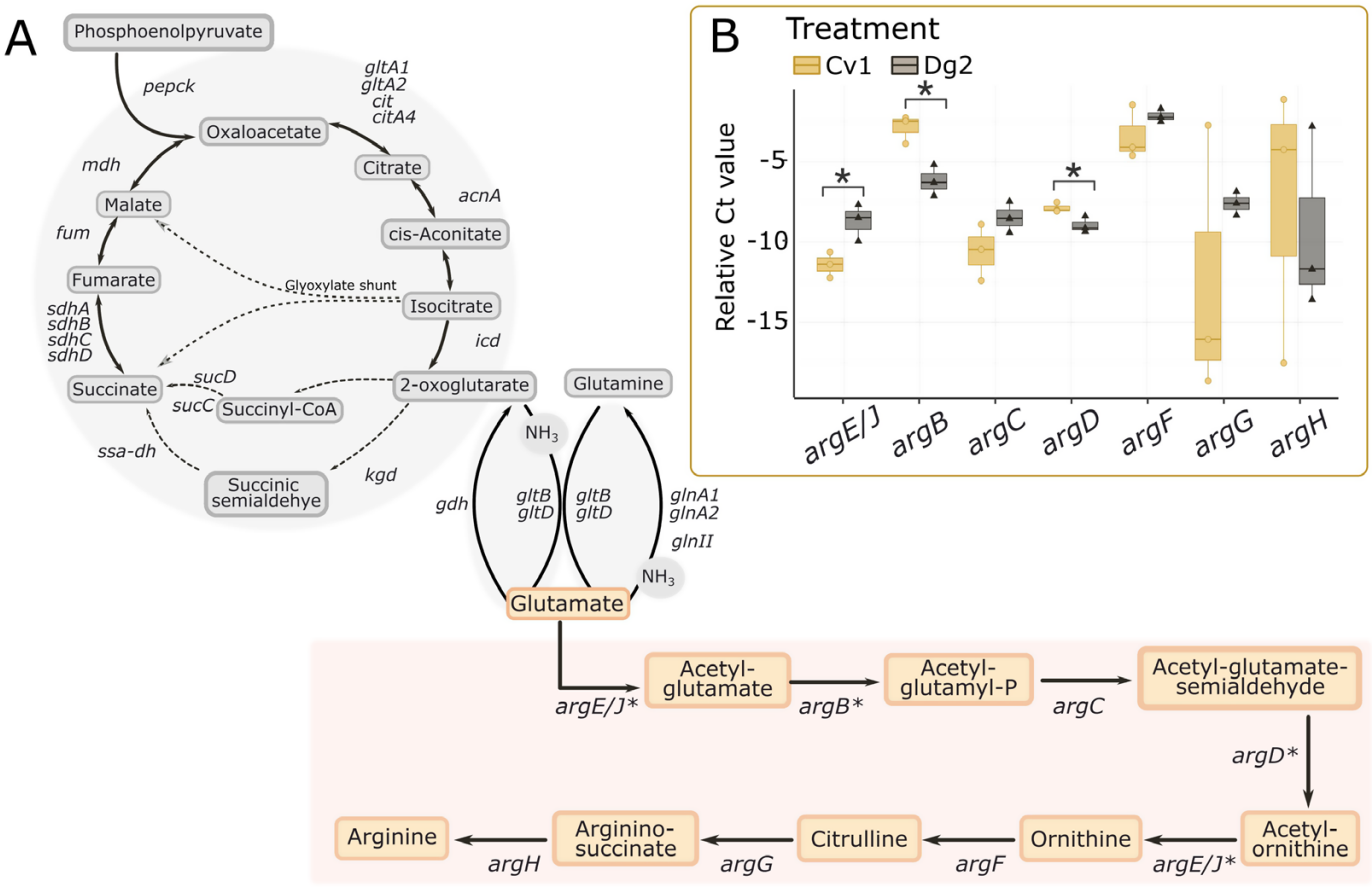

Fig. 2. Relative expression levels (ΔCt value) of genes involved in the arginine biosynthesis pathway. (A) Biosynthesis pathway and connection to the TCA and GS/GOGAT cycle, illustrated in grey. (B) Gene expression levels. The Ct value is normalized against the gene *infC*, encoding the translation initiation factor IF-3 (Alloisio *et al.*, 2010), and the nitrogenase subunit (MoFe protein) gene *nifD*. An asterisk indicates a significant difference (P<0.5), based on Student's *t*-test (Cv1 and Dg2), of gene expression of four technical repeats of three biological repeats of nodules from *Ceanothus thyrsiflorus* induced by *Candidatus* Frankia californiensis Cv1 (yellow, left), and *Datisca glomerata* induced by *Candidatus* Frankia californiensis Dg2 (black, right). Individual data points of the biological repeats are presented. Abbreviations: *argB*, acetylglutamate kinase; *argC*, *N*-acetyl-γ-glutamyl-phosphate reductase; *argD*, acetylornithine/succinyldiaminopimelate aminotransferase; *argE/argJ*, bifunctional gene acetylornithine deacetylase; *argF*, ornithine carbamoyltransferase; *argG*, argininosuccinate synthase; *argH*: argininosuccinate lyase.

used for the production of arginine in *Frankia* in nodules of *D. glomerata* compared with *C. thryrsiflorus*, despite the higher expression of *argB* and *argD* in *C. thyrsiflorus*. This is supported by previous work based on the expression levels of plant genes (Salgado *et al.*, 2018). Based on the gene expression data of *Frankia*, we conclude that arginine cannot be the main export source in nodules of *C. thyrsiflorus*.

Metabolic profiling of *Ceanothus* roots and nodules, and glutamine synthetase enzyme activity

The above findings raise the question: which metabolite is exported from the *Frankia* to the plant in nodules of *C. thyrsiflorus*? To address this question, the levels of different amino acids and N metabolites were compared between nodules, inoculated roots, and uninoculated roots of *C. thyrsiflorus* (Fig. 3). For a comparison, data from a previous study on *D. glomerata* and *A. glutinosa* were included (Persson *et al.*, 2016). Only the top eight most abundant metabolites were included in the analysis.

The results show that, in general, in samples of *C. thyrsiflorus*, the concentrations of all N metabolites were significantly higher in nodules compared with roots (P<0.05; Fig. 3). Consistent with the results on gene expression, and in contrast with the results for nodules of *D. glomerata* (Berry *et al.*, 2004; Persson *et al.*, 2016) and *Coriaria myrtifolia* (*Coriariaceae*, *Cucurbitales*; Wheeler and Bond, 1970), arginine did not accumulate at high levels in *C. thyrsiflorus*. Asparagine and glutamate were the most dominant nitrogenous solutes in nodules. Glutamine could not be detected in uninoculated and inoculated roots, but was found to accumulate in nodules of *C. thyrsiflorus*. This is in agreement with gene expression data for *glnII* (Fig. 1), where the gene was significantly more highly expressed in *Frankia* in nodules of *C. thyrsiflorus* compared with nodules of *A. glutinosa* or *D. glomerata*. It is important to consider that no distinction between accumulation in plant and

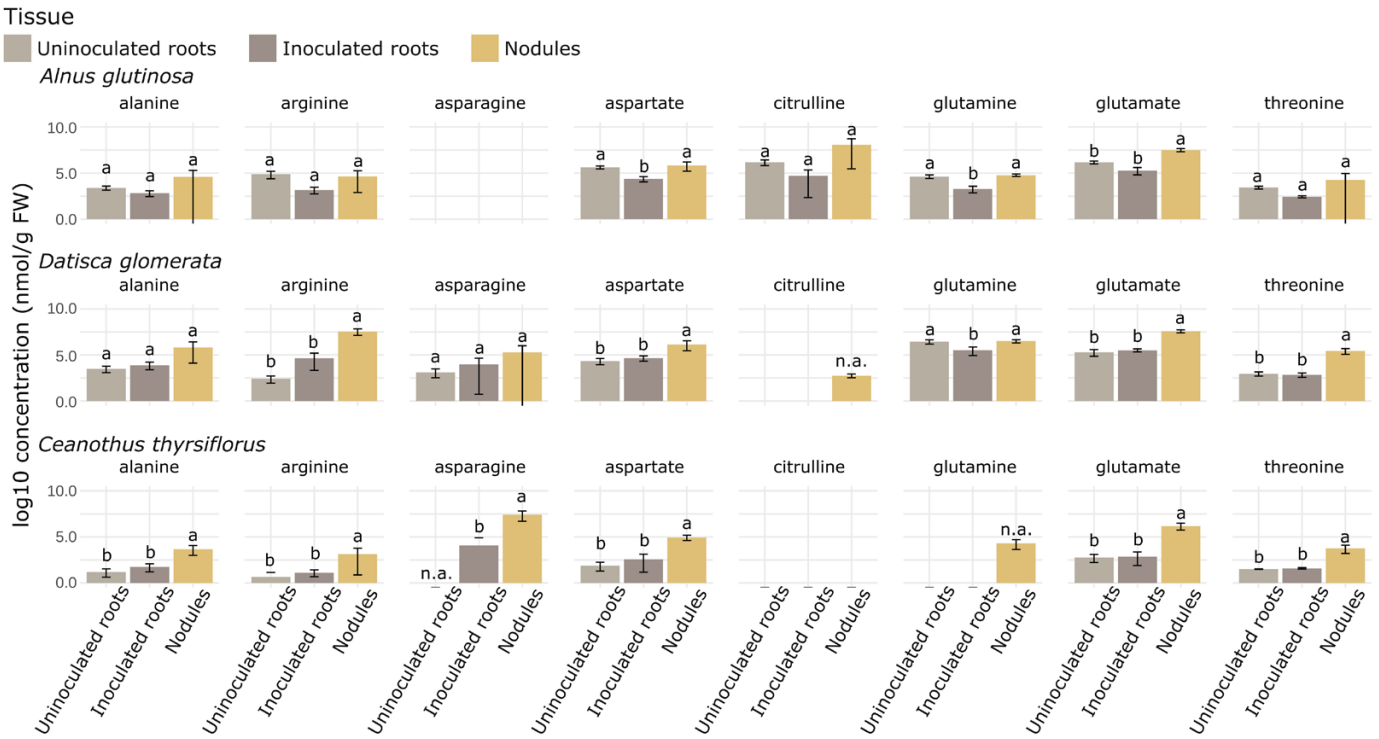

Fig. 3. Concentrations of nitrogen metabolites in *Ceanothus thyrsiflorus* compared with previously published data on *Datisca glomerata* and *Alnus glutinosa* (Persson *et al.*, 2016). Data are presented to compare the concentration (log10 nmol g FW^{-1}) per tissue: uninoculated roots (grey), inoculated roots (brown), and nodules (mustard). Significant differences, based on one-way ANOVA followed by Tukey post-hoc analysis, are indicated with compact letter display.

bacterial cells can be made based on the metabolite data (Fig. 3). Glutamate, for instance, is present at significantly higher levels in nodules than in uninoculated or inoculated roots, in all three nodule types. For nodules of *D. glomerata* and *C. thyrsiflorus*, this could be because glutamate is an intermediate for the production of arginine or glutamine, respectively. In nodules of *A. glutinosa*, it was shown that assimilation of ammonium by *Frankia* occurs only at low levels, but the plant assimilates ammonium via the GS/GOGAT pathway, followed by synthesis of citrulline, which is then transported via the xylem (Guan *et al.*, 1996; Alloisio *et al.*, 2010).

Given the importance of post-transcriptional regulation, we wanted to conduct an enzyme activity assay of GS to verify the results of the expression analysis and the metabolic profiling. As *Frankia* cluster-2 cannot be cultivated *in vitro*, and we were interested in activity under symbiotic conditions, enzyme activities needed to be conducted on nodule material. Unlike plants, certain prokaryotes such as rhizobia and *Frankia* (Edmands *et al.*, 1987; de Bruijn *et al.*, 1989) have been shown to contain two variants of GS: a heat-stable GSI, encoded by *glnA*, and a heat-labile GSII, encoded by *glnII* (Huss-Danell, 1997). Aside from their different heat sensitivity, the two variants also differ in their regulation. GlnA is controlled post-translationally by reversible adenylylation, whereas *glnII* is under transcriptional control (de Bruijn *et al.*, 1989). Eukaryotes, such as plants, commonly only contain the heat-labile variant of GS (GSII). This allows for the distinction between plant and bacterial activity of GS.

In nodules of *A. glutinosa*, we found significant activities of GS in the crude extract, but not after exposure to heat treatment at 40, 50, or 60 °C. (Fig. 4). This indicates that, here, only the heat-labile GS is responsible for GS activity, most probably from plant origin. This would be in agreement with our gene expression data (Fig. 1), as well as with the metabolite data (Fig. 3) (Persson *et al.*, 2016), if we assume most fixed N is exported to, and assimilated by, the plant. These results are supported by previous work, showing that GlnII could not be detected immunologically in nodules of *Alnus incana*, while it could be detected in N-fixing cultures of the infective *Frankia* (Lundquist and Huss-Danell, 1992), and gene expression data on nodules of *A. glutinosa* (Guan *et al.*, 1996). In nodules of *D. glomerata*, we found the highest activity in the crude extract. After heat exposure at 60 °C, GS activity was significantly reduced to the levels of the negative control. Activity was not statistically reduced compared with the crude extract after exposure to 40 °C or 50 °C, but there was also no significant difference from the negative control. This would indicate that both a heat-stable and a heat-labile copy of GS are active in nodules of *D. glomerata*. In nodules of *C. thyrsiflorus*, we found significant GS activity in the crude extract. However, this activity was not significantly reduced after heat treatment. This would indicate that most, if not all, nodule GS activity is

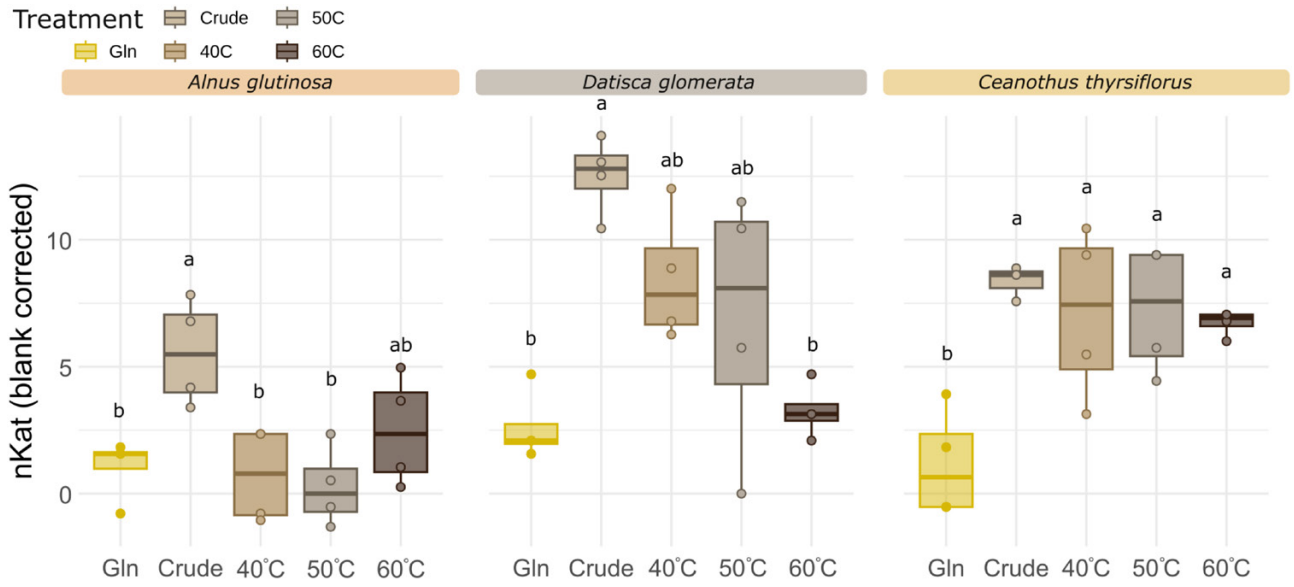

Fig 4. GS transferase activity assay in nodules of *Alnus glutinosa*, *Datisca glomerata*, and *Ceanothus thyrsiflorus*. Nodules of *A. glutinosa* were induced by *Frankia alni* ACN14a, of *D. glomerata* induced by *Candidatus* Frankia californiensis Dg2, and of *C. thyrsiflorus* by *Candidatus* F. californiensis Cv1. Activity measurements are based on the amount of enzyme required to produce 1 µmol of γ-glutamyl hydroxamate (nKat). The activity was measured in the negative control in crude protein extract without glutamine (red), and otherwise either using crude protein extracts (yellow), or in protein extract exposed for 10 min to 40, 50, or 60 °C (brown, light grey, and dark grey), to distinguish between *Frankia* GSI and plant GS/*Frankia* GSII activity. The absorbance at 530 nm was corrected against the background absorbance of total denatured protein (boiled for 10 min at 95 °C). The assay was conducted on two technical replicates of two biological replicates. The individual data points are given. Different letters indicate significant differences ($P<0.05$), based on one-way ANOVA followed by Tukey post-hoc analysis.

performed by a heat-stable version of GS, namely by *Frankia* GSI.

Another N metabolite which accumulated in nodules of *C. thyrsiflorus* was asparagine. The biosynthesis of asparagine, as well as aspartate, relies on a high input of oxaloacetate, an intermediate of the TCA cycle (Fig. 5B). We investigated the gene expression levels of phosphoenolpyruvate carboxykinase (*pepck*), as well of the genes encoding the bidirectional malate dehydrogenase (*mdh*), and citrate synthase (*gltA1/A2* and *citA/citA4*) (Fig. 5A). We found that while *pepck* was not expressed at significantly higher levels in *Frankia* in nodules of either *D. glomerata* or *C. thyrsiflorus*, the genes *mdh*, *gltA2*, and *citA* were expressed at significantly higher levels in *Frankia* in nodules of *C. thyrsiflorus.* Potentially, this could indicate an accumulation of oxaloacetate, which would be shuttled into asparagine biosynthesis. This would be supported by the metabolite data (Fig. 3), where asparagine accumulated in nodules, and could not be detected in uninoculated roots.

The amino acid interconversions of glutamine–glutamate and asparagine–aspartate can occur rapidly and play important roles in central metabolism. Analysis of the full transcriptome might provide some more insights; however, enzyme activity assays would be more compelling. These would have to be conducted under symbiotic conditions and, unlike the heat-liable and heat-stable forms of GS, such assays could not distinguish between plant and bacterial activity. Altogether, our results suggest that asparagine and/or glutamine could be exported by *Frankia* cluster-2 in nodules of *Ceanothus* spp. These metabolites might also play a role as N transport forms in the xylem. Taking together our gene expression, metabolite, and enzyme activity data, our results suggest that in *C. thyrsiflorus* glutamine is the most likely metabolite to be exported by symbiotic *Frankia* to the host plant. As the *Frankia* strains under investigation here, *Candidatus* F. californiensis Cv1 and *Candidatus* F. californiensis Dg2, belong to the same species, this would indicate that the export of assimilated N is a common feature for cluster-2 *Frankia*, but the host plant determines which metabolite is exported.

Gene loss of isocitrate lyase indicates lack of the glyoxylate shunt in *Frankia* cluster-2

The export of assimilated N, whether in the form of arginine or asparagine, would require a high input of carbon skeletons. Carboxylates provided by the host plant can be used to provide energy or carbon skeletons for ammonium assimilation, both by being shuttled into the TCA cycle. Genome analyses, facilitated through BLAST, revealed that several genes of TCA cycle-related enzymes were lacking in *Frankia* cluster-2 (indicated by red crosses in Fig. 5).

Firstly, the gene *icl*, encoding isocitrate lyase which is responsible for catalysing the first step of the glyoxylate shunt (Fig. 5), was lacking in all *Frankia* cluster-2 genomes available (Supplementary Table S2; Supplementary Fig. S2). Further

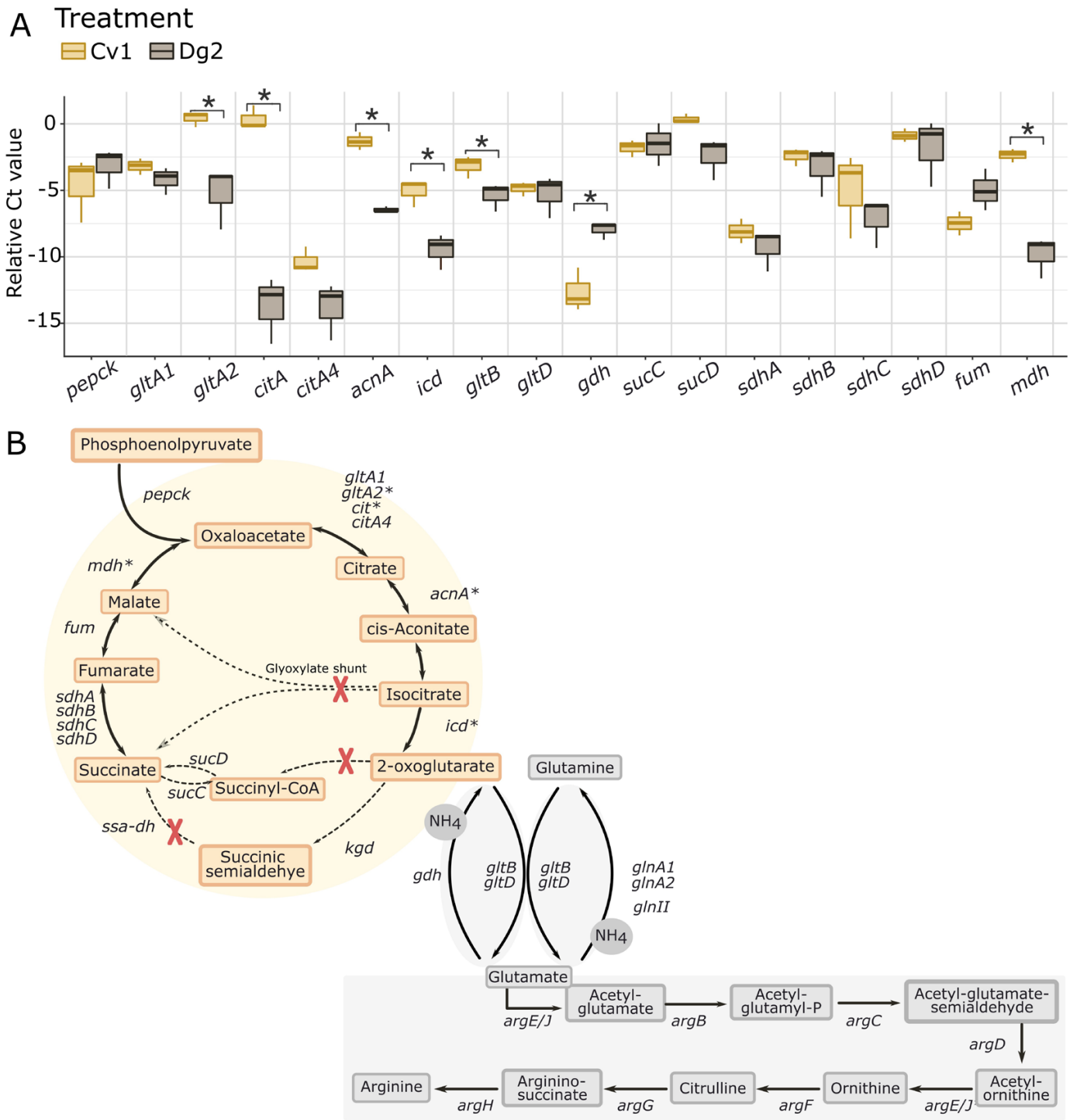

Fig. 5. Relative gene expression (ΔCt value) of genes encoding enzymes involved in the TCA cycle. The Ct value is normalized against the gene *infC*, encoding the translation initiation factor IF-3 (Alloisio *et al.*, 2010), and the nitrogenase subunit (MoFe protein) gene *nifD*. An asterisk indicates a significant difference (P<0.5), based on Student's *t*-test (Cv1 and Dg2), of gene expression of four technical repeats of three biological repeats of nodules from *Ceanothus thyrsiflorus* induced by *Candidatus* Frankia californiensis Cv1 (yellow, left), and *Datisca glomerata* induced by *Candidatus* F. californiensis Dg2 (black, right). Individual data points of each biological repeat are presented. The pathway with enzymes interacting is illustrated below. Abbreviations used: *pepck*, phosphoenolpyruvate carboxykinase; *gltA/gltA2*, citrate synthase; *citA/citA4*, citrate synthase; *acnA*, aconitate hydratase A; *icd*, isocitrate dehydrogenase; *gltB*, glutamate synthase, large subunit; *gltD*, glutamate synthase, small subunit; *gdh*, glutamate dehydrogenase; *sucC/sucD*, succinate-CoA ligase subunit alpha/beta; *sdhA/shdB/sdhC/sdhD*, succinate dehydrogenase complex subunit A/B/C/D; *fum*, fumarate hydratase; *mdh*: malate dehydrogenase.

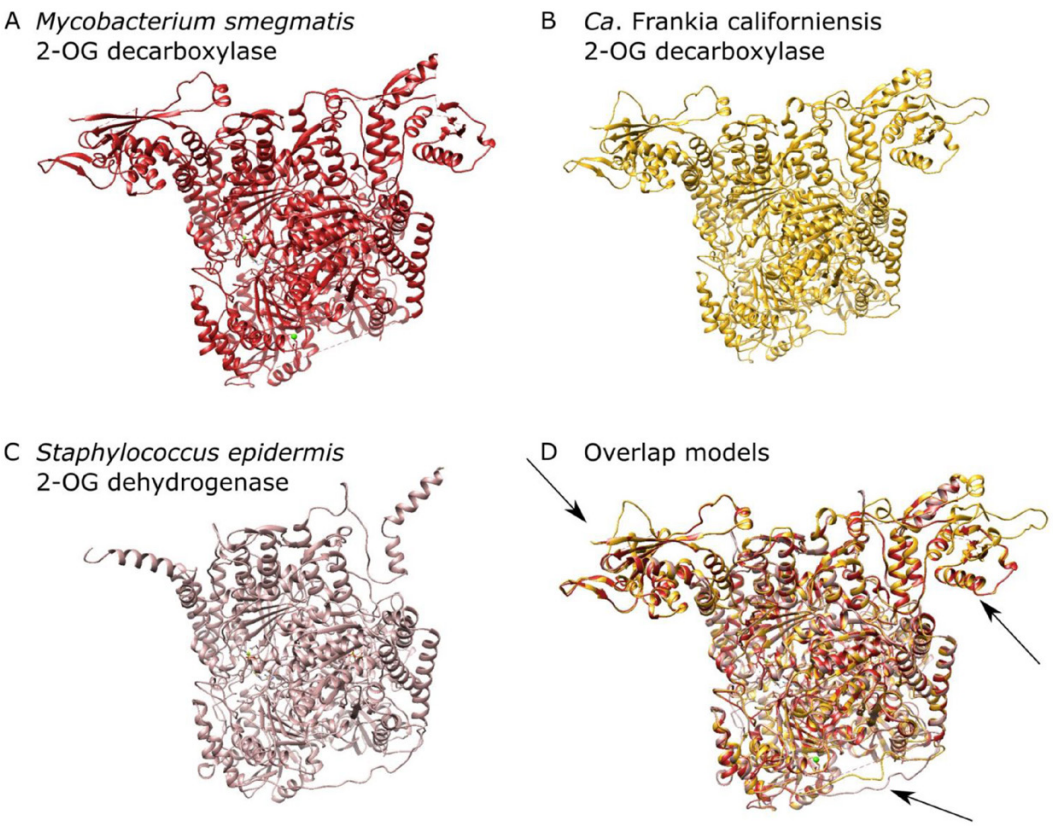

Fig. 6. Protein modelling of 2-oxoglutarate decarboxylase. (A) The solved crystal structure of *Mycobacterium smegmatis* (red, reference A0R2B1). (B) The model of *Candidatus* Frankia californiensis Dg2 (yellow, global model quality estimate QMEANDisCo >0.70). (C) The solved crystal structure of 2-oxoglutarate dehydrogenase of *Staphylococcus epidermis* (rosy brown, reference Q5HPC6). (D) Overlay of three different models, with arrows indicating the major differences between the 2-OG decarboxylase and 2-OG dehydrogenase.

analysis of *Frankia* genomes revealed this gene to be present in most cluster-1 genomes, but lacking in most cluster-3 and cluster-4 genomes. Instead, using the sequence from *F. alni* ACN14a (Genbank accession CAJ61004.1) led to the identification of methylisocitrate lyase as the closest homologue. The corresponding enzyme has been shown to be unable to catalyse the same reaction (Dunn *et al.*, 2009). For *Candidatus* F. meridionalis, no significant hit for the *icl* sequence could be found. This suggests that the production 2-OG cannot be avoided in most *Frankia* strains, except for most cluster-1 strains.

All *Frankia* strains exhibit a variant TCA cycle

2-OG can have different fates as a metabolite and signalling molecule, as reviewed by Huergo and Dixon (2015). Within the TCA cycle, 2-OG will be converted to succinate. This can happen through one of three pathways, as explained before: by 2-OG dehydrogenase, 2-OG decarboxylase, or through the GABA shunt (Green *et al.*, 2000; Tian *et al.*, 2005; Xiong *et al.*, 2014). The latter two pathways require the activity of SSA-DH. The 2-OG dehydrogenase complex is composed of three enzymes: the E1 2-OG dehydrogenase, the E2 dihydrolipoyl succinyltransferase, and the E3 dihydrolipoyl dehydrogenase.

Frankia cluster-2 genomes have previously been shown to contain a high abundance of transposable elements (Nguyen *et al.*, 2019). We therefore continued to utilize BLAST to identify genes present in *Frankia* genomes. Using the corresponding amino acid sequence of *Mycobacteriodes abscessus* and *Staphylococcus epidermis*, the closest homologue of E1 in all *Frankia* genomes examined was found to be a multifunctional 2-OG decarboxylase/oxoglutarate dehydrogenase with <45% amino acid sequence identity with the query sequences (accession numbers listed in Supplementary Table S2). To fully understand if the *Frankia* gene encodes E1 2-OG dehydrogenase or 2-OG decarboxylase, homology modelling of the amino acid sequence of the *Frankia* proteins was performed on the SWISS Model portal (Fig. 6; Supplementary Fig. S3). It was compared with the solved crystal structures of 2-OG decarboxylase from *M. smegmatis* and 2-OG dehydrogenase of *S. epidermis*. The best homology model for all the *Frankia* models was found to be 2-OG decarboxylase (global model quality estimate QMEANDisCo >0.70) while it showed much less similarity to 2-OG dehydrogenase (QMEANDisCo <0.50). The model was highly conserved for all *Frankia* genomes examined, which included all declared species of the different clusters (Fig. 6; Supplementary Fig. S3). In addition, a

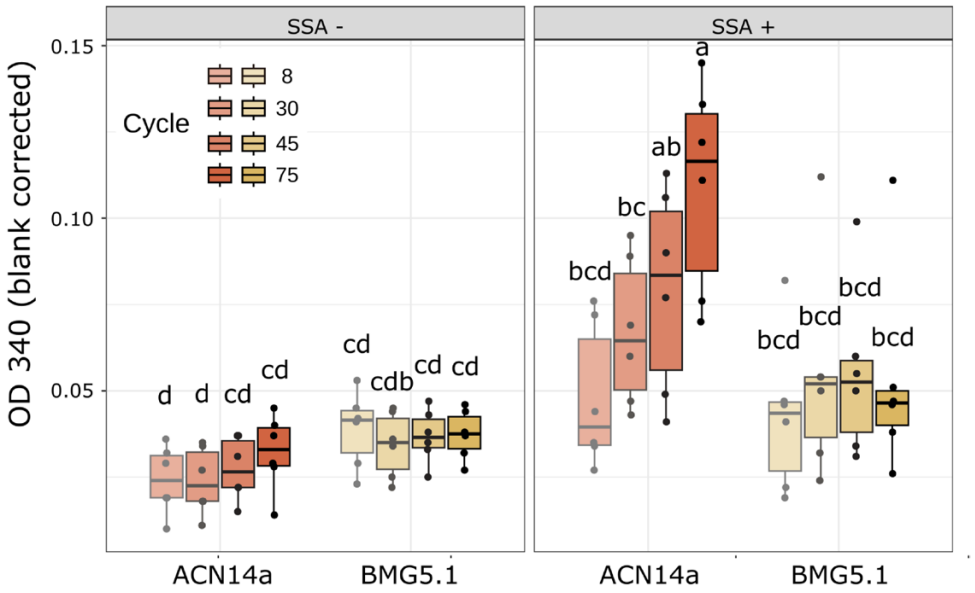

Fig. 7. Enzyme activity assay for succinic semialdehyde dehydrogenase (SSA-DH) in cluster-1 *Frankia alni* ACN14a compared with cluster-2 *Frankia coriariae* BMG5.1. Absorbance was measured at 340 nm, to detect the production of NADPH from $NADP^+$; absorbance was blank corrected. Protein extracts were allowed to acclimate for 30 min in the reaction buffer with $NADP^+$ but without substrate, after which SSA or solvent only was injected. The absorbance was measured for an additional 50 min. Time is represented in cycles, with increasing darkness (red for ACN14a and yellow for BMG5.1) indicating that more time has passed. Boxplot indicates the absorbance at the beginning, before injection, after injection, and at the end of the experiment, based on two technical replicates of three biological replicates. The compact letters display indicates significant differences based on three-way ANOVA (sample, treatment, cycle), followed by Tukey post-hoc analysis.

phylogenetic tree was built based on the amino acid sequences, which separated the 2-OG dehydrogenase sequences from the 2-OG decarboxylase sequences (Supplementary Fig. S4). Taken together, we conclude that all *Frankia* genomes available contain a 2-OG decarboxylase gene instead of a 2-OG dehydrogenase. This would imply that the variant TCA cycle is active in which succinate is synthesized from 2-OG via SSA and not succinyl-CoA (Fig. 5), as has been shown for some other bacteria. Interestingly, this pathway has been suggested to be distributed widely among rhizobia as an adaptation to the microaerobic conditions for catabolizing dicarboxylic acids (Green *et al.*, 2000). We suggest this adaptation to have evolved in *Frankia* as well.

The production of succinate is thus dependent on the activity of SSA-DH. However, the gene encoding this enzyme could not be identified in *Frankia* cluster-2 genomes. The $NADP^+$-dependent variant was found to be common in genomes of strains from other *Frankia* clusters, whereas the NAD^+-dependent variant was found only in some of these genomes (Supplementary Table S2). Using a pBLAST query with the corresponding sequence from *F. alni* ACN14a, the closest homologue in cluster-2 genomes was found to be a protein belonging to the aldehyde dehydrogenase family (listed in Supplementary Table S2). Yet, this dehydrogenase had <45% amino acid sequence identity with the $NADP^+$-dependent SSA-DH in *Frankia* genomes of other clusters. The SSA-DH could also not be identified in the proteome data available for *F. coriariae* BMG5.1 (Ktari *et al.*, 2017), the only *Frankia* cluster-2 species which has been cultivated to date.

The lack of strong homologues of $NADP^+$-dependent or NAD^+-dependent SSA-DH alone cannot eliminate the possibility for another enzyme to catalyse the reaction. To determine if any activity was present in *Frankia* cluster-2 strains, an SSA-DH activity assay as described by Tian *et al.* (2005) was conducted for *F. coriariae* BMG5.1. *F. alni* ACN14a was used as a positive control because its genome contains genes encoding $NADP^+$-dependent as well as NAD^+-dependent SSA-DH (Supplementary Table S2). The absorbance significantly increased after the addition of SSA in *F. alni* ACN14a (Fig. 7) in a buffer containing $NADP^+$, indicating that the $NADP^+$ variant was active. However, a successful assay could not be established for the NAD^+ variant. In extracts of *F. coriariae* BMG5.1, no activity could be measured after the addition of SSA, in the presence of either $NADP^+$ or NAD^+, thus confirming the lack of SSA-DH activity (Fig. 7). *F. coriariae* BMG5.1 grows considerably more slowly than any other cultivable *Frankia* strain. The medium recommended by the DSMZ (medium 1589) contains sodium succinate. This indicates that external supplementation with an intermediate from the reductive part of the TCA cycle is required for this strain to show sufficient growth. The strain was maintained in medium without external succinate for at least 3 weeks before the activity assay to ensure that SSA-DH activity was not down-regulated.

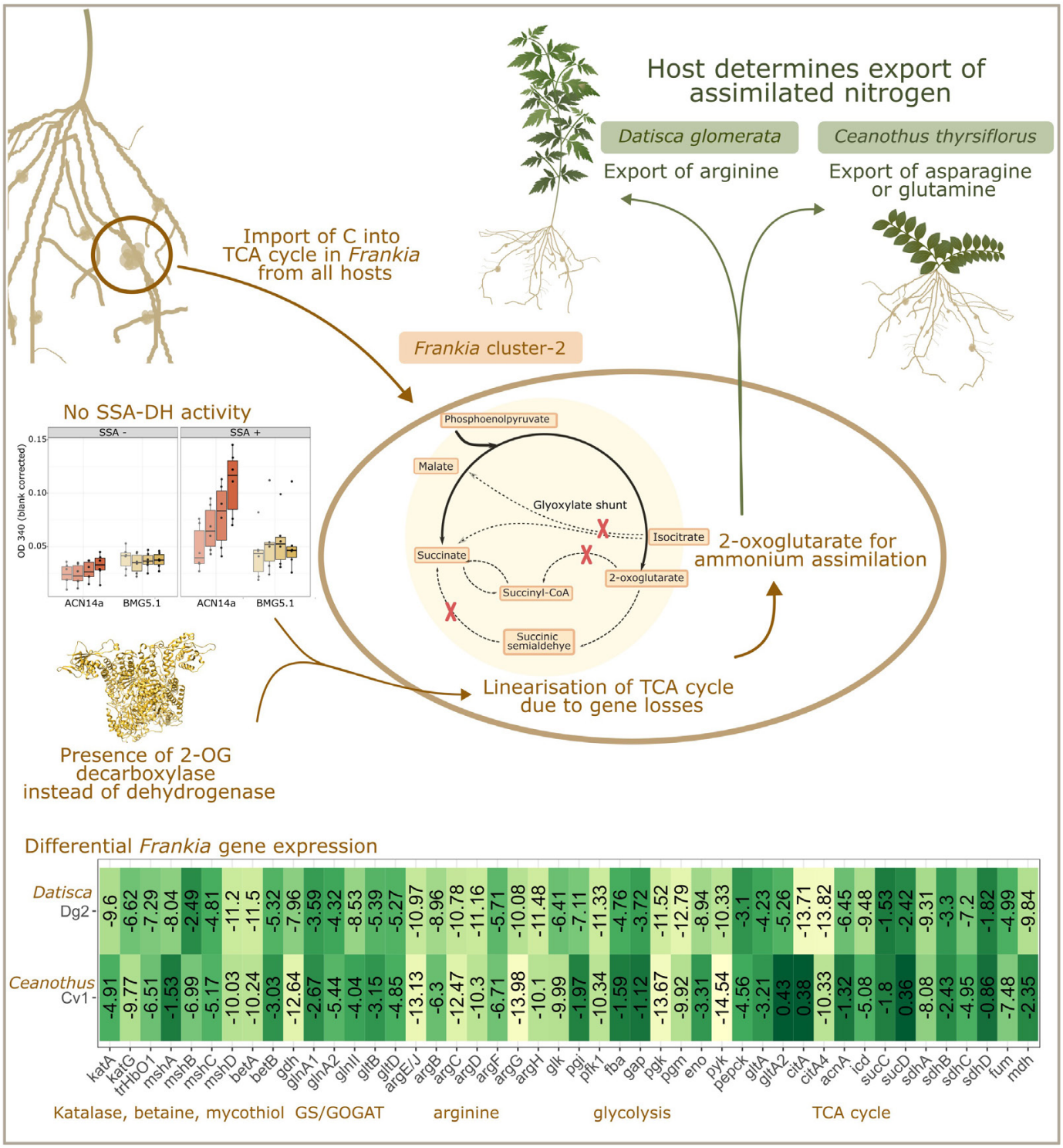

Fig. 8. Summary of the data presented. The host plant provides *Frankia* with carbon sources which are shuttled into the TCA cycle. Several gene losses were identified, resulting in a linearization of the cycle. As a result, the produced 2-oxoglutarate is used for ammonium assimilation. The export of an assimilated nitrogen source is dependent on the host plant. A heatmap represents all *Frankia* gene expression analysed, including genes which were not discussed in the manuscript. Gene expression levels are indicated in light (lower) to dark (higher) green.

Apart from acting as a precursor of succinate, 2-OG can be used for the synthesis of glutamate via the GS/GOGAT cycle, and support ammonium assimilation. As shown above, our results suggest that the export of an assimilated form of N from the host is a common feature in nodules induced by *Frankia* cluster-2 strains. This would explain the loss of the genes for the glyoxylate shunt and SSA-DH observed in *Frankia* cluster-2, as in symbiosis ammonium assimilation would require the majority of 2-OG to be used for ammonium assimilation, leading to glutamate biosynthesis. Most reactions of the TCA cycle are reversible, except for the step catalysed by isocitrate dehydrogenase (encoded by *icd*). The carboxylate provided by the plant as carbon source, which thus far has only been identified for *A. glutinosa* where it represents malate (Jeong *et al.*, 2004), could be converted into any compound on the reductive side of the TCA cycle (Fig. 5), and thus into fumarate and succinate by the reverse reactions of succinic dehydrogenase; and into fumarate, as well as into oxaloacetate, by malate dehydrogenase. Under symbiotic conditions in nodules, the high input of presumably malate to the reductive side of the TCA cycle would allow 2-OG to be drawn out of the oxidative side of the TCA cycle at a high rate. Hence, the TCA cycle would not be required to work as a cycle, but only as a linear pathway.

We hypothesized that *Frankia* cluster-2 strains have this carbon metabolism: the TCA cycle acts as a linear pathway instead of a cycle to keep up with the constant removal of 2-OG for ammonium assimilation (Fig. 8). Gene losses within the TCA cycle are not unique to *Frankia*, as shown by the symbiotic cyanobacteria UCYN-A and *Trichormus azollae* (previously *Nostoc azollae*), which even lack the entire TCA cycle (Ran *et al.*, 2010; Tripp *et al.*, 2010), as well as some rhizobial mutants (Green *et al.*, 2000; Schulte *et al.*, 2021). *Frankia* cluster-2 strains represent the earliest divergent symbiotic clade within the *Frankia* genus (Sen *et al.*, 2014; Persson *et al.*, 2015). Their metabolism might be based on an ancient form of symbiotic metabolite exchange. The export of an assimilated form of N is not energy efficient: it requires a much larger supply of carbon skeletons from the host to the endosymbiont. In contrast to ammonia, which can leak through the bacterial membrane and be converted to ammonium in the acidic perisymbiont space, assimilated forms of N would have to be transported across the bacterial membrane at the expense of energy, which has to be provided by the host (Fig. 8).

The divergence between the symbiotic cluster-2 and the non-symbiotic cluster-4 took place early in the evolution of *Frankia*, followed by the divergence of the precursor of cluster-1 and cluster-3 from cluster-4 (Sen *et al.*, 2014; Berckx *et al.*, 2024). *Frankia* cluster-1 strains export ammonia or ammonium to the host, as opposed to assimilated N (Guan *et al.*, 1996). This indicates that in different symbiotic clusters, the preferred export N form evolved differently. The *de facto* linear version of the TCA cycle from the provided carbon source to 2-OG might have led to the loss of genes involved in the production of succinate from 2-OG. The loss of the SSA-DH gene in *Frankia* cluster-2 might have prevented the evolution of a more energy-efficient nutrient exchange system. Given their adaptation of the TCA cycle to the metabolite exchange in symbiosis, it is not surprising that cluster-2 stains have such a low saprotrophic potential and are rarely found in the soil in the absence of a host plant (Persson *et al.*, 2015; Battenberg *et al.*, 2017).

Conclusions

Based on the data presented in this study, the export of an assimilated form of N by *Frankia* cluster-2 strains in symbiosis is a common feature of the clade. In *Cucurbitales* host plants, such as *D. glomerata* and *C. myrtifolia*, this export form is arginine, while in *Rosales*, such as *C. thyrsiflorus*, it is asparagine or glutamine. The assimilation of fixed N for export during symbiosis puts a high demand on 2-OG. The TCA cycle therefore seems to work linearly from the carbon source(s) provided by the host to 2-OG. Due to this special metabolism, the need for the glyoxylate shunt is obviated, as well as the production of succinate from 2-OG. This led to gene losses which can explain the low saprotrophic potential of *Frankia* cluster-2 strains.

Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Gene expression data of *nifDHK* genes.

Fig. S2. ML tree of icocitrate lyase proteins.

Fig. S3. Protein models of 2-OG decarboxylase in different *Frankia* species.

Fig. S4. RAxML phylogeny for 2-oxoglutarate dehydrogenase and 2-oxoglutarate decarboxylase.

Table S1. Primer sequences used for all RT–qPCRs.

Table S2. The accession numbers in NCBI BLAST searchers of *icl*, E1 component of 2-OG DH, and SSA-DH for *M. abscessus*, *S. epidermis*, and *Frankia*.

Acknowledgements

We would like to thank Anna Pettersson and Ingela Lundwall (Stockholm University) for taking care of the plants in the greenhouse, and Rachel Foster and Edouard Pesquet (Stockholm University) for helpful discussions.

Author contributions

FB, TVN, and KP: conceptualization; FB, RH, and DW: formal analysis; FB, TVN, and KB: investigation; FB: visualization; FB and TVN: writing—original draft; AB and KP: supervision; JK, AB and KP: funding; FB, TVN, RH, DW, KB, JK, AB and KP: writing—review & editing.

Conflict of interest

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

Funding

This project was supported by grants from the Swedish Research Council Vetenskapsradet (VR2012-03061 and VR2019-05540 to KP) and the Carl Tryggers Foundation (CTS 9:925 to KP). AMB's research received support from the U.S. Department of Agriculture (USDA-NIFA-CA-D-PLS-2173-H); KB was supported in part by UC Davis Graduate Research Fellowship in Plant Sciences. The bioinformatics support by the BMBF-funded project 'Bielefeld-Giessen Center for Microbial Bioinformatics-BiGi (grant no. 031A533)' within the German Network for Bioinformatics Infrastructure (deNBI.de) is gratefully acknowledged.

Data availability

Protein modelling files are available online on the Swedish National Data service, found through following https://doi.org/10.5878/wxkv-s592. The raw data and the R code are available upon request to FB.

References

Alloisio N, Queiroux C, Fournier P, Pujic P, Normand P, Vallenet D, Médigue C, Yamaura M, Kakoi K, Kucho K. 2010. The *Frankia alni* symbiotic transcriptome. Molecular Plant-Microbe Interactions **23**, 593–607.

Battenberg K, Wren JA, Hillman J, Edwards J, Huang L, Berry AM. 2017. The influence of the host plant is the major ecological determinant of the presence of nitrogen-fixing root nodule symbiont cluster II *Frankia* species in soil. Applied and Environmental Microbiology **83**, e02661-16.

Behrmann I, Hillemann D, Pühler A, Strauch E, Wohlleben W. 1990. Overexpression of a *Streptomyces viridochromogenes* gene (*glnII*) encoding a glutamine synthetase similar to those of eucaryotes confers resistance against the antibiotic phosphinothricyl-alanyl-alanine. Journal of Bacteriology **172**, 5326–5334.

Bender RA, Janssen KA, Resnick AD, Blumenberg M, Foor F, Magasanik B. 1977. Biochemical parameters of glutamine synthetase from *Klebsiella aerogenes*. Journal of Bacteriology **129**, 1001–1009.

Benoist P, Müller A, Diem HG, Schwencke J. 1992. High-molecular-mass multicatalytic proteinase complexes produced by the nitrogen-fixing actinomycete *Frankia* strain BR. Journal of Bacteriology **174**, 1495–1504.

Berckx F, Nguyen TV, Bandong CM, *et al*. 2022. A tale of two lineages: how the strains of the earliest divergent symbiotic *Frankia* clade spread over the world. BMC Genomics **23**, 602.

Berckx F, Wibberg D, Brachmann A, Morrison C, Obaid NB, Blom J, Kalinowski J, Wall LG, Pawlowski K. 2024. Genome analysis and biogeographic distribution of the earliest divergent *Frankia* clade in the southern hemisphere. FEMS Microbiology Ecology **100**, fiae042.

Berry AM, Mendoza-Herrera A, Guo Y-Y, *et al*. 2011. New perspectives on nodule nitrogen assimilation in actinorhizal symbioses. Functional Plant Biology **38**, 645–652.

Berry AM, Murphy TM, Okubara PA, Jacobsen KR, Swensen SM, Pawlowski K. 2004. Novel expression pattern of cytosolic Gln synthetase in nitrogen-fixing root nodules of the actinorhizal host, *Datisca glomerata*. Plant Physiology **135**, 1849–1862.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. Analytical Biochemistry **72**, 248–254.

de Bruijn FJ, Rossbach S, Schneider M, Ratet P, Messmer S, Szeto WW, Ausubel FM, Schell J. 1989. *Rhizobium meliloti* 1021 has three differentially regulated loci involved in glutamine biosynthesis, none of which is essential for symbiotic nitrogen fixation. Journal of Bacteriology **171**, 1673–1682.

Dunn MF, Ramírez-Trujillo JA, Hernández-Lucas I. 2009. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. Microbiology (Reading, England) **155**, 3166–3175.

Edmands J, Noridge NA, Benson DR. 1987. The actinorhizal root-nodule symbiont *Frankia* sp. strain CpI1 has two glutamine synthetases. Proceedings of the National Academy of Sciences, USA **84**, 6126–6130.

Gifford I, Vance S, Nguyen G, Berry AM. 2019. A stable genetic transformation system and implications of the type IV restriction system in the nitrogen-fixing plant endosymbiont *Frankia alni* ACN14a. Frontiers in Microbiology **10**, 2230.

Goddard TD, Huang CC, Ferrin TE. 2007. Visualizing density maps with UCSF Chimera. Journal of Structural Biology **157**, 281–287.

Gouy M, Tannier E, Comte N, Parsons DP. 2021. Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. Methods in Molecular Biology **2231**, 241–260.

Green LS, Li Y, Emerich DW, Bergersen FJ, Day DA. 2000. Catabolism of alpha-ketoglutarate by a *sucA* mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. Journal of Bacteriology **182**, 2838–2844.

Griesmann M, Chang Y, Liu X, *et al*. S2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science **361**, eaat1743.

Gtari M, Ghodhbane-Gtari F, Nouioui I, *et al*. 2015. Cultivating the uncultured: growing the recalcitrant cluster-2 *Frankia* strains. Scientific Reports **5**, 13112.

Guan C, Ribeiro A, Akkermans ADL, Jing Y, van Kammen A, Bisseling T, Pawlowski K. 1996. Nitrogen metabolism in actinorhizal nodules of *Alnus glutinosa*: expression of glutamine synthetase and acetylornithine transaminase. Plant Molecular Biology **32**, 1177–1184.

Gueddou A, Swanson E, Hezbri K, *et al*. 2019. Draft genome sequence of the symbiotic *Frankia* sp. strain BMG5.30 isolated from root nodules of *Coriaria myrtifolia* in Tunisia. Antonie Van Leeuwenhoek **112**, 67–74.

Hacham Y, Avraham T, Amir R. 2002. The N-terminal region of *Arabidopsis* cystathionine gamma-synthase plays an important regulatory role in methionine metabolism. Plant Physiology **128**, 454–462.

Hay A-E, Herrera-Belaroussi A, Rey M, Fournier P, Normand P, Boubakri H. 2020. Feedback regulation of N fixation in *Frankia–Alnus* symbiosis through amino acids profiling in field and greenhouse nodules. Molecular Plant-Microbe Interactions **33**, 499–508.

Hoagland DR, Arnon DI. 1938. The water-culture method for growing plants without soil. California Agricultural Experiment Station Circular No. 347.

Huang CC, Meng EC, Morris JH, Pettersen EF, Ferrin TE. 2014. Enhancing UCSF Chimera through web services. Nucleic Acids Research **42**, W478–W484.

Huergo LF, Dixon R. 2015. The emergence of 2-oxoglutarate as a master regulator metabolite. Microbiology and Molecular Biology Reviews **79**, 419–435.

Huss-Danell K. 1997. Actinorhizal symbioses and their N_2 fixation. New Phytologist **136**, 375–405.

Jeong J, Suh S, Guan C, Tsay Y-F, Moran N, Oh CJ, An CS, Demchenko KN, Pawlowski K, Lee Y. 2004. A nodule-specific dicarboxylate transporter from alder is a member of the peptide transporter family. Plant Physiology **134**, 969–978.

Ktari A, Gueddou A, Nouioui I, Miotello G, Sarkar I, Ghodhbane-Gtari F, Sen A, Armengaud J, Gtari M. 2017. Host plant compatibility shapes the proteogenome of *Frankia coriariae*. Frontiers in Microbiology **8**, 720.

Kucho KI, Tobita H, Utsumi S, Uchiumi T, Yamanaka T. 2022. Biology of actinorhizal symbiosis from genomics to ecology: the 20th International Meeting on Frankia and Actinorhizal Plants. Journal of Forest Research **27**, 96–99.

Lundquist P-O, Huss-Danell K. 1992. Immunological studies of glutamine synthetase in *Frankia–Alnus incana* symbioses. FEMS Microbiology Letters **91**, 141–146.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop, 1–8.

Nguyen TV, Wibberg D, Battenberg K, Blom J, Vanden Heuvel B, Berry AM, Kalinowski J, Pawlowski K. 2016. An assemblage of *Frankia* Cluster II strains from California contains the canonical *nod* genes and also the sulfotransferase gene *nodH*. BMC Genomics **17**, 796.

Nguyen TV, Wibberg D, Vigil-Stenman T, *et al*. 2019. *Frankia*-enriched metagenomes from the earliest diverging symbiotic *Frankia* cluster: they come in teams. Genome Biology and Evolution **11**, 2273–2291.

Normand P, Lapierre P, Tisa LS, *et al*. 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Research **17**, 7–15.

Normand P, Nguyen TV, Battenberg K, Berry AM, Heuvel BV, Fernandez MP, Pawlowski K. 2017. Proposal of '*Candidatus* Frankia californiensis', the uncultured symbiont in nitrogen-fixing root nodules of a phylogenetically broad group of hosts endemic to western North America. International Journal of Systematic and Evolutionary Microbiology **67**, 3706–3715.

Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, Evtushenko L, Misra AK. 1996. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. International Journal of Systematic Bacteriology **46**, 1–9.

Nouioui I, Ghodhbane-Gtari F, Rohde M, Klenk H-P, Gtari M. 2017. *Frankia coriariae* sp. Nov., an infective and effective microsymbiont isolated from *Coriaria japonica*. International Journal of Systematic and Evolutionary Microbiology **67**, 1266–1270.

Pawlowski K, Demchenko KN. 2012. The diversity of actinorhizal symbiosis. Protoplasma **249**, 967–979.

Persson T, Battenberg K, Demina IV, *et al*. 2015. *Candidatus* Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical *nod* genes *nodABC* in symbiosis with its host plant. PLoS One **10**, e0127630.

Persson T, Van Nguyen T, Alloisio N, Pujic P, Berry AM, Normand P, Pawlowski K. 2016. The N-metabolites of roots and actinorhizal nodules from *Alnus glutinosa* and *Datisca glomerata*: can *D. glomerata* change N-transport forms when nodulated? Symbiosis **70**, 149–157.

Pesce C, Oshone R, Hurst SG, Kleiner VA, Tisa LS. 2019. Stable transformation of the actinobacteria *Frankia* spp. Applied and Environmental Microbiology **85**, e00957–e00919.

Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K, Zheng W-W, Lapidus A, Lowry S, Haselkorn R, Bergman B. 2010. Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. PLoS One **5**, e11486.

Rognstad R. 1979. Rate-limiting steps in metabolic pathways. Journal of Biological Chemistry **254**, 1875–1878.

Romanov VI, Gordon AJ, Minchin FR, Witty JF, Skøt L, James CL, Tikhonovich IA. 1998. Physiological and biochemical characteristics of FN1, a 'fixation impaired' mutant of pea (*Pisum sativum* L.). Journal of Experimental Botany **49**, 1789–1796.

RStudio Team. 2022. RStudio: integrated development environment for R. RStudio, PBC. http://www.rstudio.com/

Salgado MG, van Velzen R, Nguyen TV, Battenberg K, Berry AM, Lundin D, Pawlowski K. 2018. Comparative analysis of the nodule transcriptomes of *Ceanothus thyrsiflorus* (Rhamnaceae, Rosales) and *Datisca glomerata* (Datiscaceae, Cucurbitales). Frontiers in Plant Science **9**, 1629.

Schulte CCM, Borah K, Wheatley RM, *et al*. 2021. Metabolic control of nitrogen fixation in rhizobium–legume symbioses. Science Advances **7**, eabh2433.

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P. 2014. Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov. International Journal of Systematic and Evolutionary Microbiology **64**, 3821–3832.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proceedings of the National Academy of Sciences, USA **92**, 2647–2651.

Tian J, Bryk R, Itoh M, Suematsu M, Nathan C. 2005. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of alpha-ketoglutarate decarboxylase. Proceedings of the National Academy of Sciences, USA **102**, 10670–10675.

Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, Affourtit JP, Zehr JP. 2010. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature **464**, 90–94.

Udvardi M, Poole PS. 2013. Transport and metabolism in legume–rhizobia symbioses. Annual Review of Plant Biology **64**, 781–805.

van Velzen R, Doyle JJ, Geurts R. 2019. A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. Trends in Plant Science **24**, 49–57.

van Velzen R, Holmer R, Bu F, *et al*. 2018. Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proceedings of the National Academy of Sciences, USA **115**, E4700–E4709.

Wheeler CT, Bond G. 1970. The amino acids of non-legume root nodules. Phytochemistry **9**, 705–708.

Wickham H, Averick M, Bryan J, *et al*. 2019. Welcome to the Tidyverse. Journal of Open Source Software **4**, 1686.

Xiong W, Brune D, Vermaas WFJ. 2014. The γ-aminobutyric acid shunt contributes to closing the tricarboxylic acid cycle in *Synechocystis* sp. PCC 6803. Molecular Microbiology **93**, 786–796.

Xu M, Tang M, Chen J, Yang T, Zhang X, Shao M, Xu Z, Rao Z. 2020. PII signal transduction protein GlnK alleviates feedback inhibition of N-acetyl-L-glutamate kinase by L-arginine in *Corynebacterium glutamicum*. Applied and Environmental Microbiology **86**, e00039-20.

Zhang S, Bryant DA. 2015. Biochemical validation of the glyoxylate cycle in the Cyanobacterium *Chlorogloeopsis fritschii* strain PCC 9212. Journal of Biological Chemistry **290**, 14019–14030.