

# Proteomic analysis dissects the impact of nodulation and biological nitrogen fixation on *Vicia faba* root nodule physiology

Beate Thal, Hans-Peter Braun, and Holger Eubel\*

Institut für Pflanzengenetik, Leibniz Universität Hannover, Herrenhäuser Str. 2, 30419

Hannover, Germany

\* To whom correspondence should be addressed: Holger Eubel, Institute of Plant Genetics, Leibniz Universität Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany, Tel.: +495117622699, Fax: +4951176214351, Mail: [heubel@genetik.uni-hannover.de](mailto:heubel@genetik.uni-hannover.de)

Running title

Cellular metabolism of root nodules

Keywords

Symbiotic nitrogen fixation; nodulation; cellular metabolism; *Vicia faba*; *Rhizobium leguminosarum*; shotgun mass spectrometry

Acknowledgements

We thank Prof. Dr. Ursula B. Priefer and Prof. Dr. Allan Downie for providing the *Rhizobium leguminosarum* strains Fix<sup>+</sup> and Fix<sup>-</sup>. Our research was supported by the Deutsche Forschungsgemeinschaft (DFG) in the frame of the Graduiertenkolleg *Signaling at the plant Soil interface* (GRK1798).

## 1. Abstract

Root nodules are specialized plant organs housing and regulating the mutual symbiosis of legumes with nitrogen fixing rhizobia. As such, these organs fulfill unique functions in plant metabolism. Identifying the proteins required for the metabolic reactions of nitrogen fixation and those merely involved in sustaining the rhizobia:plant symbiosis, is a challenging task and requires an experimental setup which allows to differentiate between these two physiological processes. Here, quantitative proteome analyses of nitrogen fixing and non-nitrogen fixing nodules as well as fertilized and non-fertilized roots were performed using *Vicia faba* and *Rhizobium leguminosarum*. Pairwise comparisons revealed altered enzyme abundance between active and inactive nodules. Similarly, general differences between nodules and root tissue were observed. Together, these results allow distinguishing the proteins directly involved in nitrogen fixation from those related to nodulation. Further observations relate to the control of nodulation by hormones and provide supportive evidence for the previously reported correlation of nitrogen and sulfur fixation in these plant organs. Additionally, data on altered protein abundance relating to alanine metabolism imply that this amino acid may be exported from *V. faba* root nodules in addition to ammonia. Data are available via ProteomeXchange with identifier PXD008548.

## Supporting Information

Suppl. Table 1: Proteins changing significantly between lines 0 and N

Suppl. Table 2: Proteins changing significantly between lines Fix<sup>+</sup> and N

Suppl. Table 3: Proteins changing significantly between lines Fix<sup>+</sup> and Fix<sup>-</sup>

Suppl. Table 4: Proteins changing significantly between lines 0 and Fix<sup>-</sup>

Suppl. Table 5: Top five proteins derived from pairwise comparisons of Fix<sup>+</sup>/Fix<sup>-</sup>, Fix<sup>+</sup>/N, Fix<sup>-</sup>/0, and N/0

Suppl. Figure 1: Testing of protein sequence databases for cross species protein identification

Suppl. Figure 2: Altered cellular processes in Fix<sup>+</sup> compared to Fix<sup>-</sup> and N

Suppl. Figure 3: Pairwise comparisons and meta-comparisons between the four sample groups

Suppl. Figure 4: Mean abundance of core plant and bacteroid SNF enzymes

Suppl. Figure 5: Response of seven proteins commonly found in MC1 and MC2 to SNF activity (Fix<sup>+</sup>/Fix<sup>-</sup>) as well as nodulation and SNF activity (Fix<sup>+</sup>/N)

## 2. Introduction

Biological nitrogen fixation within the *Fabaceae* family is of considerable significance for both natural ecosystems and agriculture. Legume plants including peas, lentils, soybean, peanuts, and other legume crops contribute 27% of the world's food and forage production (Graham 2003). Fixation of molecular nitrogen in legumes is based upon the symbiosis with bacteria of the genus *Rhizobium*. It enables plants to grow and to produce protein rich seeds without the need for external application of nitrogen, making legumes a valuable 'green' fertilizer. For this and their positive impact on soil quality legumes are often used in crop rotation practices (Biederbeck *et al.* 1998).

The onset of the legume:rhizobia symbiosis is a complex process initiated by Nod factor signaling events between the host plant and its bacterial partner. Bacteria initially infect the plant roots and trigger the formation of new root organs, the nodules (Oldroyd *et al.* 2008). Within nodules, the symbiotic bacteria (now termed bacteroids) reduce the atmospheric dinitrogen (N<sub>2</sub>) by the enzyme nitrogenase to produce ammonia. The latter is then provided to the plant in exchange for organic nutrients (Hoffmann *et al.* 2014). The plant is thus required to invest carbohydrates from photosynthesis to support initiation and maintenance of symbiosis, nodule growth and senescence, production of leghemoglobin (a protein regulating the oxygen concentration of the N<sub>2</sub>-fixing cell), and potentially the assimilation of the reduced nitrogen into organic compounds for the export from the nodule. Exchange of metabolites between the plant cell and the bacteroids is facilitated by the peribacteroid membrane surrounding the bacteroids. It contains a range of different transporters, including an ammonium transport protein (Tyerman *et al.* 1995). Ammonia is the primary product of SNF and also seems to be the main form in which nitrogen is exported from bacteroids (Li *et al.* 2002, Ludwig *et al.* 2004, Kumar *et al.* 2005, Udvardi and Poole 2013). In addition, amino acids such as alanine, aspartate and glutamate may also contribute to varying degrees towards the nitrogen export. Plants inoculated with *R. leguminosarum* strains mutated in broad-specificity Aap and Bra amino acid transporters are growth retarded and suffer from nitrogen starvation (Lodwig *et al.* 2003), thus implying the need for amino acids to be moved between bacteroids and plants. Prell *et al.* (2009) later reported that *R. leguminosarum* bacteroids are auxotrophic for branched chain amino acids (LIV) and depend on the import of plant-produced LIV. Mutations in Aap and Bra therefore most likely do not directly induce nitrogen starvation in the plants (by preventing organic nitrogen being supplied to the plant) but indirectly by negatively affecting bacterial metabolism and nodule development. This, however, does not exclude that, besides ammonium, organic nitrogen is supplied to the plant cell by bacteroids. Indeed, alanine production in isolated bacteroids was reported but the rate of its synthesis varies with the density of the isolated bacteroids in the experimental setup (Appels and Haaker 1991, Waters *et al.* 1998, Allaway *et al.* 2000, Day *et al.* 2001). It is produced from SNF-derived ammonia and pyruvate by the action of bacteroids alanine dehydrogenase (AlaDH) in pea nodules and can be transported across the peribacteroid membrane (Rosendahl *et al.* 1992, Allaway *et al.* 2000). Plants inoculated with a *R. leguminosarum* AlaDH knock-out mutant accumulate 20% less dry weight within six weeks when compared to the wild-type strain. Since ammonia-production of both strains is similar and only the alanine production is hampered, the higher plant biomass in the wild-type may be the direct result of alanine export from the bacteroids (Allaway *et al.* 2000, Day *et al.* 2001). Interestingly, over-expression of AlaDH in *R. leguminosarum* also reduced the biomass of inoculated plants (Lodwig *et al.* 2004). It is therefore possible that bacteroid-produced alanine plays a role in delivering nitrogen in reduced form to the plant but the extent of its contribution is currently unclear as is the fate of alanine after its import into the plant cell.

Considering the profound impact SNF has on plants and their bacterial symbionts, it is not surprising that the accompanying alterations in root tissue morphology and metabolism are driven to a large extent by changes in the abundance of enzymes. Proteomics has been used intensively to characterize the transformation of bacteria into symbiotic, temporary organelles of root nodule cells (Guerreiro *et al.* 1997; Djordjevic 2004; Delmotte *et al.* 2010; Koch *et al.* 2010; Oldroyd *et al.* 2011). Likewise, the protein content of the symbiosome membrane has been studied to identify the transporters mediating symbiotic metabolite crosstalk (Udvardi *et al.* 1997; Wienkoop and Saalbach 2003; Clarke *et al.* 2015). The protein content of nodules and roots grown under different conditions have also been analyzed to gain insight into the special physiology of nodules (Bestel-Corre *et al.* 2002; Larrainzar *et al.* 2007; van Noorden *et al.* 2007; Gil-Quintana *et al.* 2015; Marx *et al.* 2016). Finally, the impact of nodulation on soybean mitochondria as central compartments of non-photosynthetic cellular energy metabolism was studied by Hoa *et al.* (2004). Given that root nodules constitute a plant organ on their own, which is different to the root tissue it originated from, it is often difficult to tell if altered protein abundance is tissue specific or the consequence of nitrogen fixation. In the past, research has mainly focused on the comparison of either two different tissues (roots and nodules) or nitrogen supply conditions (sufficient/insufficient) and is thus of limited use for discriminating between SNF and nodulation related proteins. The pairwise comparisons and the meta-analyses between fertilized and non-fertilized roots as well as nitrogen fixing and non-nitrogen fixing nodules, as presented in this study, provides the means to assign altered protein abundance to either process more precisely. Given that the comparisons conducted here were performed on tissues in steady states, analyses of protein abundance will provide a broad overview of the alterations in the physiology of tissue types as well as nitrogen status. It should, however, be acknowledged that altered enzymatic activities cannot be detected by this approach.

### 3. Materials and Methods

#### 3.1 Plant material and plant cultivation conditions

*Vicia faba* (var. 'Fuego') seeds were sterilized in 70% [v/v] EtOH (for 2 min), washed in ddH<sub>2</sub>O for 10 min, and incubated in NaClO (for 10 min) before being soaked in ddH<sub>2</sub>O for 24 h in the dark. Plants were then transferred to pots (12 cm ID) filled with broken foamed clay (Lamstedt Lecaton, 2 – 4 mm, Lamstedt, Germany). The Lecaton substrate lacks any nutrients and can easily be removed from root material at harvest. Plants were grown for two weeks under an 8 h light/ 16 h dark regime at 20° C and fertilized with 0.5x Hoagland nutrient solution without nitrogen (1 mM MgSO<sub>4</sub>, 40 μM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NaFeEDTA, 0.1 mM CaCl<sub>2</sub>, 0.2 μM Na<sub>2</sub>MoO<sub>4</sub>, 10 μM H<sub>2</sub>BO<sub>3</sub>, 0.2 μM NiSO<sub>4</sub>, 1 μM ZnSO<sub>4</sub>, 2 μM MnCl<sub>2</sub>, 0.5 μM CuSO<sub>4</sub>, 0.2 μM CoCl<sub>2</sub>, Hoagland and Arnon 1950) twice a week. In addition, plants were watered with demineralized water according to demand. After two weeks of growth, bacteria were added to Fix<sup>+</sup> (*Rhizobium leguminosarum* bv. *viciae*) and Fix<sup>-</sup> (*Rhizobium leguminosarum* mutant deficient in nitrogen fixation, NifH78::Tn5, Ma *et al.* 1982) plant groups. No bacteria were added to the plants in groups N and 0. Only N plants were fertilized after two weeks with 0.5x Hoagland (without 0.1 mM CaCl<sub>2</sub> but including 2.5 mM Ca(NO<sub>3</sub>)<sub>2</sub>, and 2.5 mM KNO<sub>3</sub>). Group 0 plants were neither inoculated with bacteria nor fertilized with NO<sub>3</sub><sup>-</sup>. Harvest of nodules (Fix<sup>+</sup>, Fix<sup>-</sup>) and roots (N, 0) took place after four weeks of growth. Three biological replicates (n=3) were performed for all analyses being conducted with this plant material with the exception of nitrogenase enzyme assay (n=1).

### 3.2 Element analysis

Leaf material from three plants of each group was dried at 37° C and ground in a cryo mill (Retsch, Haan, Germany). The carbon and nitrogen content of the powder was determined in an element analyzer by dry oxidation at 1500° C and excess O<sub>2</sub> (Vario EL, Elementar, Hanau, Germany).

### 3.3 Nitrogenase activity

Nitrogenase activity was assayed by detecting the molecular hydrogen (H<sub>2</sub>) being generated as a side product of the reaction. Root systems of four weeks old plants were detached from shoots and placed in a 50 ml Falcon tubes. The lid contained a chemical hydrogen sensor (Euro-Gas, Devon, United Kingdom) attached to a micro-computer. Hydrogen accumulation in the tube was monitored for 300 s.

### 3.4 Sample preparation, liquid chromatography-coupled tandem MS analyses and data processing

Proteins of root and nodules of three biological replicates for each group were extracted in SDS-containing buffer and proteins were concentrated on a SDS gel, before in-gel trypsin digestion. Prior to shotgun tandem MS analyses the resulting peptides were separated chromatographically on a reverse-phase column. Identification of peptides selected for MS/MS spectra was achieved by the Andromeda search engine integrated into the MaxQuant computational proteomics platform (comparative analyses between the sample groups) or by Mascot (Matrix Science, London, UK) search employing the proteome Discoverer platform (Thermo Fisher Scientific, Dreieich, Germany). Minimum requirement for a positive protein identification was a single peptide with a p-value ≤ 0.01. For a comprehensive protein profiling the data were analyzed with the Perseus software. Details of the procedure are given below. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD008548.

#### 3.4.1 Protein extraction and MS sample preparation

Five grams of plant material were shock frozen in liquid nitrogen and ground in a cryo mill (Retsch, Haan, Germany) to fine powder before proteins were extracted with SDS buffer (4% [w/v] SDS; 125 mM Tris-HCl, pH 6.8; 20% [v/v] glycerol) for 5 min at 60° C. After centrifugation (10 min at 18.000 g) the protein concentration of the supernatant was determined using the Pierce BCA-200 Protein Assay Kit (Thermo Fisher Scientific, Dreieich, Germany). For purification and concentration of the proteins, a volume corresponding to 50 µg of protein was mixed with β-mercaptoethanol and bromophenol blue to yield final concentrations of 5% [v/v] and 0.05% [w/v], respectively). Samples were loaded on a glycine / sodium dodecyl sulfate (SDS) polyacrylamide gel consisting of a 4% [v/v] acrylamide stacking and a 14% [v/v] acrylamide separating gel. The run was stopped as soon as the running front reached the border between the two gel phases. Gels were subsequently incubated in fixation solution (15% [v/v] ethanol; 10% [v/v] acetic acid) for 30 min and stained for 1 h with Coomassie Brilliant Blue G250 (Neuhoff *et al.* 1985). The protein band at the interface between both gel phases was cut into cubes with edge lengths of approximately 1 mm, transferred into low-binding Eppendorf tubes (Eppendorf, Wesseling-Berzdorf, Germany) and dried in a vacuum centrifuge. Proteins were in-gel Trypsin digested and resulting peptides were extracted according to Fromm *et al.* (2016).

#### 3.4.2 Label-free quantitative shotgun mass spectrometry and protein identification

Tandem mass spectrometry (MS/MS) analyses were performed using a Q-Exactive (Thermo Fisher Scientific, Dreieich, Germany) mass spectrometer coupled to an Ultimate 3000 (Thermo Fisher Scientific, Dreieich, Germany) UPLC as described in Fromm *et al.* (2016). In brief, 2 to 5 µl of sample solution were injected onto a 2 cm, C18, 5 µm, 100 Å reverse phase trapping column

(Acclaim PepMap100, Thermo Fisher Scientific, Dreieich, Germany) at a rate of 4  $\mu\text{l min}^{-1}$ . Peptide separation was achieved on a 50 cm, C18, 3  $\mu\text{m}$ , 100  $\text{\AA}$  reverse phase analytical column (Acclaim PepMap100, Thermo Fisher Scientific, Dreieich, Germany). Peptides were eluted by a non-linear 5% to 30% [v/v] acetonitrile gradient in 0.1% [v/v] formic acid at a flow rate of 300  $\text{nl min}^{-1}$  and over a period of 60 min and at 35° C. Transfer of eluted peptides into the mass spectrometer was performed using a NSI source (Thermo Fisher Scientific, Dreieich, Germany) equipped with stainless steel nano-bore emitters (Thermo Fisher Scientific, Dreieich, Germany). Spray voltage was set to 2.2 kV, capillary temperature to 275° C, and S-lens RF level to 50%. The MS was run in positive ion mode, MS/MS spectra (top 10) were recorded from 20 min to 100 min. Full MS scans were performed at a resolution of 70,000 whereas 17,500 was used for MS/MS scans. Automatic gain control (AGC) targets for MS and MS/MS were set to 1E6 and 1E5, respectively. Only peptides with 2, 3, or 4 positive charges were considered.

### 3.4.3 Data processing

For assessing the best protein identification strategy, data were analyzed using the Proteome Discoverer software (Thermo Fisher Scientific, Dreieich, Germany) utilizing the Mascot (Matrix Science, London, UK) search engine. Raw files were queried against three different databases: i) NCBIInr, ii) SwissProt, iii) an in-house database of integrated *Medicago truncatula* and *Rhizobium leguminosarum* proteins including common contaminations (Trypsin, BSA, Keratin) using the following parameters for spectrum selection: Precursor selection: Use MS1 precursor; Lower and upper RT limit: 0; First and last scan: 0; Lowest charge state: 1; Highest charge state: 5, Min. precursor mass: 350 kDa, Max. precursor mass: 5000 kDa; Total intensity threshold: 0; Minimum peak count: 1; Mass analyzer: is FTMS; MS order: is MS2; activation t6ype, is HCD; Min. collision energy: 0; Max. collision energy: 1000; Scan type: is full; Ionization source: is nanospray; Polarity mode: is +; S/N threshold: 1.5; unrecognized charges replacement: 1, 2, 3, 4, 5; Unrecognized mass analyzer replacement: FTMS; Unrecognized MS order replacement: MS2; Unrecognized activation type replacement: HCD; Unrecognized polarity replacement: +. Mascot Parameters were as follows: Enzyme name: Trypsin; Maximum missed cleavage sites: 1; Instrument: default; Taxonomy: all entries; Precursor tolerance: 10 ppm; Fragment mass tolerance: 0.8 Da; Use average precursor mass: false; Dynamic modifications: acetyl (N-term), oxidation (M); Static modifications: carbamidomethyl (C). The Target Decoy Validator was set to a false discovery rate (FDR) of 0.01. Only peptides matching this requirement were considered for protein identification.

For quantitative analyses of the samples in the four groups MS/MS spectra were evaluated using the MaxQuant software package (version 1.5.5.1) including the Andromeda search engine (Cox and Mann 2008) against an in-house database of combined *M. truncatula* and *R.m leguminosarum* protein sequences which also included common contaminants. Variable modifications for the Andromeda search were oxidation of methionine residues and N-terminal acetylations. Carbamidomethylation of cysteine residues was selected as fixed modification. The false discovery rate (based on at the peptide spectrum match level, determined by target decoy approach, was set to 0.01. To increase the identification rate of peptides for which no MS/MS data were recorded in some runs but not in others, information on precursor masses and retention times of successfully identified peptides were allowed to be superimposed on those for which no MS/MS information was recorded. Label free quantification (LFQ) values were calculated using elution peak areas of peptides based on MS intensities. The sum of all peptide intensities within a sample was used to calculate normalization factors which were then automatically applied to each sample. In detail, Max Quant and Andromeda settings were selected as follows: Group specific parameters: 1. - General: Standard LC-MS run, Multiplicity 1, variable modifications, acetyl (N-term), oxidation (M); 1.1 - Instrument: Instrument type: Orbitrap; first search peptide tolerance, 20; main

search tolerance, 4.5; peptide tolerance unit, ppm; individual peptide mass tolerance, chosen; isotope match tolerance, 2 ppm; centroid match tolerance, 8 ppm; centroid half width, 35 ppm; time valley factor, 1.4; isotope time correlation, 0.6; theoretical isotope correlation, 0.6; recalibration unit, ppm; use MS1 centroids, not chosen; use MS2 centroids, not chosen; intensity dependent calibration, not chosen; min. peak length, 2; max. charge, 7; min. score for recalibration, 70, digestion mode, specific; enzyme, Trypsin/P; max. number of missed cleavages, 2; match type, match from and to; number of threads, 3; max. instrument type, cut peaks, chosen; gap scans, 1; advanced peak splitting, not chosen; intensity threshold 0; intensity determination, value at maximum. 1.3 - Label free quantification (LFQ): LFQ min. ratio count, 2; Fast LFQ, chosen; LFQ min. number of neighbors, 3; LFQ average number of neighbors, 6. 1.4 - Advanced: Max. number of modifications per peptide, 5; min. time, NaN; max. time, NaN; additional var. mods for special proteins, not chosen; separate variable modifications for first search, not chosen; separate enzyme for first search, not chosen.

Global parameters were chosen as follows: 2.1 - General: FASTA in-house file of combined *M. truncatula* and *R. leguminosarum* protein sequences including common contaminants; fixed modifications, carbamidomethyl (C); re-quantify, not chosen; match between runs, chosen; 2.2 - Sequences: decoy mode, revert; special AAs, KR; include contaminants, chosen; I=L, not chosen; max peptide mass, 4600 Da; min. peptide length for unspecific search, 8; max. peptide length for unspecific search, 25; 2.3 - Identification: PSM FDR, 0.01; protein FDR, 0.01; Site decoy fraction, 0.01; min. peptide length, 7; min. peptides, 1; min. razor + unique peptides, 1; min. unique peptides, 0; min. score for unmodified peptides, 0; min. score for modified peptides, 40; min. delta score for unmodified peptides, 0; min. delta score for modified peptides, 6; base FDR calculation of delta score, not chosen; razor protein FDR, chosen; split protein groups by taxonomy ID, not chosen; filter labelled amino acids, not chosen; second peptides, chosen; dependent peptides, not chosen; 2.4 - Protein quantification: min ratio count, 2; peptides for quantification, unique + razor; use only unmodified peptides and selected modifications, chosen; modifications used in protein quantification, acetyl (N-term), oxidation (M); discard unmodified counterpart peptide, chosen; 2.5 - LFQ: separate LFQ in parameter groups, not chosen; stabilize large LFQ ratios, chosen; require MS/MS for LFQ comparisons, chosen; iBAQ, not chosen, advanced site intensities, chosen; 2.6 - MS/MS - Fourier transformation mass spectrometry (FTMS): match tolerance, 20 ppm; de novo tolerance, 10 ppm; de-isotoping tolerance, 7 ppm; top peaks per 100 Da, 12; FTMS de-isotoping, higher charges, water loss, ammonia loss, dependent losses all chosen, FTMS recalibration not chosen 2.7 - Advanced: Use normalized ratios for occupancies; advanced ratio estimation, chosen; top x mass window, 100 Da.

#### 3.4.4 Data processing with Perseus

LFQ values as produced by the MaxQuant software were used for statistical analyses using the Perseus software (version 1.5.5.3, Tyanova *et al.* 2016). Imported MaxQuant data were cleared of proteins only identified by a single peptide containing a modification, decoy (reversed) entries, and common contaminations. Intensity values were log (base 2) transformed. MS result files were grouped according to sample origin (Fix<sup>+</sup>, Fix<sup>-</sup>, N, 0) and protein entries were classified as 'quantifiable', if they were identified at least three times within one of the four groups. The resulting dataset was submitted to analysis of variation (ANOVA) and two-sample t-tests for further validation using a threshold of  $p < 0.05$ .

## 4. Results

The four sample groups are similar in development, but differ in respect to SNF and nitrogen content. *Vicia faba* plants were harvested after 4 weeks. Plants were either supplied with mineral fertilizer ("N") or infected with *Rhizobium leguminosarum*



capable or incapable of SNF (“Fix<sup>+</sup>” and “Fix<sup>-</sup>”, respectively). Control plants (“0”) were neither fertilized nor infected with *R. leguminosarum* (Figure 1). At harvest time (after two weeks of treatment) plants in all four groups displayed the same growth, without showing phenotypic signs of nitrogen starvation. Only after continued cultivation was nitrogen starvation in 0 and Fix<sup>-</sup> discernable (data not shown). At harvest, Fix<sup>+</sup> nodules were of reddish color whereas Fix<sup>-</sup> nodules were whitish (Figure 2A). Production of molecular hydrogen (H<sub>2</sub>) in roots (including nodules) was monitored at harvest to test for nitrogenase activity. Only Fix<sup>+</sup> showed nitrogenase activity whereas no H<sub>2</sub>-production was detectable for the other lines (Figure 2B). To evaluate the impact of nitrogenase activity and fertilization on the plants, shoots were harvested, dried, and their carbon-to-nitrogen ration (C/N) was determined. The calculated C/N ratios provide information on nitrogen supply to the plants by SNF or mineral fertilization. Plants of group 0 possess the highest C/N ratio, followed by Fix<sup>-</sup>, Fix<sup>+</sup>, and N (Figure 2C). Together, these results indicate that the four groups do not suffer from nitrogen stress, nor do they differ in their development. At the same time, all four groups differ in their respective C/N ratios. In case of Fix<sup>+</sup> and Fix<sup>-</sup> this can be related to active/inactive nodules. The experimental setup is hence ideally suited to study the impact of symbiotic nitrogen fixation and nodulation without the need to consider secondary effects such as development and stress.

#### 4.1 General differences in the proteomes of the four sample groups

Fix<sup>+</sup> and Fix<sup>-</sup> nodule samples, as well as N and 0 root samples, were homogenized and proteins were subsequently extracted for proteomic analyses. Three biological replicates were prepared for each group. Since sequence information on *V. faba* is scarce, a cross-species identification approach had to be used. To ensure best proteome coverage, all twelve raw files were queried against three different databases: NCBIInr, SwissProt, and an in-house database combining *Medicago truncatula* and *R. leguminosarum* protein sequences. The proteome coverages for each individual database was then compared to each other (Suppl. Fig. 1). The genome of the model plant *M. truncatula* is sequenced and well annotated. Furthermore, it is closely related to *V. faba* and also performs SNF by engaging in symbiosis with rhizobium (Young and Udvardi 2009; Wojciechowski *et al.* 2004). As such, it was deemed a suitable data background for protein identifications of *V. faba* proteins. Results clearly indicate that the combined *M. truncatula* and *R. leguminosarum* database provides a better protein coverage in all samples and is thus the database of choice.

For a quantitative comparison of the proteomes in the four sample groups, the Andromeda search engine implemented in the MaxQuant software (Cox and Mann 2008) identified and quantified 2695 protein groups in total. These are referred to as ‘raw dataset’. To check for intra and inter group variance a three-dimensional Principal Component Analysis (PCA) was performed on this dataset using the Perseus software (Tyanova *et al.* 2016, Figure 3). The analysis revealed tight clustering of the biological replicates within each group and, with the exception of N and 0, a clear separation between the individual groups. The distance of Fix<sup>+</sup> and Fix<sup>-</sup> from N and 0 is expected, given that nodules and roots are different tissues. This may also explain the close proximity of N and 0. Considering this, the big difference between Fix<sup>+</sup> and Fix<sup>-</sup> cannot be easily explained since both tissues are of the same type.

The bulk of the 2695 identified proteins could be assigned to *M. truncatula* (1982 proteins, 74%) and only a minor part (621 proteins, 23%) are of *R. leguminosarum* origin. Ninety-two proteins (3%) are most likely contaminations or were excluded for other reasons (proteins only identified by a single modified peptide or decoy proteins, Figure 4, left chart). Proteins with low abundance were often not identified across all twelve samples, but were lacking in some of them. To ensure a solid statistical base for subsequent protein quantitation a filter was applied, which required the identification of the same protein in all biological

replicates of at least one of the four treatments (Fix<sup>+</sup>, Fix<sup>-</sup>, N, or 0). This ensures, that the proteins, which enter the quantitation process, are detectable and are therefore amenable to statistical validation of the quantitation. Application of the filter reduced the number of 'quantifiable' proteins to 817 (658 and 159 for the plant and bacterial datasets, respectively) of which 434 proteins (53%) appear in all four groups, 554 (68%) in at least three groups and 781 (96%) in at least two of the four groups. ANOVA resulted in the identification of 301 and 55 proteins differing significantly among the plant and the rhizobium proteins, respectively (Figure 4, right chart). As such, approximately 45% of the total quantifiable plant proteins are altered in response to nitrogen status and/or nodulation.

#### 4.2 Specific differences in the proteomes of the four sample groups

Proteins with significantly changed abundance in four of the six conceivable pairwise comparisons were identified using t-tests (Figure 5). Only those proteins with p-values  $\leq 0.05$  were considered (Suppl. Tables 1-4). In each comparison, the bulk of proteins (73% - 83%) did not differ, suggesting a broadly similar metabolism between nodules and roots, as well as between nitrogen availability and nitrogen deficiency within each tissue type. The most striking changes can be observed between Fix<sup>+</sup> and N (173 proteins, 27%) and Fix<sup>-</sup> and 0 (160 proteins, 23%). Both comparisons document the differences between root and nodule tissue, albeit at different nitrogen availabilities (high in N and Fix<sup>+</sup>, low in 0 and Fix<sup>-</sup>, Figure 2). When comparing within the same tissue types (Fix<sup>+</sup>/Fix<sup>-</sup> and N/0), the numbers of altered proteins were considerably less. Only 79 proteins (17%) changed between Fix<sup>+</sup> and Fix<sup>-</sup> while 127 proteins (22%) were of altered abundance between N and 0. Interestingly, the PCA analysis (Figure 3) performed on the unfiltered (quantifiable as well as not quantifiable proteins) showed a higher difference between Fix<sup>+</sup> and Fix<sup>-</sup>, indicative of a high influence of not-quantifiable proteins on the analysis.

To assess the most significant changes in the root and nodule proteomes, the five proteins with the highest and lowest fold changes in each of the four pairwise comparisons are summarized in Figure 6. Within this set of 40 proteins, 26 accessions are unique to one comparison, while seven appear in more than one comparison, thus reducing the set to 33 non-redundant accessions. Each of the four comparisons highlights different aspects of SNF. Comparing Fix<sup>+</sup> and Fix<sup>-</sup> allows identification of plant proteins involved in SNF, such as asparagine synthetase (two isoforms; 1, 2) and aspartate aminotransferase (3), all three of which are upregulated in Fix<sup>+</sup>. Enzymes of aspartate and asparagine metabolism are also more abundant in Fix<sup>+</sup> if compared to N. Phosphoenolpyruvate carboxylase (PEPC, 5) supports nitrogen fixation by converting the glycolytic intermediate PEP into oxaloacetate, which is the precursor for malate, the main substrate of bacteroid energy metabolism (and other processes). The enzyme thus links nitrogen fixation to carbon metabolism and respiration (Nomura 2006). The finding, that four of the top five proteins upregulated in Fix<sup>+</sup> are directly connected to SNF strongly indicates that the 12-oxophytodienoate reductase-like protein (OPR2, 4) is involved in SNF as well. The enzyme is proposed to function in the production of oxylipins (i.e. jasmonate) and is also increased in soybean and *M. truncatula* nodules in drought stress (Gil-Quintana *et al.* 2015). Mechanostimulation also led to increased levels of jasmonate in *M. truncatula* root nodules and was suggested to be the result of nodule harvest (Zdyb *et al.* 2011). A similar effect may have also led to the induction of the OPR here, but it is remarkable that this effect was not found in the Fix<sup>-</sup>/0 comparison.

Proteins upregulated in Fix<sup>-</sup> are probably linked to increased nodule formation and/or nodule senescence. Four proteins are involved in protein biosynthesis and protein fate: prolyl-tRNA synthetase family protein (6), 26S proteasome regulatory particle triple-A ATPase protein (7), translation elongation factor EF-2 subunit (8), and polyadenylate-binding protein (10). The fifth protein (clathrin heavy chain, CHC, 9) is a central component of clathrin-mediated endocytosis in eukaryotic cells and therefore

possibly involved in the uptake of bacteria into the host cells. A *CHC1* gene was found to be highly expressed in rhizobium-infected root hairs (Wang *et al.* 2015).

The comparison of  $\text{Fix}^+$  and N reveals differences between nodule and root tissue, when both are sufficiently supplied with nitrogen. Proteins more abundant in  $\text{Fix}^+$  contain typical enzymes of SNF: asparagine synthase, aspartate aminotransferase, and phosphoenolpyruvate carboxylase (1, 3, and 5). In addition, S-adenosylmethionine synthase-like protein (13), and 3,4-dihydroxy-2-butanone 4-phosphate synthase (DHBP, 14) are upregulated. No typical enzymes of  $\text{NO}_3^-$  assimilation were among the higher abundant proteins in N, but two proteins related to energy metabolism were found: NADP-dependent malic enzyme-like protein (18), and glucose-6-phosphate isomerase (20), indicative of an altered energy physiology in  $\text{Fix}^+$  nodules and roots.

The comparison between 0 and  $\text{Fix}^-$  showcases tissue specific differences under nitrogen deficiency (Figure 2B), which could explain the increased abundance of stress related proteins in 0, for instance the stress-inducible protein (26) and the peroxidase family protein (27). Other proteins present in higher amounts in 0 (17, 29 & 30) are involved in several unspecific metabolic and transport processes.  $\text{Fix}^-$  contains higher levels of proteins involved in protein biosynthesis (GTP-binding elongation factor Tu family protein, 23; 40S ribosomal S26-like protein, 24). Sucrose synthase (SuSy, 25) hydrolyzes sucrose to provide energy and carbon skeletons for SNF. As observed in  $\text{Fix}^+/\text{N}$ , the S-adenosylmethionine synthase-like protein (SAM, 13) is present in higher amounts in the nodule sample. Also of increased abundance in  $\text{Fix}^-$  is annexin D8 (22).

The comparison of 0 and N reveals the impact of nitrogen status on the root tissue proteome. Alterations in protein abundance are generally low and no coherent picture of the physiological impact of nitrogen deficiency emerges.

#### 4.3 Proteins with altered abundance participate in diverse cellular functions

A broader overview of the physiological differences between nodules capable and incapable of SNF, as well as between nodules and roots, is provided by concentrating on the two comparisons  $\text{Fix}^+/\text{Fix}^-$  and  $\text{Fix}^+/\text{N}$ . For this, proteins with significantly altered abundance were grouped according to cellular processes (Suppl. Figure 2). Proteins participating in carbohydrate metabolic processes, as well as starch, ammonium, amino acid, and sulfur metabolism are more abundant in  $\text{Fix}^+$  when compared to  $\text{Fix}^-$  and N. Except for sulfur metabolism, these processes are directly involved in the assimilation of the fixed nitrogen and the provision of energy to the bacteroids. The term ‘protein biosynthesis’ is noticeably more often associated with  $\text{Fix}^+$  and  $\text{Fix}^-$ , supporting the importance of this process in the formation of nodules. In contrast, proteolysis related proteins and chaperones are more abundant in  $\text{Fix}^-$  and N suggesting higher protein turnover rates in nodules incapable of SNF and in roots sufficiently supplied with nitrogen. Similarly, fatty acid metabolism, as well as stress related and transport processes, also seem to be enhanced in N and  $\text{Fix}^-$ . In addition, N shows a higher abundance of TCA cycle members and components of oxidative phosphorylation (OXPHOS) pointing towards an altered mode of mitochondrial energy metabolism in nodules.

##### 4.3.1 Meta comparisons

To identify enzymes involved in either SNF or maintenance of the symbiosis, proteins with altered abundance in more than one pairwise comparison were identified. Proteins upregulated in  $\text{Fix}^+$  when compared to  $\text{Fix}^-$  and N (meta comparison 1, MC1, Suppl. Figure 3) are very likely involved in biological nitrogen fixation. Similarly, proteins upregulated in both nodule forming sample groups, when compared to the root groups, are considered necessary for the maintenance of the symbiosis (MC2, Figure 5). In total, eight proteins of higher abundance in  $\text{Fix}^+$  in MC1, while 42 proteins were identified in MC2. Seven of the eight MC1 derived proteins were also identified in MC2 (Table 1).

## 5. Discussion

### 5.1 Experimental setup

The complexity of SNF stems from its presence within specialized plant organs, the root nodules, which harbor the temporary bacterial endosymbionts required for SNF. Drastic changes in cellular metabolism accompany the formation of the nodules, as well as their maintenance, and require an adjustment of cellular proteomes to support the bacterial nitrogen reduction and plant ammonia assimilation. A vast number of proteins have already been linked to SNF (Salvati *et al.* 2013). For many, however, it is not clear if they are directly involved in fixation of atmospheric nitrogen or in nodulation related processes. The pairwise comparisons and the meta analyses across the set of the four sample groups, as presented in this study, provide the means to assign altered protein abundance to either of the two functions with higher precision. However, it should be noted that the analyses presented here cannot discriminate between infected and non-infected nodule cells in the SNF zone, nor between this zone and the nodule meristem or the senescence zone. Altered protein abundance are thus average values across the entire plant organ. They are nevertheless in concordance with previous results and allow generating new hypotheses, which need further testing in the future. Likewise, the regulation of enzyme activity by post-translational modifications to the proteins cannot be assessed by this approach.

#### 5.1.1 Presence and absence of classical SNF-related enzymes

In the course of this study, several enzymes of SNF were identified, some of them in multiple isoforms (Suppl. Figure 4). These include glutamine synthase (GS), glutamate synthase (GOGAT), aspartate aminotransferase (AspAT), and asparagine synthase (AS). Most of these enzymes are of increased abundance in Fix<sup>+</sup>. Noticeably, no peptides of Fd-GOGAT were found in Fix<sup>+</sup>, while several peptides of this enzyme were found in N, O, and Fix<sup>-</sup>. Fd-GOGAT in roots of maize and rice is mainly active in the assimilation of NH<sub>4</sub><sup>+</sup>, derived from reduction of nitrate (NO<sub>3</sub><sup>-</sup>), or has other functions in amino acid metabolism (Hayakawa *et al.* 2003). Hence, it is not involved in SNF and therefore of only limited use in active root nodules. NADH-GOGAT is the main species of this enzyme employed for SNF (Vance *et al.* 1995). Peptides for NADH-GOGAT were only found in a single (Fix<sup>+</sup>) MS-run but were absent in all other analyses. The absence of NADH-GOGAT and some other expected proteins in the comparisons and meta-comparisons presented here may have multiple reasons. Proteins with low abundance are notoriously hard to detect and are often missing from MS analyses. Moreover, low-abundant peptides often fail to meet the stringency criteria for quantitation and are hence withdrawn from further analyses. In addition, the cross-species identification approach using a *M. truncatula* protein database for the identification of *V. faba* proteins may further reduce the number of positive peptide identifications. The lack of proteins in the result tables presented here is therefore of lack of a comprehensive detection of these proteins due to biological and technical reasons and no conclusions can be drawn on the relative abundance of said proteins failing to meet the quantitation criteria.

#### 5.1.2 Meta comparisons identify commonly regulated proteins in pairwise comparisons and allow to distinguish SNF from nodulation related proteins.

Eight proteins are commonly regulated in Fix<sup>+</sup> when compared to Fix<sup>-</sup> or N (MC1). In contrast, 42 proteins are upregulated when nodules and root tissue are compared at high nitrogen supply (Fix<sup>+</sup> vs N) and low nitrogen availability (Fix<sup>-</sup> vs 0, MC2). Seven of

the MC1 proteins were also found in MC2. This may indicate, that the expression of the corresponding genes is already increased in response to the nodulation event, but further boosted by the onset of SNF. Indeed, all of these proteins change more between Fix<sup>+</sup> and N (response to nodulation and SNF activity) than they do between Fix<sup>+</sup> and Fix<sup>-</sup> (response to SNF activity, Suppl. Fig. 5).

Five of the proteins commonly found in MC1 and MC2 (aspartate aminotransferase, two isoforms of phosphoenolpyruvate carboxylase, neutral/alkaline invertase, and sucrose synthase) have been linked to SNF in previous studies (Küster *et al.* 1997; Farnham *et al.* 1990; Nomura 2006; Gordon *et al.* 1999; Welham *et al.* 2009). As such, the two remaining proteins (UDP-glucuronic acid decarboxylase, and glycogen/starch/alpha-glucan phosphorylase family protein) can be expected to be important for SNF as well, but their functions are not established yet.

In contrast to the other seven MC1 proteins also found in MC2, asparagine synthetase is exclusively found in MC1 pointing towards an increased abundance of this protein directly in response to SNF activity. Among the residual 35 proteins in MC2, 17 are involved in protein biosynthesis (grey background in Table 1). Indeterminate nodules, as found in *V. faba*, are growing continuously. The presence of protein biosynthesis related proteins is therefore expected. Fourteen proteins (blue background in Table 1) participate in carbohydrate metabolism and are most likely involved in a modified carbon metabolism within infected cells. Six proteins (yellow background in Table 1) are related to sulfur metabolism (OAS-TL, 26, 27; SAM, 36, 37, 38; sulfite reductase, 42), further strengthening the notion of a connection between SNF and sulfur metabolism (Kalloniati *et al.* 2015). Since the presence of these proteins is also increased in the Fix<sup>-</sup>/0 comparison, sulfur fixation seems to be largely independent from SNF. Four proteins are engaged in amino acid metabolism, two of which are reported to be involved in SNF (AS, AST). Two proteins, exclusively found in MC2, are alanine aminotransferase (AlaAT) and an enzyme involved in lysine metabolism, LL-diaminopimelate aminotransferase (DAP). The putative role of AlaAT is discussed below.

### 5.1.3 Ammonium assimilation in *V. faba*:*R. leguminosarum* symbiosis

Assimilation of ammonium derived from nitrogenase is primarily performed by the plant enzymes glutamine synthetase (GS) and glutamate synthase (GOGAT, Ohyama and Kumazawa 1980, Rawsthorne *et al.* 1980, Vance *et al.* 1995). Amino groups are subsequently transferred to aspartate (by AST) and asparagine (by AS), which is the major form of nitrogen exported from nodules (Li *et al.* 2002, Ludwig *et al.* 2004, Kumar *et al.* 2005). All enzymes except for NADH-dependent GOGAT, which was below the detection limit in all but one (Fix<sup>+</sup>) sample, were identified in the course of this study (suppl. Fig. 4). Most of these enzymes are of higher average abundance in nodules when compared to root material but differences are often not significant. This may either be due to high variations in protein abundance within the same group or due to the complete absence of positive identifications within a single group or even several groups.

However, besides this main pathway of nitrogen export, alternative routes may exist. Alanine was reported to be exported from the bacteroids of pea (Waters *et al.* 1998). Later, *R. leguminosarum* was found to excrete both alanine and NH<sub>4</sub><sup>+</sup> (Allaway *et al.* 2000), whereas NH<sub>3</sub>/NH<sub>4</sub><sup>+</sup> were found to be the major products of determinate soybean bacteroid SNF (Li *et al.* 2002). Most of these results were obtained on *ex planta* bacteroids and therefore may not represent the *in vivo* situation. Transport of alanine from the bacteroids into the cytosol of the plant cell is possible (Waters *et al.* 1998; Day *et al.* 2001). An isoform of AlaAT is present in higher amounts in nodules when compared to roots (MC2) and was also more abundant in Fix<sup>+</sup> compared to Fix<sup>-</sup> (p=0.065). The identified AlaAT shows strong homology to AlaAT2 from *Arabidopsis thaliana*, which is predicted to be located in mitochondria (Liepman and Olsen 2003). Export of alanine from bacteroids and its breakdown in the mitochondria would enable the production

of glutamate which could subsequently be used in the cytosol to transaminate oxaloacetate and glutamate to produce aspartate and glutamine, respectively (Figure 7). A bacterial alanine dehydrogenase (AD) is exclusively present in the Fix<sup>+</sup> bacteroid dataset, producing alanine from pyruvate and ammonia. Plants infected with a mutant *R. leguminosarum* strain defective in alanine catabolism grew more slowly than plants nodulated by the wild-type (Allaway et al. 2000). However, over-expression of the enzyme did not improve the ability of pea plants to fix nitrogen (Lodwig et al. 2004). Our data thus highlight the potential existence of an alanine-driven pathway for asparagine production, which is largely in concordance with previous reports. No statements of the flux of nitrogen through this system, and thus on its importance, can be made from the data presented here.

#### 5.1.4 Allocation of sucrose into starch synthesis and glycolysis in nodules

Roots and nodules are heterotrophic tissues and rely on the provision of carbohydrates from photosynthesis, mostly in the form of sucrose. In the sink cells sucrose is either used to fuel glycolysis or is transformed into starch. Starch accumulation in nodules was observed for *V. faba* (Fred et al. 2002). In *Phaseolus vulgaris* starch is mainly present in uninfected nodule cells (Tate et al. 1994) and is remobilized when the shoot does not deliver sucrose to the root (Vance and Heichel 1991). The activity of starch-bound starch synthase in Fix<sup>+</sup> nodules declines over time in soybean nodules and a positive correlation between N<sub>2</sub> fixation and starch degradation in nodules was reported (Sharon et al. 1991). Concordantly, starch accumulates in young soybean nodules and is gradually metabolized during the progression of symbiosis, most likely to fulfil the energy demands of the bacteroids (Forrest et al. 1991). In contrast, soybean plastids are not emptied of starch granules until nodule senescence (Puppo et al. 2005).

Our results provide additional information on starch utilization in nodules. Starch related processes appear to be especially important in Fix<sup>+</sup>. MC1 contains proteins involved in starch metabolism, comprising sucrose synthase (SuSy), a neutral/alkaline invertase (INV), as well as a glycogen/starch/alpha-glucan phosphorylase family protein (GPase). These findings imply a role for starch in SNF, which is yet to be specified in more detail. As nitrogen fixation follows a 24 h rhythm it is possible, that stored starch in nodules is degraded to support SNF and bacteroid metabolism in the dark, in the same fashion it is used to drive respiration during the night (Wheeler 1969; Weise et al. 2011).

SuSy is essential for SNF (Gordon et al. 1999). As an alternative to starch synthesis, SuSy also provides substrates for glycolysis and synthesis of cell wall components (Barratt et al. 2001). Due to low fructose concentrations within the cells it is expected to operate predominantly in the cleaving direction (Morell and Copeland 1985). The meta-analysis detected several enzymes of the glycolytic pathway (glyceraldehyde-3-phosphate dehydrogenase, GAPDH; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, PGM; and phosphopyruvate hydratase, ENO). All of them are more abundant in Fix<sup>+</sup> and Fix<sup>-</sup>. The product of GAPDH, 1,3-bisphosphoglycerate also mediates the oxygen release from the heme group of hemoglobin in the blood (Stefansson et al. 2016), a similar function in relation to leghemoglobin may be hypothesized for nodules. Another glycolysis intermediate is 2-phosphoglycerate, which is converted to phosphoenolpyruvate by ENO. Phosphoenolpyruvate (PEP) may then be metabolized to oxaloacetate by phosphoenolpyruvate carboxylase (PEPC) or to pyruvate by pyruvate kinase. Expression of the PEPC gene was found to be positively correlated with nitrogenase activity (Nomura 2006) and is featured in both meta-analyses. It is therefore possible that the glycolytic pathway in root nodules is drained to a large degree at this position. This may have negative effects for downstream metabolic pathways, i.e. plant TCA cycle and respiration. Notably, plants in the sample group N possess higher amounts of TCA cycle enzymes and OXPHOS subunits than Fix<sup>+</sup>.

The actively growing indeterminate nodules of *V. faba* have an increased need for synthesizing cell walls. UDP-Glucose produced by SuSy can be converted into UDP glucuronic acid, an important substrate for plant cell wall polysaccharides such as xyloglucans

and plant glycoproteins. UDP glucuronic acid decarboxylase (UXS) is among the seven proteins present in both meta-analyses. The enzyme catalyzes the conversion of UDP-glucuronic acid to UDP-xylose. Additionally present in higher amounts in Fix<sup>+</sup> and Fix<sup>-</sup> is the rhamnose biosynthetic-like enzyme (Medtr2g073240.1, rhamnose) which is also involved in cell wall biosynthesis. Lignin, another cell wall component, is a metabolic sink for S-adenosylmethionine (Shen *et al.* 2002). Correspondingly, other enzymes related to lignin biosynthesis were found. One of them, a caffeic acid O-methyltransferase (Medtr4g038440.2), is also nodule specific (Fix<sup>+</sup> and Fix<sup>-</sup>).

#### 5.1.5 Sulfur and sulfur-related metabolism

Nodules are an important source of reduced sulfur (Kalloniati *et al.* 2015). This correlates with the frequent occurrence of enzymes involved in sulfur metabolism in the meta-analyses. Among these proteins the glycolytic enzyme D-3-phosphoglycerate dehydrogenase (PHGDH, Medtr4g005880.1) may take a central position since it controls the biosynthesis of serine. Serine, in turn, is a precursor for cysteine, the first organic form of reduced sulfur. It is produced from O-acetyl-L-serine and hydrogen sulfide (provided by sulfite dehydrogenase) by the O-acetylserine (thiol) lyase (OAS-TL) of which three isoforms are more abundant in Fix<sup>+</sup> and are also featured in the meta-comparisons. Kalloniati *et al.* (2015) proposed the incorporation of sulfate into glutathione. Our study provides no evidence for elevated levels of enzymes for glutathione synthesis. Enzymes of glutathione biosynthesis were either found in equal amounts in all four sample groups (glutamate-cysteine ligase B, Medtr8g098350.1) or did not occur in the dataset (glutathione synthetase). We therefore suggest the conversion of serine, with cysteine and methionine as intermediates, into S-adenosylmethionine (AdoMet), which is used as a general donor for methyl groups. AdoMet is synthesized by S-adenosylmethionine synthase (SAM), which was found twice in the meta-comparisons. AdoMet related proteins are a general feature of the proteome of Fix<sup>+</sup> plants (Figures 6, 8). SAM is a precursor for a diverse group of metabolites including cyclopropyl fatty acids, biotin, tRNA, lignin, ethylene, and polyamines (PA, Fontecave *et al.* 2004; Ravanel *et al.* 1998). The methyl transferases involved in these processes also produce S-Adenosyl-L-homocysteine (AdoHcy) which can be regenerated to AdoMet by the action of AdoHcy-Hydrolase (AdoHcyH), found in MC2 (Tab. 2). Together, the altered protein abundance suggests an important role for sulfur metabolism in root nodules which is possibly aimed at the production of AdoMet while the downstream utilization of this compound remains unclear.

#### 5.1.6 Control of nodulation, signaling

To cut unnecessary losses, nodule growth (and activity) is tightly controlled. In Fix<sup>-</sup> the development of nascent nodules is stopped during nodule development and the formation of new nodules is initiated. This results in a higher number of nodules of smaller size and tan color (Figure 2A).

Two proteins, potentially involved in regulating nodulation and SNF, were identified: annexin D8 (ANND8) and 12-oxophytodienoate reductase-like protein (OPR). Transcripts of annexin 1 (MtAnn1) are upregulated in root tissues in response to rhizobial Nod factors (Carvalho-Niebel *et al.* 2002). An analogous *M. truncatula* protein, MtAnn8, was upregulated in nodules when compared to roots, and is hypothesized to play a role in nodule growth and development (Marx *et al.* 2016). The appearance of annexin D8 in MC2 is therefore in support of this. OPR appears in the Top five comparison Fix<sup>+</sup>/Fix<sup>-</sup> (Figure 7). The protein is most likely involved in the conversion of linolenic acid to jasmonic acid (Li *et al.* 2009) in response to pathogen attack or infection by rhizobial bacteria (van Wees *et al.* 2008). Jasmonic acid negatively affects nodulation and transcripts of jasmonate biosynthesis

related genes, such as allene oxide cyclase and 12-oxophytodienoate reductase are highly increased in early infection stages (Sun *et al.* 2006; Kouchi 2004). The presence of OPR in MC1 also suggests a role of jasmonic acid in regulating SNF and nodulation events.

As described in the section concerning sulfur and sulfur-related metabolism (6.1.5) the metabolic fate of AdoMet is unclear. However, polyamine (spermidine, spermin) synthesis requires AdoMet and polyamines accumulate in *Lotus japonicus* nodules, providing a possible explanation for the increased AdoMet production in Fix<sup>+</sup> and Fix<sup>-</sup> (Flemetakis *et al.* 2004). Spermidine and spermin have been linked to autoregulation, a mechanism controlling nodule number in legumes (Terakado *et al.* 2006).

### 5.1.7 Summary

The proteomic comparisons, performed during the course of this study, allow to differentiate between alterations in root and nodule proteomes either induced by the fixation of atmospheric nitrogen and those which are a consequence of the bacteroid:plant symbiosis. They confirm results of previous studies but also extend them and provide new angles for further, targeted analyses of nodule physiology. Particularly interesting are the areas of nitrogen fixation and intracellular transport, sulfur fixation, starch metabolism and the role of plant hormones in establishing and maintaining nodules. Figure 7 provides an overview of newly suggested metabolic alterations in SNF-active root nodules (Fix<sup>+</sup>) of *V. faba*.

### 5.2 Outlook

Understanding SNF in detail is of high value for agricultural applications and for its potential to promote improvements in legume crop yields. Future perspectives to harness the benefits of SNF in non-legume crop plants are directly linked to the identification of the enzymatic infrastructure supporting nitrogenase function and incorporation of ammonia into organic compounds, as well as establishing and maintaining the plant:*R. leguminosarum* symbiosis. Knowledge of the proteins mediating the complex metabolic and regulatory interplay between plants and bacteroids may foster the transfer of legume SNF to non-legume plants. Several important aspects of nodule metabolism, besides amino acid metabolism, have been highlighted in this study: assimilation of ammonia and the potentially also of alanine derived from bacteroids, allocation of sucrose into starch synthesis and glycolysis, sulfur metabolism, and the control of nodulation. Proteomic studies often struggle in delivering direct evidence for altered metabolic pathways. Instead, they excel in generating new candidate proteins for more targeted research. Focusing on the topics and the cellular compartments outlined above, the results produced in this manuscript may form the basis for further studies employing the full suite of biochemical and genetic tools available today.

## 6. References

- Abreu EFM, Aragao FJL (2006) Isolation and Characterization of a myo-inositol-1-phosphate Synthase Gene from Yellow Passion Fruit (*Passiflora edulis* f. *flavicarpa*) Expressed During Seed Development and Environmental Stress. *Ann Bot* 99: 285–292. doi: 10.1093/aob/mcl256
- Allaway D, Ludwig EM, Crompton LA, Wood M, Parsons R, Wheeler TR, Poole PS (2000) Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol Microbiol* 36: 508–515.
- Appels MA, Haaker H (1991) Glutamate Oxaloacetate Transaminase in Pea Root Nodules. *Plant Physiol* 95: 740-747.



- Barratt DHP, Barber L, Kruger NJ, Smith AM, Wang TL, Martin C (2001) Multiple, Distinct Isoforms of Sucrose Synthase in Pea. *Plant Physiol* 127: 655–664.
- Bestel-Corre G, Dumas-Gaudot E, Poinso V, Dieu M, Dierick J-F, van Tuinen D, Remacle J, Gianinazzi-Pearson V, Gianinazzi S (2002) Proteome analysis and identification of symbiosis-related proteins from *Medicago truncatula* Gaertn. by two-dimensional electrophoresis and mass spectrometry. *Electrophoresis* 23: 122-137. doi: 10.1002/1522-2683(200201)23:1<122:AID-ELPS122>3.0.CO;2-4
- Biederbeck VO, Campbell CA, Rasiah V, Zentner RP, Wen G (1998) Soil quality attributes as influenced by annual legumes used as green manure. *Soil Biol Biochem* 30: 1177–1185. doi: 10.1016/S0038-0717(97)00150-8
- Carvalho-Niebel F de, Timmers ACJ, Chabaud M, Defaux-Petras A, Barker DG (2002) The Nod factor-elicited annexin MtAnn1 is preferentially localised at the nuclear periphery in symbiotically activated root tissues of *Medicago truncatula*. *Plant J* 32: 343–352. doi: 10.1046/j.1365-313X.2002.01429.x
- Clarke VC, Loughlin PC, Gavrin A, Chen C, Brear EM, Day DA, Smith PMC (2015) Proteomic analysis of the soybean symbiosome identifies new symbiotic proteins. *Mol Cell Proteomics* 14: 1301–1322. doi: 10.1074/mcp.M114.043166
- Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J* 39: 487–512. doi: 10.1111/j.1365-313X.2004.02150.x
- Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367–1372. doi: 10.1038/nbt.1511
- Day DA, Poole PS, Tyerman SD, Rosendahl L (2001) Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell Mol Life Sci* 58: 61–71. doi: 10.1007/PL00000778
- Delmotte N, Ahrens CH, Knief C, Qeli E, Koch M, Fischer H-M, Vorholt JA, Hennecke H, Pessi G (2010) An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *J Proteomics* 10: 1391–1400. doi: 10.1002/pmic.200900710.
- Djordjevic MA (2004) *Sinorhizobium meliloti* metabolism in the root nodule: A proteomic perspective. *J Proteomics* 4: 1859–1872. doi: 10.1002/pmic.200300802
- Durner J, Klessig DF (1999) Nitric oxide as a signal in plants. *Curr Opin Plant Biol* 2: 369–374. doi: 10.1016/S1369-5266(99)00007-2
- Farnham MW, Miller SS, Griffith SM, Vance CP (1990) Aspartate Aminotransferase in Alfalfa Root Nodules: II. Immunological Distinction between Two Forms of the Enzyme. *Plant Physiol* 93: 603–610. doi: 10.1104/pp.93.2.603
- Flemetakis E, Efroze RC, Desbrosses G, Dimou M, Delis C, Aivalakis G, Udvardi MK, Katinakis P (2004) Induction and spatial organization of polyamine biosynthesis during nodule development in *Lotus japonicus*. *Mol Plant Microbe Interact* 17: 1283–1293. doi: 10.1094/MPMI.2004.17.12.1283

- Fontecave M, Atta M, Mulliez E (2004) S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* 29: 243–249. doi: 10.1016/j.tibs.2004.03.007
- Forrest SI, Verma DPS, Dhindsa RS (1991) Starch content and activities of starch-metabolizing enzymes in effective and ineffective root nodules of soybean. *Can J Bot* 69: 697–701. doi: 10.1139/b91-094
- Fred EB, Baldwin IL, McCoy E, Triplett EW (2002) Root nodule bacteria and leguminous plants. University of Wisconsin studies in science, No. 5. Parallel Press, Madison, Wisc.
- Fromm S, Senkler J, Eubel H, Peterhansel C, Braun H-P (2016) Life without complex I: proteome analyses of an Arabidopsis mutant lacking the mitochondrial NADH dehydrogenase complex. *J Exp Bot* 67: 3079–3093. doi: 10.1093/jxb/erw165
- Gálvez S, Hirsch AM, Wycoff KL, Hunt S, Layzell DB, Kondorosi A, Crespi M (2000) Oxygen Regulation of a Nodule-Located Carbonic Anhydrase in Alfalfa. *Plant Physiol* 124: 1059–1068. doi: 10.1104/pp.124.3.1059
- Gil-Quintana E, Lyon D, Staudinger C, Wienkoop S, González EM (2015) Medicago truncatula and Glycine max: Different Drought Tolerance and Similar Local Response of the Root Nodule Proteome. *J Proteome Res* 14: 5240–5251. doi: 10.1021/acs.jproteome.5b00617
- Gordon AJ, James CL (1997) Enzymes of carbohydrate and amino acid metabolism in developing and mature nodules of white clover. *J Exp Bot* 48: 895–903. doi: 10.1093/jxb/48.4.895
- Graham PH (2003) Legumes: Importance and Constraints to Greater Use. *Plant Physiol* 131: 872–877. doi: 10.1104/pp.017004
- Guerreiro N, Redmond JW, Rolfe BG, Djordjevic MA (1997) New Rhizobium leguminosarum flavonoid-induced proteins revealed by proteome analysis of differentially displayed proteins. *MPMI* 10: 506–516. doi: 10.1094/MPMI.1997.10.4.506
- Györgyey J, Vaubert D, Jiménez-Zurdo JI, Charon C, Troussard L, Kondorosi Á, Kondorosi É (2000) Analysis of Medicago truncatula Nodule Expressed Sequence Tags. *MPMI* 13: 62–71. doi: 10.1094/MPMI.2000.13.1.62
- Hayakawa T, Sakai, T, Ishiyama K, Hirose, N, Nakajima H, Takezawa M, Naito K, Hino-Nakayama M, Akagawa T, Goto S Yamaya T (2003) Organization and Structure of Ferredoxin-Dependent Glutamate Synthase Gene and Intracellular Localization of the Enzyme in Rice Plants. *Plant Biotech* 20: 43-55.
- Herz S, Eberhardt S, Bacher A (2000) Biosynthesis of riboflavin in plants. The ribA gene of Arabidopsis thaliana specifies a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase. *Phytochemistry* 53: 723–731
- Hoa LT-P (2004) Proteomic Analysis on Symbiotic Differentiation of Mitochondria in Soybean Nodules. *Plant Cell Physiol* 45: 300–308. doi: 10.1093/pcp/pch035
- Hoagland D. R., Arnon D. I. (1950) The water-culture method for growing plants without soil. Circular. California Agricultural Experiment Station 347:1-32.
- Hoffman BM, Lukoyanov D, Yang Z-Y, Dean DR, Seefeldt LC (2014) Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem Rev* 114: 4041–4062. doi: 10.1021/cr400641x

- Hulsen T, Vlieg J de, Alkema W (2008) BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9: 488. doi: 10.1186/1471-2164-9-488
- Kalloniati C, Krompas P, Karalias G, Udvardi MK, Rennenberg H, Herschbach C, Flemetakis E (2015) Nitrogen-Fixing Nodules Are an Important Source of Reduced Sulfur, Which Triggers Global Changes in Sulfur Metabolism in *Lotus japonicus*. *The Plant Cell* 27: 2384-400. doi: 10.1105/tpc.15.00108
- Knights BA, McKinley WM, Wheeler C (1977) Sterols of roots and nitrogen fixing root nodules of *vicia faba*. *Phytochemistry* 16: 727-728. doi: 10.1016/S0031-9422(00)89241-9
- Koch M, Delmotte N, Rehrauer H, Vorholt JA, Pessi G, Hennecke H (2010) Rhizobial adaptation to hosts, a new facet in the legume root-nodule symbiosis. *MPMI* 23: 784-790. doi: 10.1094/MPMI-23-6-0784
- Küster H, Albus U, Frühling M, Tchetkova SA, Tikhonovitch IA, Pühler A, Perlick AM (1997) The asparagine synthetase gene *VfAS1* is strongly expressed in the nitrogen-fixing zone of broad bean (*Vicia faba* L.) root nodules. *Plant Sci* 124: 89-95. doi: 10.1016/S0168-9452(97)04607-4
- Kumar S, Bourdès A, Poole P (2005) De Novo Alanine Synthesis by Bacteroids of *Mesorhizobium loti* Is Not Required for Nitrogen Transfer in the Determinate Nodules of *Lotus corniculatus*. *J Bacteriol* 187: 5493-5495. doi: 10.1128/JB.187.15.5493-5495.2005
- Laemmli UK (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227:680-685. doi: 10.1038/227680a0
- Larrainzar E, Wienkoop S, Weckwerth W, Ladrera R, Arrese-Igor C, Gonzalez EM (2007) *Medicago truncatula* Root Nodule Proteome Analysis Reveals Differential Plant and Bacteroid Responses to Drought Stress. *Plant Physiol* 144: 1495-1507. doi: 10.1104/pp.107.101618
- Lea PJ, Mifflin BJ (2003) Glutamate synthase and the synthesis of glutamate in plants. *Plant Physiol Biochem* 41: 555-564. doi: 10.1016/S0981-9428(03)00060-3
- Lee J-K, Hurwitz J (2001) Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. *Proceedings of the National Academy of Sciences* 98: 54-59. doi: 10.1073/pnas.98.1.54
- Li Y, Parsons R, Day DA, Bergersen FJ (2002) Reassessment of major products of N<sub>2</sub> fixation by bacteroids from soybean root nodules. *Microbiology* 148: 1959-1966. doi: 10.1099/00221287-148-6-1959
- Li W, Liu B, Yu L, Feng D, Wang H, Wang J (2009) Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. *BMC Evol Biol* 9: 90. doi: 10.1186/1471-2148-9-90
- Liepman AH, Olsen LJ (2003) Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxisomes of *Arabidopsis*. *Plant Physiol* 131: 215-227. doi: 10.1104/pp.011460
- Lodwig EM, Hosie AH, Bourdès A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS. (2003) Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature* 422: 722-6. doi: 10.1038/nature01527

- Lodwig E, Kumar S, Allaway D, Bourdès A, Prell J, Priefer U, Poole P (2004) Regulation of l-alanine dehydrogenase in *Rhizobium leguminosarum* bv. *viciae* and its role in pea nodules. *J Bacteriol* 186: 842-849. doi: [10.1128/JB.186.3.842-849.2004](https://doi.org/10.1128/JB.186.3.842-849.2004)
- Ma Q-S, Johnston AWB, Hombrecher G, Downie JA (1982) Molecular genetics of mutants of *Rhizobium leguminosarum* which fail to fix nitrogen. *Molec Gen Genet* 187: 166-171. doi: [10.1007/BF00384401](https://doi.org/10.1007/BF00384401)
- Madsen O, Sandal L, Sandal NN, Marcker KA (1993) A soybean coproporphyrinogen oxidase gene is highly expressed in root nodules. *Plant Mol Biol* 23: 35-43. doi: [10.1007/BF00021417](https://doi.org/10.1007/BF00021417)
- Marx H, Minogue CE, Jayaraman D, Richards AL, Kwiecien NW, Siahpirani AF, Rajasekar S, Maeda J, Garcia K, Del Valle-Echevarria AR, Volkening JD, Westphall MS, Roy S, Sussman MR, Ané J-M, Coon JJ (2016) A proteomic atlas of the legume *Medicago truncatula* and its nitrogen-fixing endosymbiont *Sinorhizobium meliloti*. *Nat Biotechnol* 34: 1198-1205. doi: [10.1038/nbt.3681](https://doi.org/10.1038/nbt.3681)
- Morell M, Copeland L (1985) Sucrose Synthase of Soybean Nodules. *Plant Physiol* 78: 149-154. doi: [10.1104/pp.78.1.149](https://doi.org/10.1104/pp.78.1.149)
- Neuhoff V, Stamm R, Eibl H (1985) Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: A systematic analysis. *Electrophoresis* 6: 427-448. doi: [10.1002/elps.1150060905](https://doi.org/10.1002/elps.1150060905)
- Nomura M (2006) Phosphoenolpyruvate Carboxylase Plays a Crucial Role in Limiting Nitrogen Fixation in *Lotus japonicus* Nodules. *Plant Cell Physiol* 47: 613-621. doi: [10.1093/pcp/pcj028](https://doi.org/10.1093/pcp/pcj028)
- Ohyama T, Kumazawa K (1980) Nitrogen assimilation in soybean nodules. *Soil Science and Plant Nutrition* 26: 109-115. doi: [10.1080/00380768.1980.10433217](https://doi.org/10.1080/00380768.1980.10433217)
- Oldroyd GED, Murray JD, Poole PS, Downie JA (2011) The rules of engagement in the legume-Rhizobial symbiosis. *Annu Rev Genet* 45: 119-144. doi: [10.1146/annurev-genet-110410-132549](https://doi.org/10.1146/annurev-genet-110410-132549)
- Oldroyd GE, Downie JA (2008) Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annu Rev Plant Biol* 59: 519-546. doi: [10.1146/annurev.arplant.59.032607.092839](https://doi.org/10.1146/annurev.arplant.59.032607.092839)
- Peiter E, Yan F, Schubert S (2004) Amino acid export from infected cells of *Vicia faba* root nodules: Evidence for an apoplastic step in the infected zone. *Physiol Plant* 122: 107-114. doi: [10.1111/j.1399-3054.2004.00382.x](https://doi.org/10.1111/j.1399-3054.2004.00382.x)
- Poole P, Allaway D (2000) Carbon and nitrogen metabolism in *Rhizobium*. *Adv Microb Physiol* 43: 117-163.
- Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS (2009) Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. *Proc Natl Acad Sci U S A* 106: 12477-12482. doi: [10.1073/pnas.0903653106](https://doi.org/10.1073/pnas.0903653106)
- Puppo A, Groten K, Bastian F, Carzaniga R, Soussi M, Lucas MM, Felipe MR de, Harrison J, Vanacker H, Foyer CH (2005) Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytol* 165: 683-701. doi: [10.1111/j.1469-8137.2004.01285.x](https://doi.org/10.1111/j.1469-8137.2004.01285.x)
- Ravanel S, Gakiere B, Job D, Douce R (1998) The specific features of methionine biosynthesis and metabolism in plants. *Proc Natl Acad Sci U S A* 95: 7805-7812

- Rawsthorne S, Minchin FR, Summerfield RJ, Cookson C, Coombs J (1980) Carbon and nitrogen metabolism in legume root nodules. *Phytochemistry* 19: 341–355. doi: 10.1016/0031-9422(80)83181-5
- Reynolds PHS, Hine A, Rodber K (1988) Serine metabolism in legume nodules: Purification and properties of phosphoserine aminotransferase. *Physiol Plant* 74: 194–199. doi: 10.1111/j.1399-3054.1988.tb04963.x
- Rosendahl L, Dilworth MJ, Glenn AR. (1992) Exchange of metabolites across the peribacteroid membrane in pea root nodules. *J Plant Physiol* 139: 635–38.
- Salavati A, Shafeinia A, Klubicova K, Bushehri AAS, Komatsu S (2013) Proteomic insights into intra- and intercellular plant-bacteria symbiotic association during root nodule formation. *Front Plant Sci* 4: 28. doi: 10.3389/fpls.2013.00028
- Šamaj J, Thelen JJ (2007) *Plant proteomics*. Springer, Berlin, New York
- Schubert (1986) *Products of Biological Nitrogen Fixation in Higher Plants: Synthesis, Transport, and Metabolism*. *Annu Rev Plant Biol* 37: 539-574. doi.org/10.1146/annurev.pp.37.060186.002543
- Sellstedt A, Reddell P, Rosbrook Pa, Ziehr A (1991) The Relations of Haemoglobin and Lignin-like Compounds to Acetylene Reduction in Symbiotic Casuarina. *J Exp Bot* 42: 1331–1337. doi: 10.1093/jxb/42.10.1331
- Sharon I. Forrest, Desh Pal S. Verma, Rajinder S. Dhindsa (1991) Starch content and activities of starch-metabolizing enzymes in effective and ineffective root nodules of soybean. *Can J Bot* 69: 697-701. doi.org/10.1139/b91-094
- Shen B, Li C, Tarczynski MC (2002) High free-methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl-L-methionine synthetase 3 gene. *Plant J* 29: 371–380
- Shimajima M, Hoffmann-Benning S, Garavito RM, Benning C (2005) Ferredoxin-dependent glutamate synthase moonlights in plant sulfolipid biosynthesis by forming a complex with SQD1. *Arch Biochem Biophys* 436: 206–214. doi: 10.1016/j.abb.2005.02.005
- Stefansson S, Chung DS, Yoon J, Yoo WS, Park YW, Kim G, Hahn D, Le H, Chung S-J, Bruttig SP, Ho DH (2016) Improving Oxygen Binding of Desiccated Human Red Blood Cells. *ABB* 07: 47–54. doi: 10.4236/abb.2016.72006
- Stewart DC, Copeland L (1999) Kinetic properties of UDP-glucose dehydrogenase from soybean nodules. *Plant Sci* 147: 119–125. doi: 10.1016/S0168-9452(99)00110-7
- Sun J, Cardoza V, Mitchell DM, Bright L, Oldroyd G, Harris JM (2006) Crosstalk between jasmonic acid, ethylene and Nod factor signaling allows integration of diverse inputs for regulation of nodulation. *Plant J* 46: 961–970. doi: 10.1111/j.1365-313X.2006.02751.x
- Tate R, Patriarca EJ, Riccio A, Defez R, Iaccarino M (1994) Development of *Phaseolus vulgaris* root nodules. *MPMI* 7: 582–589
- Temple SJ, Vance CP, Stephen Gantt J (1998) Glutamate synthase and nitrogen assimilation. *Trends Plant Sci* 3: 51–56. doi: 10.1016/S1360-1385(97)01159-X

- Terakado J, Yoneyama T, Fujihara S (2006) Shoot-applied polyamines suppress nodule formation in soybean (*Glycine max*). *Plant Physiol* 163: 497–505. doi: 10.1016/j.jplph.2005.05.007
- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Meth* 13: 731–740. doi: 10.1038/nmeth.3901
- Tyerman SD, Whitehead LF, Day DA (1995) A channel-like transporter for  $\text{NH}_4^+$  on the symbiotic interface of  $\text{N}_2$ -fixing plants. *Nature* 378: 629–632. doi: 10.1038/378629a0
- Udvardi MK, Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. *Annu Rev Plant Biol* 48: 493–523. doi: 10.1146/annurev.arplant.48.1.493
- Udvardi M, Poole PS (2013) Transport and Metabolism in Legume-Rhizobia Symbioses. *Annu Rev Plant Biol* 64: 781–805. doi: 10.1146/annurev-arplant-050312-120235
- van Noorden GE, Kerim T, Goffard N, Wiblin R, Pellerone FI, Rolfe BG, Mathesius U (2007) Overlap of proteome changes in *Medicago truncatula* in response to auxin and *Sinorhizobium meliloti*. *Plant Physiol* 144: 1115–1131. doi: 10.1104/pp.107.099978
- van Wees SCM, van der Ent S, Pieterse CMJ (2008) Plant immune responses triggered by beneficial microbes. *Curr Opin Plant Biol* 11: 443–448. doi: 10.1016/j.pbi.2008.05.005
- van Wyk SG, Du Plessis M, Cullis CA, Kunert KJ, Vorster BJ (2014) Cysteine protease and cystatin expression and activity during soybean nodule development and senescence. *BMC Plant Biol* 14: 243. doi: 10.1186/s12870-014-0294-3
- Vance CP, Bryan JW (1981) Purification and properties of caffeic acid O-methyltransferase from alfalfa root nodules. *Phytochemistry* 20: 41–43. doi: 10.1016/0031-9422(81)85215-6
- Vance CP, Gregerson RG, Robinson DL, Miller SS, Gantt JS (1994) Primary assimilation of nitrogen in alfalfa nodules: Molecular features of the enzymes involved. *Plant Sci* 101: 51–64. doi: 10.1016/0168-9452(94)90164-3
- Vance CP, Heichel GH (1991) Carbon in  $\text{N}_2$  fixation: limitation or exquisite adaptation. *Annual review of plant biology* 42(1):373–390.
- Vance CP, Miller SS, Gregerson RG, Samac DA, Robinson DL, Gantt JS (1995) Alfalfa NADH-dependent glutamate synthase: structure of the gene and importance in symbiotic  $\text{N}_2$  fixation. *Plant Journal* 8: 345-58. DOI: 10.1046/j.1365-313X.1995.08030345.x
- Vázquez-Lobo A, Roujol D, Zuñiga-Sánchez E, Albenne C, Piñero D, Buen AGd, Jamet E (2012) The highly conserved spermatophyte cell wall DUF642 protein family: Phylogeny and first evidence of interaction with cell wall polysaccharides in vitro. *Mol Phylogenet Evol* 63: 510–520. doi: 10.1016/j.ympev.2012.02.001
- Vercauteren I, Almeida Engler J de, Groodt R de, Gheysen G (2002) An *Arabidopsis thaliana* Pectin Acetyltransferase Gene Is Upregulated in Nematode Feeding Sites Induced by Root-knot and Cyst Nematodes. *MPMI* 15: 404–407. doi: 10.1094/MPMI.2002.15.4.404

- Wang C, Xu X, Hong Z, Feng Y, Zhang Z (2015) Involvement of ROP6 and clathrin in nodulation factor signaling. *Plant Signal Behav* 10: e1033127. doi: 10.1080/15592324.2015.1033127
- Waters JK, Hughes BL, Purcell LC, Gerhardt KO, Mawhinney TP, Emerich DW (1998) Alanine, not ammonia, is excreted from N(2)-fixing soybean nodule bacteroids. *Proc Natl Acad Sci U S A* 95: 12038–12042.
- Weise SE, van Wijk KJ, Sharkey TD (2011) The role of transitory starch in C(3), CAM, and C(4) metabolism and opportunities for engineering leaf starch accumulation. *J Exp Bot* 62: 3109–3118. doi: 10.1093/jxb/err035
- Welham T, Pike J, Horst I, Flegelakis E, Katinakis P, Kaneko T, Sato S, Tabata S, Perry J, Parniske M, Wang TL (2009) A cytosolic invertase is required for normal growth and cell development in the model legume, *Lotus japonicus*. *J Exp Bot* 60: 3353–3365. doi: 10.1093/jxb/erp169
- Wheeler CT (1969) The diurnal fluctuation in nitrogen fixation in the nodules of *Alnus glutinosa* and *Myrica gale*. *New Phytol* 68: 675–682.
- Wienkoop S, Saalbach G (2003) Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. *Plant Physiol* 131: 1080–1090. doi: 10.1104/pp.102.015362
- Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am J Bot* 91: 1846–1862. doi: 10.3732/ajb.91.11.1846
- Young ND, Udvardi M (2009) Translating *Medicago truncatula* genomics to crop legumes. *Curr Opin Plant Biol* 12: 193–201. doi: 10.1016/j.pbi.2008.11.005
- Zdyb A, Demchenko K, Heumann J, Mrosk C, Grzeganeck P, Göbel C, Feussner I, Pawlowski K, Hause B (2011) Jasmonate biosynthesis in legume and actinorhizal nodules. *New Phytol* 189: 568–579. doi: 10.1111/j.1469-8137.2010.03504.x

## 7. Figure captions:

Figure 1: Experimental setup. *V. faba* seeds (genotype Fuego) were cultivated in four treatment groups: in the absence of N-fertilizer but in the presence of nitrogen fixing bacteria (Fix<sup>+</sup>); in the absence of N-fertilizer but in the presence of bacteria deficient in nitrogen fixation (Fix<sup>-</sup>); in the presence of N-fertilizer and absence of bacteria (N); and in the absence of N-fertilizer and bacteria (0). During the first two weeks all four groups were cultivated under standard conditions (see material and methods). Infection with Fix<sup>+</sup> or Fix<sup>-</sup> bacteria (*R. leguminosarum* bv. *viciae* and *R. leguminosarum* IMA78, Ma *et al.* 1982) or fertilization with nitrogen (only for N) took place at day 15 and plant cultivation continued for another 13 days, during which all four groups were fertilized with 0.5x Hoagland nutrient solution (Hoagland and Arnon 1950) without nitrate twice a week and watered with demineralized water according to demand. Subsequently, nodules and roots were harvested, followed by protein extraction and sample preparation for LC-MS/MS analyses.

Figure 2: Evaluation of experimental setup. A, shoot development (upper panel), individual leaf phenotype (third leaf from the bottom, middle panel) and root development (groups N and 0, lower panel), as well as nodule cross sections (Fix<sup>-</sup> and Fix<sup>+</sup>, lower panel) of four weeks old *V. faba* plants. Fix<sup>+</sup> plants develop red colored nodules, whereas Fix<sup>-</sup> grows more but smaller nodules of tan color. N and 0 do not develop nodules. B, Assessment of nitrogenase activity as deduced from hydrogen production of representative whole root systems of 4 week old plants (sample groups: yellow, Fix<sup>-</sup>; red, Fix<sup>+</sup>; blue, N; and grey, 0). C, C/N ratio (full colored bars, left axis) and shoot length (transparent bars, right axis) of plants. Error bars of shoot lengths indicate standard deviation of >100 individual plants, error bars of C/N ratios indicate standard deviation of three biological replicates (letters indicate p<0.00001 in ANOVA).

Figure 3: Three-dimensional Principal Component Analysis (PCA) of shotgun MS datasets of three biological replicates for each of the four sample groups (Fix<sup>+</sup>, Fix<sup>-</sup>, N, 0). Abundance profiles (LFQ-values) for each identified protein in the unprocessed data obtained from MaxQuant were analyzed by the Perseus software (Tyanova *et al.* 2016). Cubes are color coded to identify the sample groups: yellow, Fix<sup>-</sup>; red, Fix<sup>+</sup>; blue, N; and grey, 0.

Figure 4: Total number of identified plant and bacterial proteins within the four sample groups and number of plant and bacterial proteins with significantly altered abundance. In total, 2695 plant and bacterial proteins were identified across the biological replicates of the four sample groups. Among them, 92 proteins were identified by one modified peptide only, were derived from the decoy database or are common contaminations and were therefore excluded from the analysis. Of the remaining 2603 proteins, 1982 proteins are of plant origin and are highlighted in dark and light blue, whereas the 621 *R. leguminosarum* proteins are shown in dark and light green (left chart). Proteins are considered “quantifiable” when they were identified three times within at least one of the four sample groups and are shown in dark blue and green. The 817 quantifiable proteins consist of 658 plant proteins (dark blue) and 159 bacterial proteins (dark green) and amount to 31.4% of all proteins. Among these quantifiable proteins 301 plant and 55 bacterial proteins were identified by an ANOVA to be significantly increased or decreased in abundance in at least one of the four sample groups (right chart).

Figure 5: Volcano plots of pairwise Student’s t-test results. Four different combinations were tested (Fix<sup>-</sup>/0; Fix<sup>+</sup>/Fix<sup>-</sup>; N/0; N/Fix<sup>+</sup>). Dots in each diagram depict all quantifiable proteins occurring across the two groups. The y-axis shows the negative common logarithm of Student’s p-value, the x-axis the Student’s t-test difference results while the dashed line marks the significance p-value cutoff (p ≤ 0.05). All proteins above this line are considered as differently abundant proteins. Numbers under the dashed line



indicate the total sample size, those above the dashed line tell the number and percentage of proteins matching the significance threshold limit. The larger the distance of the dots from the horizontal center of the graph, the bigger its difference in abundance.

Figure 6: Top-five increased and decreased proteins in the four pairwise comparisons indicated on top. Proteins are presented as circles and their regulation is expressed as fold change (log2). Superscript protein labels refer to supplemental Table 5. Protein names which appear in more than one comparison are highlighted in dark green.

Figure 7: Key metabolic processes in SNF of root nodule cells. For the sake of simplicity and a lack of spatial resolution, all physiological changes are depicted in a single (infected) cell which does not necessarily reflect the *in vivo* situation where alterations may be spread across different cell types. Those enzymes induced by the atmospheric nitrogen fixation are colored in red and those which are a consequence of the bacteria:plant symbiosis are highlighted in green. Sucrose is imported and cleaved by INV or SuSy. INV is the starting point for the ATP dependent starch formation (yellow star). Starch can be stored in the plastid and may be used at a later time point for glycolysis. Originating from SuSy, UDP-glucose forms cell wall components via UXS or precursors for malate production (via oxaloacetate, OAA) by glycolysis. OAA also is a precursor for aspartate. The bacteroids fix atmospheric dinitrogen ( $N_2$ ) and release, in turn for malate,  $NH_4^+$  (ammonium cation) and potentially alanine into the cell.  $NH_4^+$  is incorporated into aspartate to form asparagine, the transported form of reduced nitrogen, by asparagine synthase (AS). Alanine may be converted to pyruvate and glutamate by alanine aminotransferase (AlaAT) in the mitochondria. Serine metabolism also seems important in nodules. Reduced sulfur is assimilated by serine to form methionine and, subsequently, AdoMet. AdoMet is a precursor for several components such as lignin and polyamines. OPR, as well as ANN might be related to nodulation/SNF control mechanisms. AdoMet, S-adenosylmethionine; AlaAT, alanine aminotransferase; ANN, annexin D8; AS, aspartat; Asp, asparagin; AST, aspartate aminotransferase; DHBP, 3,4-dihydroxy-2-butanone 4-phosphate synthase; Glu, glutamat; INV, neutral/alkaline invertase; Leg, leghemoglobin; OAA, oxalacetate; OPR, 12-oxophytodienoate reductase-like protein; PEPC, phosphoenolpyruvate carboxylase; RL, *R. leguminosarum*; SAM, S-adenosylmethionine synthase; SuSy, sucrose synthase; UXS, UDP-glucuronic acid decarboxylase.

Suppl. Figure 1: Testing of protein sequence databases for cross species protein identification. Among the three databases, two are species independent and contain a broad spectrum of proteins from all kingdoms of life (NCBIInr, SwissProt, both of which were downloaded in Jnuauy 2018) while the third database is a custom made, in house database containing the protein sequences of *V. faba*, *R. leguminosarum*, as well as frequent contaminants. One  $Fix^-$  sample could not be analyzed using our bioinformatics platform, most likely due to memory constraints. In all cases, the in-house *M. truncatula* and *R. leguminosarum* database identified more proteins than the two broad species databases, probably due to increased statistical uncertainties inherent to large-scale databases with too many entries.

Suppl. Figure 2: Altered cellular processes in  $Fix^+$  compared to  $Fix^-$  and N. Bars indicate the number of altered proteins associated with a given cellular function. Upper panel:  $Fix^+$  (red bars) opposed to  $Fix^-$  (yellow bars). Lower panel compares  $Fix^+$  with N (blue bars). AdoMet, S-Adenosyl methionine; OXPHOS, oxidative phosphorylation system; TCA, tricarboxylic acid cycle; PPP, pentose phosphate pathway.

Suppl. Figure 3: Pairwise comparisons and meta-comparisons between the four sample groups. The four sample groups differ in two aspects i) nodule formation (true for  $Fix^+$  and  $Fix^-$ , false for N and 0) and ii) nitrogen status (high for  $Fix^+$  and N, low for  $Fix^-$  and 0). Six comparisons are conceivable between the four groups (grey full and dashed arrows) but only the full arrows allow comparison of groups differing only in a single aspect (nodulation or nitrogen status). Pairwise comparisons were thus limited to

these combinations. The comparisons of Fix<sup>+</sup> with Fix<sup>-</sup> and of Fix<sup>+</sup> with N as these allow the identification of proteins specific for symbiotic nitrogen fixation via a meta comparison (MC) approach (MC1). In a second meta comparison (MC2) the nodule forming groups are compared with N and 0 to pinpoint symbiosis related proteins.

Suppl. Figure 4: Mean abundance of core plant and bacteroid SNF enzymes. Significant differences (t-test) between the sample groups are indicated by a star (p<0.05). AS<sup>1</sup>, asparagine synthase (Medtr3g464580.1, Medtr3g464580.2, Medtr3g464580.3); AS<sup>2</sup>, glutamine-dependent asparagine synthase 1 (Medtr5g071360.1-2); AspAT<sup>1</sup>, aspartate aminotransferase (Medtr8g091280.1); AspAT<sup>2</sup> (Medtr1g013050.1); AspAT<sup>3</sup>, aspartate aminotransferase (Medtr3g110065.1-3); GOGAT, ferredoxin-dependent glutamate synthase (Medtr7g089970.1); GDH, NADP-specific glutamate dehydrogenase (Medtr7g085630.1-6); GS<sup>1</sup>, glutamine synthetase (Medtr2g021255.1-2); GS<sup>2</sup>, glutamine synthetase domain protein (Medtr6g071070.1); GS<sup>3</sup>, (Medtr3g065250.1, Medtr5g077950.1); NifD, Nitrogenase (molybdenum-iron) alpha chain (pRL100161); NifH, Nitrogenase iron protein (pRL100162); NifK, Nitrogenase (molybdenum-iron) beta chain (pRL100160); NifN, Nitrogenase FeMo-cofactor scaffold and assembly protein (pRL100158); PEPC<sup>1</sup>, phosphoenolpyruvate carboxylase (Medtr2g092930.1); PEPC<sup>2</sup> (Medtr4g079860.1); PEPC<sup>3</sup> (Medtr2g076670.1-3); PEPC<sup>4</sup> (Medtr8g463920.1-3).

Suppl. Fig. 5: Response of seven proteins commonly found in MC1 and MC2 to SNF activity (Fix<sup>+</sup>/Fix<sup>-</sup>) as well as nodulation and SNF activity (Fix<sup>+</sup>/N). Altered protein abundance are expressed in log<sub>2</sub> (fold) change between the groups as derived from pairwise comparisons.

Table 1: Proteins involved in fixation of atmospheric nitrogen and maintenance of symbiosis as identified by meta-comparisons one and two (MC1/MC2). In total 43 proteins are listed in alphabetical order with their corresponding majority protein ID (minimum half of the peptides shared between the gene identifiers in the group belong to the majority ID) and description. The presence of a given protein ID in MC1 and/or MC2 is indicated by a '+' in the columns MC1 and MC2. Grey, protein biosynthesis-related; pink, SNF related; yellow, sulfur fixation related; blue, amino acid metabolism; bold, found in MC1 and MC2.

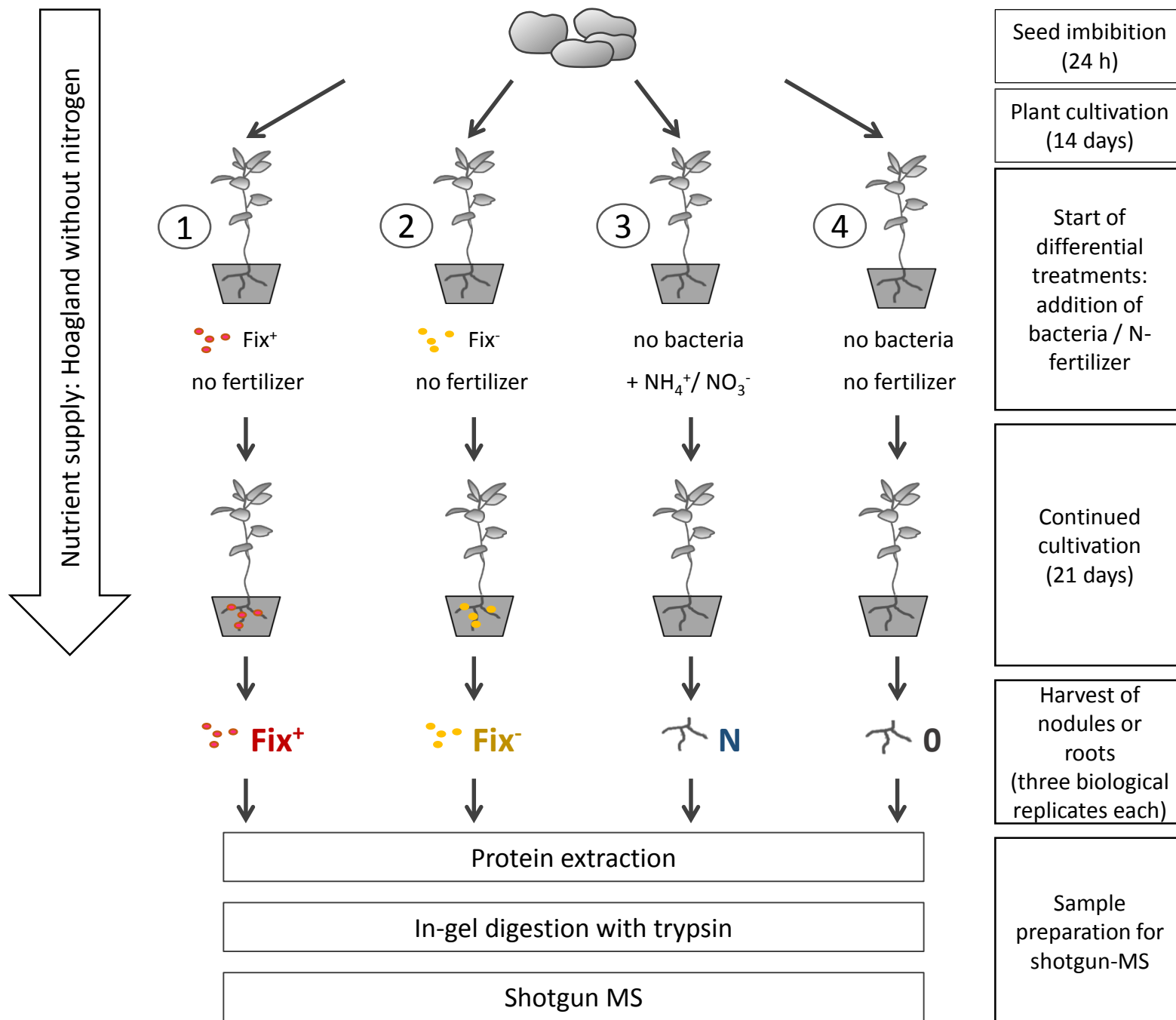
	Majority protein IDs	Description	Acronym	MC1	MC2
1	Medtr7g074570.2;Medtr7g074570.1	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	PGM		+
2	Medtr1g066860.2;Medtr7g100720.1; Medtr1g066860.1;Medtr7g061155.1	40S ribosomal protein S23-1			+
3	Medtr7g107380.1;Medtr1g054310.1	40S ribosomal protein S5-2			+
4	Medtr7g053160.2;Medtr7g053160.1	40S ribosomal protein S6-2			+
5	Medtr5g010025.1;Medtr8g098850.1; Medtr8g098850.2	40S ribosomal S10-like protein			+
6	Medtr5g097200.1;Medtr3g094220.1; Medtr3g094220.2;Medtr5g097200.2	40S ribosomal S26-like protein			+
7	Medtr8g052030.1;Medtr5g018940.2; Medtr5g018940.1	40S ribosomal S4-like protein			+
8	Medtr8g046140.1;Medtr4g080740.1; Medtr3g007700.1	50S ribosomal protein L5P			+
9	Medtr4g059400.1;Medtr1g064500.1	60S ribosomal L12-like protein			+
10	Medtr1g069905.1;Medtr0067s0060.;M edtr7g107420.1	60S ribosomal L21-like protein			+
11	Medtr8g028125.1	60S ribosomal protein L5-2			+
12	Medtr3g093110.1	60S ribosomal protein L6			+
13	Medtr8g023140.1	alanine aminotransferase	AlaAT		+
14	Medtr8g038210.2;Medtr8g038210.1	annexin D8	ANN		+
15	Medtr3g464580.1;Medtr3g464580.2;M edtr3g464580.3	asparagine synthetase [glutamine-hydrolyzing] protein	AS	+	
<b>16</b>	<b>Medtr1g013050.1</b>	<b>aspartate aminotransferase (HC)</b>	<b>AST</b>	<b>+</b>	<b>+</b>

17	Medtr5g011910.1;Medtr4g102170.1; Medtr4g024550.1;Medtr4g024265.1	cytoplasmic ribosomal protein S13			+
18	Medtr1g043040.1	cytoplasmic-like malate dehydrogenase	MDH		+
19	Medtr4g005880.1	D-3-phosphoglycerate dehydrogenase family protein	PHGDH		+
20	Medtr3g105430.1	eukaryotic translation initiation factor 3 subunit B (HC)			+
21	Medtr7g021870.1	eukaryotic translation initiation factor 3h			+
22	Medtr2g065470.3;Medtr2g065470.1; Medtr2g065470.2	glyceraldehyde-3-phosphate dehydrogenase	GAPDH		+
<b>23</b>	<b>Medtr0288s0040.1</b>	<b>glycogen/starch/alpha-glucan phosphorylase family protein</b>	<b>GPase</b>	<b>+</b>	<b>+</b>
24	Medtr4g092620.1	LL-diaminopimelate aminotransferase	DAP		+
<b>25</b>	<b>Medtr1g096140.1;Medtr1g095970.1</b>	<b>neutral/alkaline invertase</b>	<b>INV</b>	<b>+</b>	<b>+</b>
26	Medtr5g006340.4;Medtr5g006340.3; Medtr5g006340.2;Medtr5g006340.1; Medtr1g103690.1	O-acetylserine (thiol) lyase	OAS-TL		+
27	Medtr4g087520.1	O-acetylserine (thiol) lyase	OAS-TL		+
28	Medtr2g098950.1	pfkB family carbohydrate kinase			+
<b>29</b>	<b>Medtr2g076670.3;Medtr2g076670.2; Medtr2g076670.1</b>	<b>phosphoenolpyruvate carboxylase</b>	<b>PEPC</b>	<b>+</b>	<b>+</b>
<b>30</b>	<b>Medtr4g079860.1</b>	<b>phosphoenolpyruvate carboxylase</b>	<b>PEPC</b>	<b>+</b>	<b>+</b>
31	Medtr7g103620.1;Medtr7g103620.2	phosphopyruvate hydratase	Eno		+
32	Medtr3g088220.1	protein disulfide isomerase			+
33	Medtr2g073240.1	rhamnose biosynthetic-like enzyme	RH		+
34	Medtr2g086500.2;Medtr2g086500.1; Medtr3g070930.1	ribosomal protein L1p/L10e family protein			+
35	Medtr4g117470.1	ribosomal protein S8			+

36	Medtr3g084340.1	S-adenosyl-L-homocysteine hydrolase	AdoHcyH		+
37	Medtr2g046710.1	S-adenosylmethionine synthase- like protein	SAM		+
38	Medtr7g110310.1	S-adenosylmethionine synthase- like protein	SAM		+
<b>39</b>	<b>Medtr2g044070.2;Medtr2g044070.1</b>	<b>sucrose synthase</b>	<b>SuSy</b>	<b>+</b>	<b>+</b>
40	Medtr4g124660.4;Medtr4g124660.2; Medtr4g124660.1;Medtr4g124660.3	sucrose synthase	SuSy		+
41	Medtr4g077190.1	sulfite reductase [ferredoxin] protein			+
42	Medtr3g110720.1;Medtr1g106005.1; Medtr4g097830.1	tubulin beta-1 chain			+
<b>43</b>	<b>Medtr4g063600.2;Medtr4g063600.4; Medtr4g063600.3;Medtr4g063600.1</b>	<b>UDP-glucuronic acid decarboxylase</b>		<b>+</b>	<b>+</b>

Figure 1

[Click here to download Colour figure Figure 1.pdf](#)



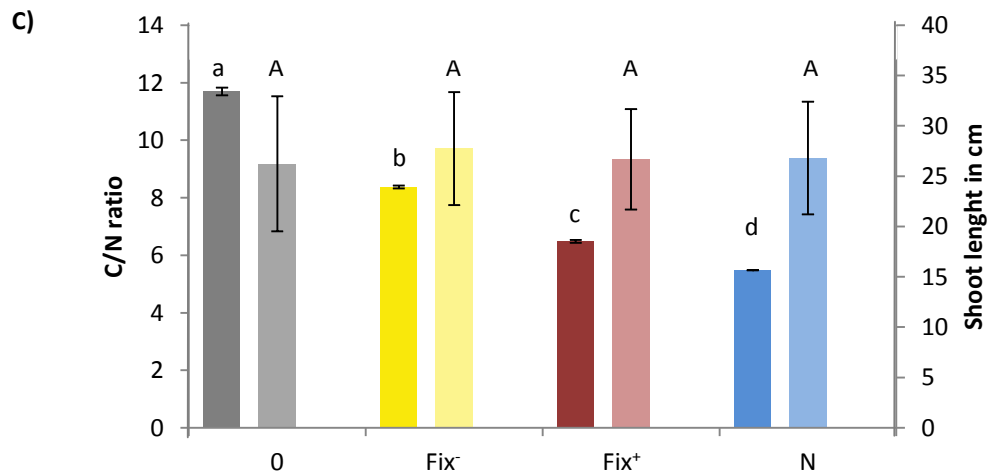
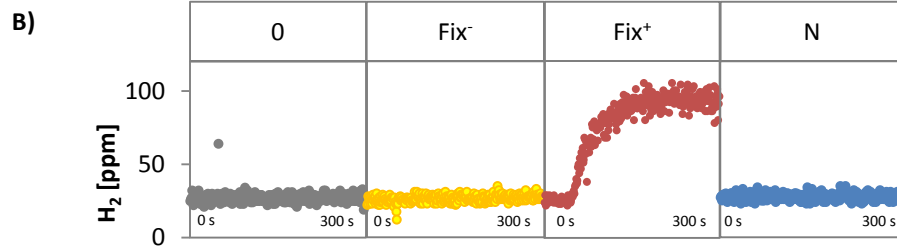
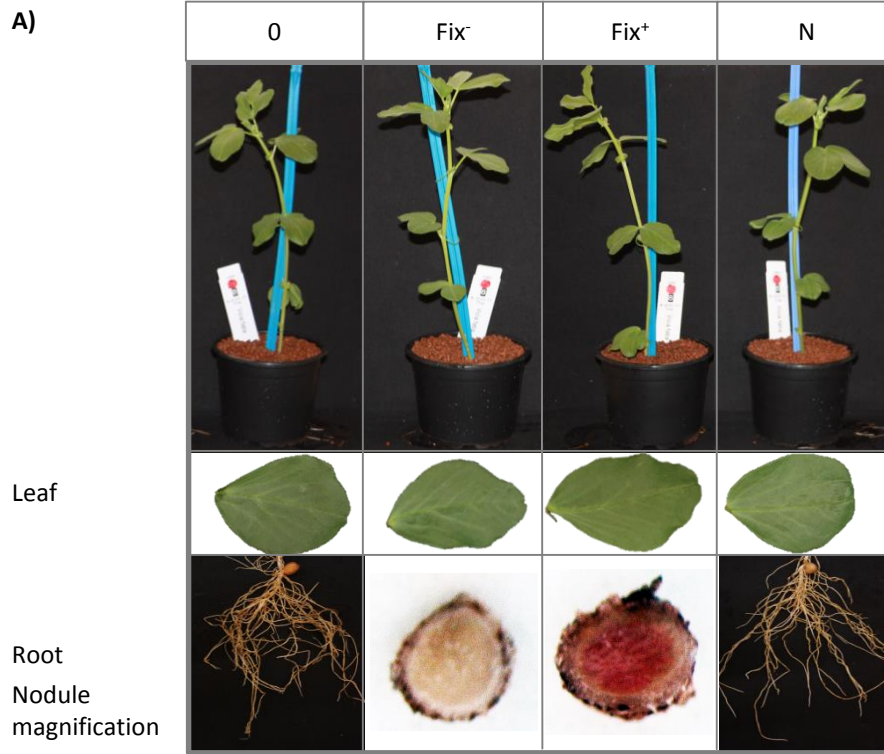
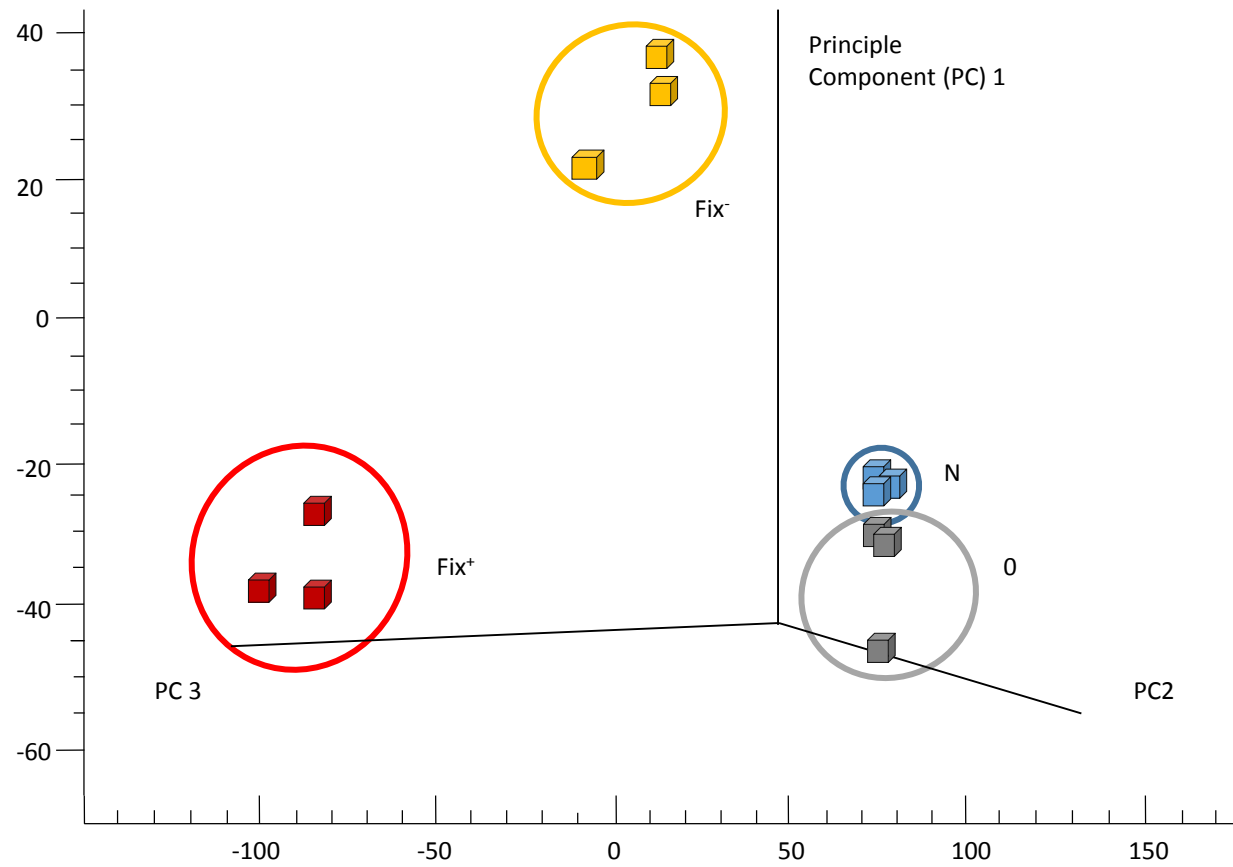
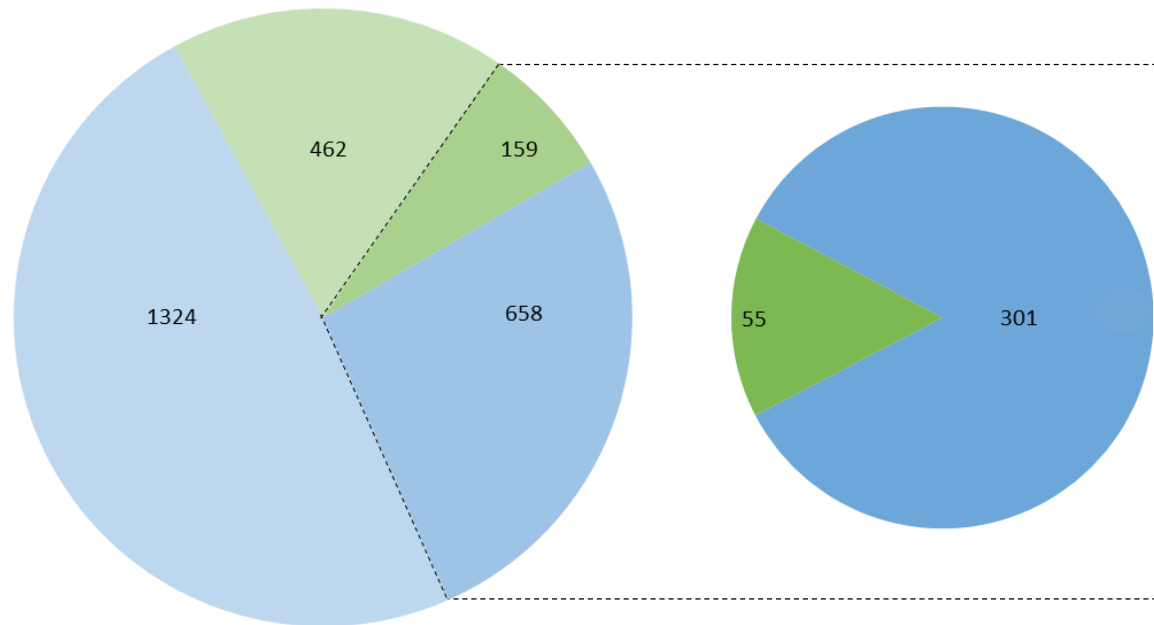


Figure 3







- *Medicago truncatula* (Mt) identified proteins not passing the filter criteria of three positive IDs within at least a single sample group
- *Rhizobium leguminosarum* (RL) identified proteins not passing the filter criteria of three positive IDs within at least a single sample group
- *Medicago truncatula* identified valid proteins (in triple replicate in at least one group)
- *Rhizobium leguminosarum* identified valid proteins (in triple replicate in at least one group)

- *Medicago truncatula* (Mt) significantly changed proteins (valid values, ANOVA  $p < 0,05$ )
- *Rhizobium leguminosarum* significantly changed proteins (valid values, ANOVA  $p < 0,05$ )

Figure 5

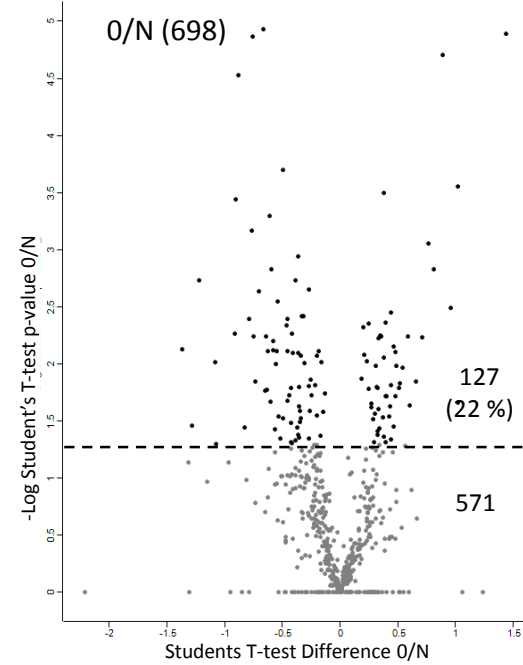
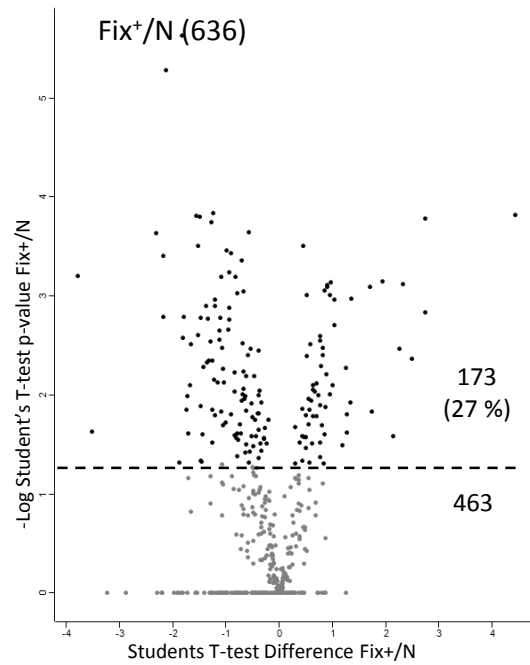
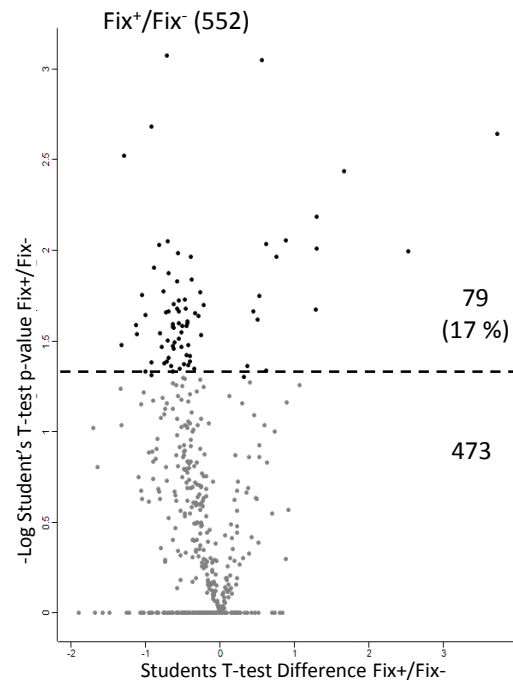
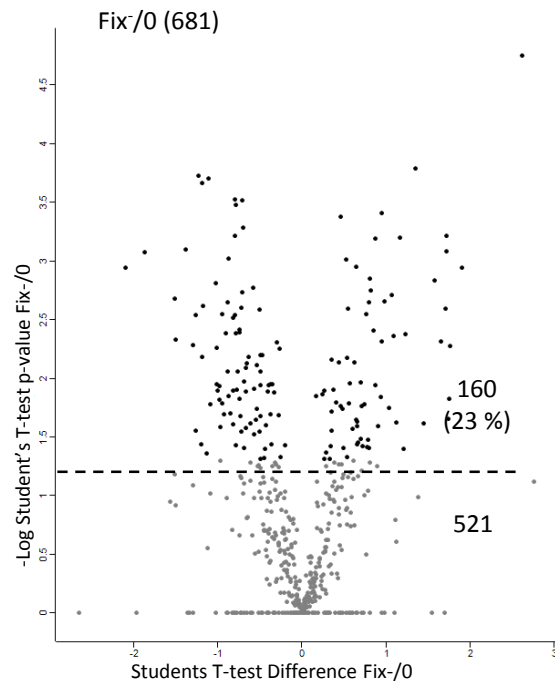


Figure 6

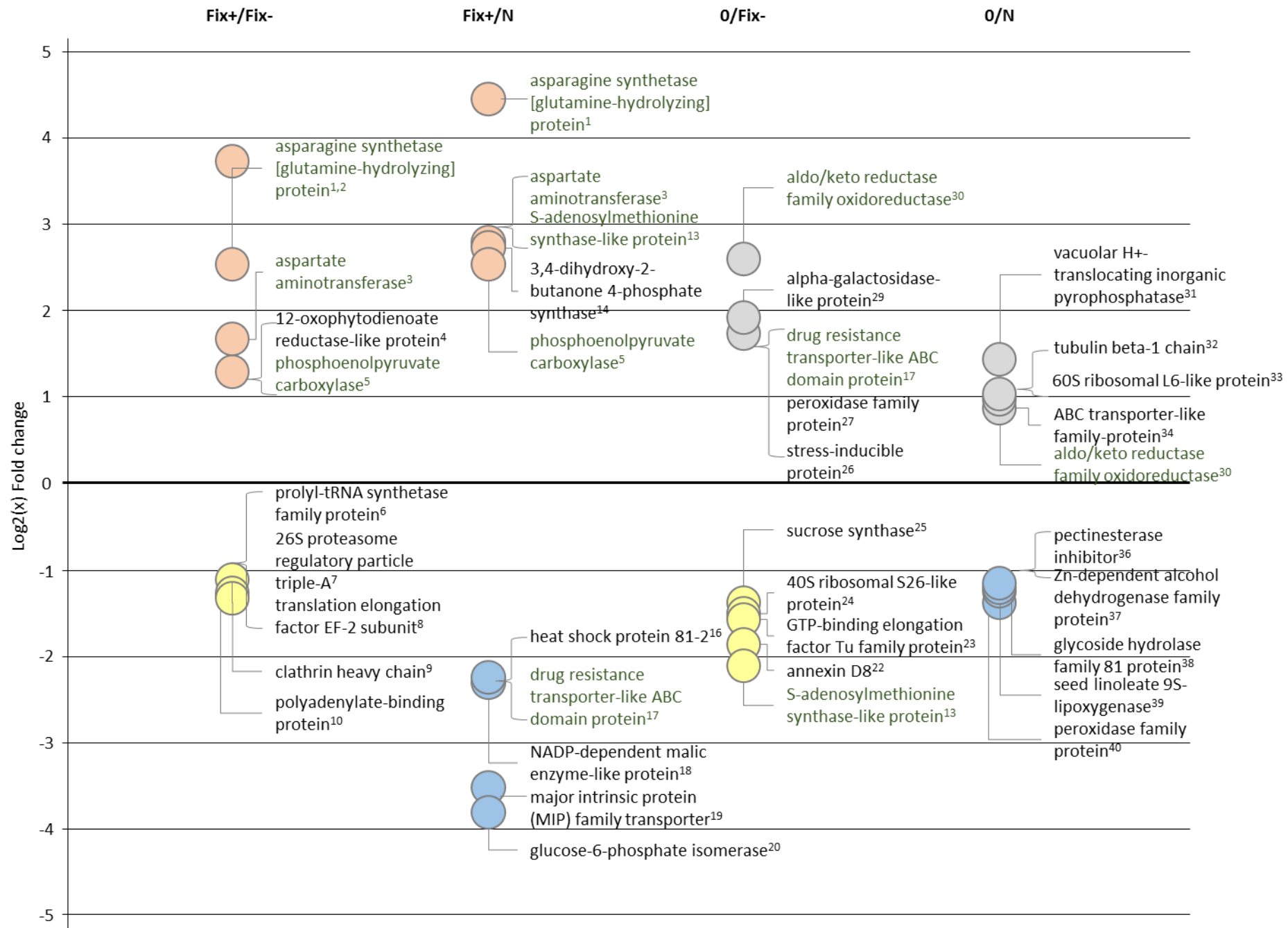
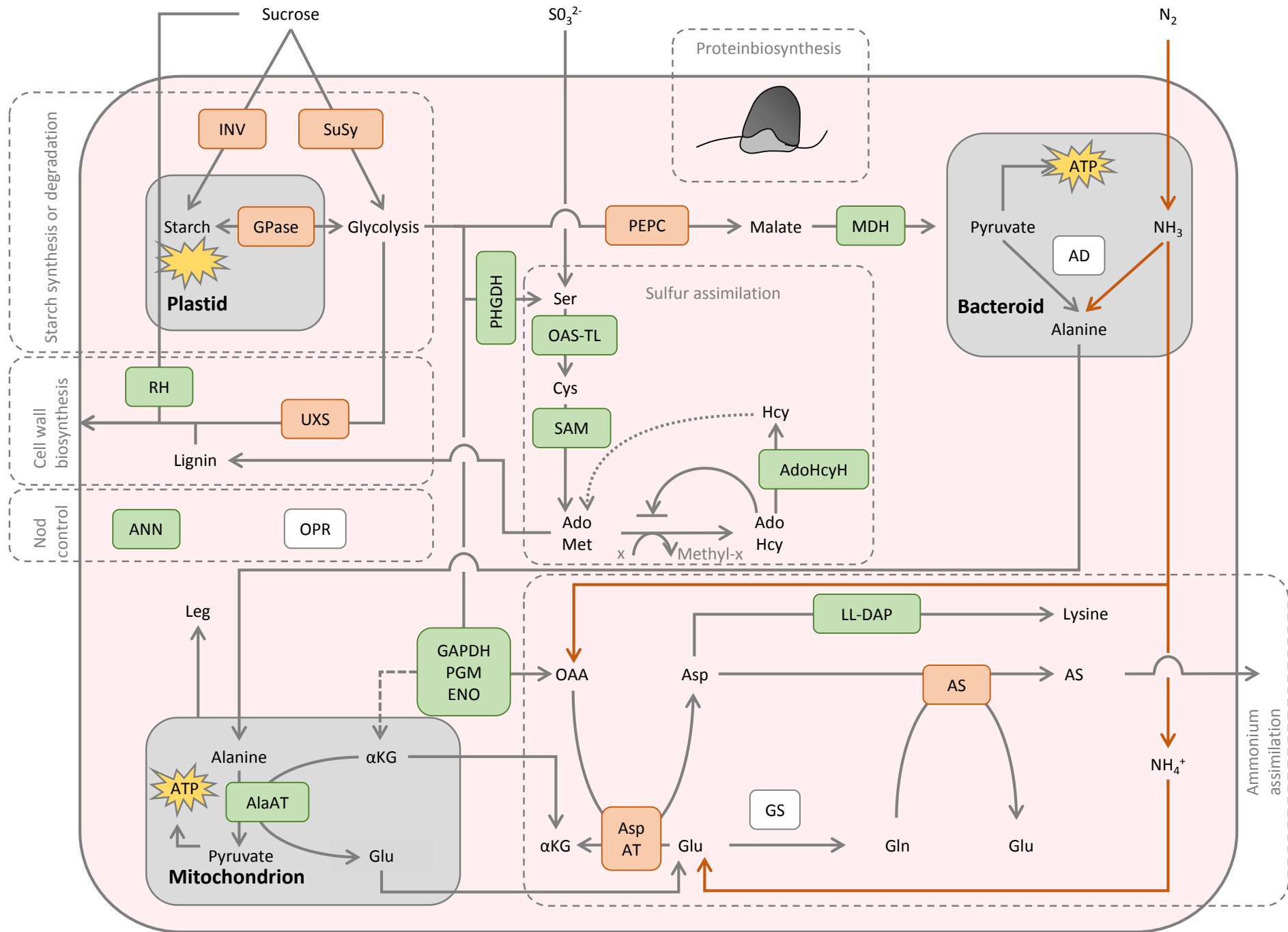


Figure 7



**Author Contribution Statement:**

B.T. conducted all experiments participated in preparation of the manuscript

H-P.B. participated in the planning of experiments and preparation of the manuscript

H.E. planned the experiments and participated in the preparation of the manuscript