

Article

Genes Associated with Biological Nitrogen Fixation Efficiency Identified Using RNA Sequencing in Red Clover (*Trifolium pratense* L.)

David Vlk ¹, Oldřich Trněný ² and Jana Řepková ^{1,*}¹ Department of Experimental Biology, Faculty of Sciences, Masaryk University, 611 37 Brno, Czech Republic² Agricultural Research, Ltd., Zahradní 1, 664 41 Troubsko, Czech Republic

* Correspondence: repkova@sci.muni.cz; Tel.: +420-549-49-6895

Abstract: Commonly studied in the context of legume–rhizobia symbiosis, biological nitrogen fixation (BNF) is a key component of the nitrogen cycle in nature. Despite its potential in plant breeding and many years of research, information is still lacking as to the regulation of hundreds of genes connected with plant–bacteria interaction, nodulation, and nitrogen fixation. Here, we compared root nodule transcriptomes of red clover (*Trifolium pratense* L.) genotypes with contrasting nitrogen fixation efficiency, and we found 491 differentially expressed genes (DEGs) between plants with high and low BNF efficiency. The annotation of genes expressed in nodules revealed more than 800 genes not yet experimentally confirmed. Among genes mediating nodule development, four nodule-specific cysteine-rich (NCR) peptides were confirmed in the nodule transcriptome. Gene duplication analyses revealed that genes originating from tandem and dispersed duplication are significantly over-represented among DEGs. Weighted correlation network analysis (WGCNA) organized expression profiles of the transcripts into 16 modules linked to the analyzed traits, such as nitrogen fixation efficiency or sample-specific modules. Overall, the results obtained broaden our knowledge about transcriptomic landscapes of red clover’s root nodules and shift the phenotypic description of BNF efficiency on the level of gene expression in situ.

Keywords: transcriptome; differentially expressed gene; nodule-specific cysteine-rich peptide; gene duplication



Citation: Vlk, D.; Trněný, O.; Řepková, J. Genes Associated with Biological Nitrogen Fixation Efficiency Identified Using RNA Sequencing in Red Clover (*Trifolium pratense* L.). *Life* **2022**, *12*, 1975. <https://doi.org/10.3390/life12121975>

Academic Editors: Kousuke Hanada and Yoshiteru Noutoshi

Received: 7 November 2022

Accepted: 22 November 2022

Published: 25 November 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Legumes are plants from the family Fabaceae. Consisting of more than 750 genera and 19,500 species, this family makes up about 7% of all flowering plant species [1,2]. This widely distributed family is the third-largest flowering plant family by number of species. From an economic point of view, it is the second-most important after Poaceae (grasses). Due to their great diversity and abundance, legumes include a number of agronomic crops (grain and fodder legumes) and others serve as genetic model organisms (*Medicago truncatula*, *Lotus japonicus*) [3–5]. In the context of sustainable agriculture, many legume species have the potential to establish symbiosis with nitrogen-fixing bacteria and obtain access to nitrogen using biological nitrogen fixation (BNF). BNF is a process whereby plants acquire atmospheric nitrogen through interacting with bacteria capable to convert this molecular nitrogen to ammonium. This symbiotic relationship, within which the plant provides the bacteria with organic compounds used as carbon and energy source and bacteria supply the plant with fixed nitrogen, is a significant competitive advantage for plants in the occupation of nitrogen-poor soil [6].

The Fabaceae genus *Trifolium* includes more than 250 species having cosmopolitan distribution around the world [7,8], with the greatest diversity occurring in the temperate Northern Hemisphere. The economic importance of this genus relates especially to those species used extensively as fodder crops for livestock (*T. pratense*, *T. hybridum*, *T. repens*)

or as a green manure plant to enhance soil fertility [9]. Due to their high content of secondary metabolites, such as isoflavonoids, some species are also being studied for potential pharmacological use [10–12]. Soil enrichment with nitrogen via growing plants utilizing BNF, such as clover, is more sustainable than using synthetic nitrogen-based fertilizers. However, not all genotypes within the fixation-capable species have the same nitrogen fixation efficiency [13]. Plant breeding directed to the enhancement of nitrogen-fixing ability is complicated by the complexity of this phenotypic trait, as an estimated several hundred genes are involved in the nodulation and nitrogen fixation [14,15].

The early phase of plant–bacteria interaction depends upon an early dialog between the host and microbes [16] when bacteria begin to produce their own lipochitoooligosaccharide signals, termed Nod factors, in response to released plant flavonoids [17]. These signal molecules determine the specificity of the interaction itself [18,19] and are recognized by nod factor receptors on the root surface, such as NFR1 and NFR5 [20], and this causes both morphological alterations on the root surface and induction of two root-specific and one systemic pathway. While the systemic reaction, known as autoregulation of the nodulation, controls the number of nodules on the roots depending upon the nodule number already formed and regulation based on the availability of nitrogen from the soil [21], the signal pathways in the roots enable nodulation initiation and nodule formation using calcium-dependent kinases and transcription factors [22] or cytokinins [23].

Nodulating bacteria use infection threads to enter the root [24,25]. The bacteria then penetrate the plant cell by endocytosis as symbiosomes, which gradually differentiate into nitrogen-fixing bacteroids and further into root nodules. The nodules are specialized organs consisting of bacteroids, meristems, and vascular bundles. Nitrogen fixation is enabled by a complex of nitrogenase-nitrate reductase enzymes [26] supported by leghaemoglobin proteins located in the nodules that provide oxygen for respiratory processes into the bacteroid membrane as well as reduce the oxygen concentration inside bacteroids [27]. Two types of root nodules are distinguished: determinate and indeterminate [24]. Meristems of indeterminate nodules remain functional (genus *Medicago* or *Trifolium*); determinate nodules, however, lose their meristematic character in later stages of development (genus *Glycine* or *Phaseolus*) [28].

Indeterminate nodules are usually created by legumes belonging to the inverted repeat-lacking clade. In this clade, bacteria released into the plant cells terminally differentiate into bacteroids that cannot be cultured, show endoreduplication of their genomes, and maintain changes in the cell wall or in expression patterns [29–31]. Many of these changes are processed using small defensin-like peptides, especially nodule-specific cysteine-rich (NCR) peptides, which are typical for legumes with indeterminate nodules and which induce bacteroid differentiation [32]. In the best-studied legume plant, *M. truncatula*, more than 600 NCRs have been identified [33], but there are large differences in the numbers of NCR peptides among various legumes, ranging from just a few NCRs to hundreds [34].

It is estimated today that hundreds of genes with differing impacts on the phenotype are associated with the BNF process [14], and nearly 200 important genes have been identified on model legume plants using both forward and reverse genetics [35]. Originally, chemical and physical mutagens (γ -rays, ethyl methanesulfonate, fast neutron bombardment) were used to enhance the frequency of mutants and to accelerate the discovery of genes connected with BNF on such model legumes as *M. truncatula* [36,37], *L. japonicus* [38], or *Glycine max* [39]. In addition, transposon mutagenesis has broadened the possibilities for obtaining mutant populations by using Ac Transposon (*L. japonicus*; [40]), transfer DNA insertions (*L. japonicus*; [41], *M. truncatula*; [42]), retrotransposon *Tnt1* (*M. truncatula*; [43,44]) or endogenous *Lotus retrotransposon 1* [45,46] for both forward and reverse genetics. Antisense RNA/RNAi methods began contributing to a better understanding of BNF's genetic background at the beginning of the 21st century [47,48], and over the years these have enabled the identification of many genes associated with BNF (see review by Arthikala et al. [49]). In recent years, CRISPR/Cas9 mediated genome editing has been established in legumes such as *G. max* [50], *L. japonicus* [51], *M. truncatula* [52], and

Cicer arietinum [53] and enabled targeted mutagenesis of BNF-associated genes. Moreover, due to the current possibilities of studying genes participating in fixation, there are also papers demonstrating the advantages of using different approaches and combining methods [13,54]. In the field of synthetic biology, there are efforts to introduce nitrogen fixation into plants that have not yet been able to do so, such as cereal crops, by transferring a system of multicistronic genes connected with nitrogen fixation [55,56].

Since their development in the early 21st century, next-generation sequencing technologies have significantly accelerated genome research, identification of gene polymorphisms, and phylogenetic analyses. From that time, too, RNA sequencing has become an important and quite universal tool for transcriptome assembly, quantification of gene expression, identification of spliced variants/fusion genes, and analysis of differentially expressed genes (DEGs) [57–60]. The last of these, identifying gene expression changes between different experimental conditions or different cell populations, is the most popular application of RNA-seq to many and various questions of interest, such as in detecting genes connected with resistance against stress factors [61,62], genes regulating development [63,64], or the genes involved in a symbiotic relationship, such as BNF, where it is used for gene expression analysis in symbiotes [15,65], transcriptome profiling of nodules [66,67], and detection of expression changes during nodule development [68]. The downstream analyses of DEGs aim at functional characterization and annotation of DEGs or finding possible common patterns among them, including enrichment of certain biosynthetic pathways or Gene Ontology (GO) terms.

The red clover genome has been de novo sequenced for the varieties *Tatra* [69] and *Milvus* [70]. In the context of BNF, Ištváněk et al. [69] identified 542 potential NCR peptides and 11 leghaemoglobin genes, and De Vega et al. [70] anchored 22,042 out of a total of 40,868 annotated genes to seven pseudomolecules and constructed a physical map enabling large-scale genomic and phylogenetic studies of traits having biological and agronomic importance. Several studies sequencing red clover transcriptomes have also been published that focus on the stress response [71,72], leaf senescence [73], splice isoforms, fusion gene and non-coding RNA [74,75], and leaf variegation [76]. Owing to the complexity of this trait, and even though red clover has a high level of BNF heritability [77], phenotypic-level understanding of nitrogen fixation is insufficient. Trněný et al. [13] identified candidate genes associated with BNF efficiency as well as polymorphisms associated with BNF and reflecting phenotype variability. Our knowledge of the genetic variation within BNF must be expanded on the level of gene expression and transcriptomic analysis.

The goals of our experiments, therefore, were to obtain red clover populations with different levels of nitrogen fixation and perform differential gene expression analysis using RNA sequencing of root nodules of red clover genotypes with contrasting nitrogen fixation levels. The annotation of differentially expressed genes between genotypes with high and low nitrogen fixation efficiency was directed to finding their functions and thereby allowing their connection with biosynthetic pathways associated with BNF. NCR peptides in nodule transcripts were identified and characterized to evaluate their connection to BNF efficiency, and evolutionary analyses were aimed at revealing the roles of different modes of duplicated genes in BNF.

2. Materials and Methods

2.1. Plant Materials and Growth Conditions

In total, 378 genotypes of two diploid (*Start*, *Global*) and three tetraploid (*Tatra*, *Tempus*, *Kvarta*) *T. pratense* varieties were grown in 2019. These genotypes were the progeny of 16 plants (8 with high and 8 with low BNF efficiency) evaluated in 2017 [13]. This progeny was used to observe the selection for nitrogen capacity in field conditions.

The red clover seeds were scarified and germinated on wet perlite. Germinated seeds were then sown in individual pots filled with perlite and inoculated with rhizobia by adding 1 mL of *Rhizobium leguminosarum* bv. *trifolii* inoculum provided by the Crop Research Institute (Prague, Czech Republic). Different rhizobia strains were used for diploid and tetraploid

varieties according to the recommendations of the collection's curator. Plants were grown in a greenhouse within individual pots filled with perlite regularly watered with nitrogen-free nutrient solution as described earlier [13]. Before evaluating nitrogen fixation efficiency, the fresh mass of the shoots and roots of analyzed plants was measured in milligrams using analytical balances after pulling them out of the pots and removing perlite.

2.2. Evaluation of Nitrogen Fixation Efficiency, Sample Preparation, and RNA-Sequencing

Nitrogen fixation efficiency was evaluated by acetylene reduction assay (ARA) while measuring nitrogenase activity in individual plants approximately 100 days after sowing [78]. ARA was performed on the sheared roots with nodules placed in a jar with added acetylene on a total of 378 plants. The results were expressed as ethylene molar concentration ($\mu\text{mol}/\text{mL}$) in a jar after 0.5 h of incubation using equations according to Unkovich et al. [79]. The ethylene level was related to particular plants as we assessed whole genotypes. After ARA, the roots were again planted in pots to let the plants regenerate, and root nodules were sheared from chosen genotypes 14 days after ARA. Eight red clover genotypes chosen for RNA sequencing were represented by two contrasting groups according to their BNF rates (low \times high BNF efficiency) and by two diploid and two tetraploid plants in both groups. Three biological replicates were collected from each genotype, and 20–30 mg of root nodules were flash-frozen in liquid nitrogen for each replicate.

RNA isolation was performed for 24 samples chosen for RNA sequencing (3 biological replicates for each of the 8 chosen genotypes) using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol, and DNase treatment was performed using TURBO DNA-free™ Kit (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity was checked on a 1.2% agarose gel and fragment analyzer system (Agilent, Santa Clara, CA, USA), and RNA concentration was quantified by the Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Library preparation and RNA sequencing were performed at the Genomics Core Facility of CEITEC MU (Brno, Czech Republic). RNA-seq library was prepared from total RNA using poly(A) enrichment of the mRNA with NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA), and the library was sequenced for 75-bp reads with paired-end sequencing on an Illumina NextSeq 500 instrument (San Diego, CA, USA) using the NextSeq 500 High Output Kit.

2.3. Bioinformatic Analysis

The basic characteristics of the obtained reads were checked in FastQC v0.10.1 [80]. Reads were qualitatively filtered, contaminant-filtered, trimmed using scripts filterbytile.sh and bbduk.sh, which are part of the BBmap scripts [81], and then aligned onto the reference genome of *T. pratense* variety *Milvus* [70] using the STAR aligner [82] and SAMtools [83] for BAM file indexing. Quality control of aligned reads was performed with QoRT [84] while using gffread, part of the Cufflinks package [85], for gff3/gtf format conversion. Aligned reads were quantified using gene-based read counting with FeatureCounts [86]. Normalizing read counts and DEG analysis were performed using the DESeq2 package in R [87] along with RStudio [88]. Prior to DEG analysis, the raw read counts were first normalized for sequencing depth differences using DESeq2 size factor and log2 transformed; the similarity of gene expression patterns in biological replicates was checked using the Pearson correlation coefficient (r), hierarchical clustering (distance measure $d = 1 - r$; complete linkage) and principal component analysis (PCA). The following DEG analysis evaluated contrast among BNF samples, with low BNF set as the default state. Genes with log2 fold change >1 and adjusted p -value < 0.05 were considered as differentially expressed.

DEGs were annotated using the reference annotation file [70] extracted from Phytozome [89] and the LegumeIP database [90]. Unannotated DEGs were further functionally annotated using blastp ver. 2.6.0+ [91] (e-value 1–10) against several databases: TrEMBL and Swiss-Prot [92], all predicted proteins from Phytozome [89], the annotated files from *T. pratense* variety *Tatra* [69] and from *T. subterraneum* [93]. For each analyzed sequence, the

hit with the highest score and the lowest e-value was chosen as an annotation. The custom Python scripts were used for filtering, extracting, and merging data during annotation. Annotation terms (Gene Ontology [GO], Kegg orthology [KO]) were assigned to proteins by manually transferring the terms from Swiss-Prot or Phytozome. Furthermore, the Blast2Go pipeline [94] in OmicsBox [95] was used for improving the functional annotation of DEGs. Using Blast2GO, protein sequences of DEGs were queried against the NCBI non-redundant protein sequences using blastp (e-value 10^{-3}). InterProScan in Blast2Go was carried out to retrieve the domains and motifs. GO terms connected with the obtained BLAST hits were retrieved and GO annotation was performed with Blast2GO (e-value hit filter: 10^{-6}). Corresponding GO terms associated with InterProScan results were transferred to the sequences and merged with already existing GO terms. Finally, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis and GO enrichment (*padj* Benjamini-Hochberg < 0.05) were carried out using Blast2Go.

Genes without experimental evidence (unconfirmed) were identified by comparing original annotation files from De Vega et al. [70] and annotation of *T. pratense* downloaded from Phytozome [89]. (Phytozome excluded any gene without experimental evidence). Extracted unconfirmed genes were annotated as described above.

2.4. Experimental Verification of Sequencing Data

Verification of acquired sequencing data was performed using quantitative polymerase chain reaction (qPCR). Ten genes with detectable expression according to RNA-seq were randomly chosen for qPCR. Primers were designed using the Primer3 tool (<http://primer3.ut.ee>, accessed on 28 October 2021; [96]); their specificity was checked using Blast+ (ver. 2.8.1; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 28 October 2021; [97]) with the *T. pratense* *Milvus* genome as a database [70]. RNA samples were isolated, and DNase treated as described above, reverse transcription was performed using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems/ThermoFisher, Foster City, CA, USA). Approximately 2 μ g of prepared cDNA was taken as the template for qPCR using SYBR[®] Select Master Mix (Applied Biosystems) and primer pairs shown in Supplementary Table S1. Cycling conditions were set as follows: 2 min at 50 °C, 2 min at 95 °C followed by 40 cycles of 3 s at 95 °C and 30 s at 60 °C. The *UBQ* (ubiquitin) gene was chosen as the reference gene. Sample cycle threshold (Ct) values were standardized for each template using the reference gene. The sample with the lowest expression was used as a calibrator and the $2^{-\Delta\Delta C_t}$ method was used to analyze the relative changes in gene expression. Three replicates per sample were used to ensure statistical credibility.

2.5. Identification of NCR Peptides in *T. pratense*

Protein sequences encoded by genes detected in nodules across all analyzed samples were inspected using a custom Python script searching conservative NCR-like structure according to Maróti et al. [98] and length < 150 bp. Sequences identified as NCR-like were merged with those detected as DE. Then, NCR-like sequences were analyzed using blastp ver 2.6.0+ (e-value 10^{-5}) [97] against the non-redundant protein sequences database [99] and conserved regions were detected using NCBI Conserved Domains Database [100]. Signal peptides were searched using the SignalP tool [101], and subcellular localization was identified using DeepLoc [102]. Physicochemical parameters were computed using ProtParam [103], and 3D structures were predicted using Phyre2 [104] and trRosetta [105].

2.6. Gene Duplication Analyses

Different modes of gene duplication in *T. pratense* were identified using Plant Duplicate Gene Database (PlantDGD [106]). The custom Python scripts were used for filtering, extracting, and merging PlantDGD data with RNA-seq results. In this way, each gene expressed in the nodule was inspected for possible duplication events and each gene was classified according to its own duplication mode (whole genome duplication [WGD], tandem, proximal, transposed, dispersed, non-duplicated). The distribution of different

duplication modes was checked among DEGs, and this distribution was compared with the global distribution of duplicate modes among all genes expressed in nodules. Statistical significance was calculated using Pearson's *chi*-squared test for global distribution and Fisher's exact test for particular duplicate modes in R while using RStudio [88].

The expression diversity between duplicated pairs was computed for those duplicated pairs in which at least one gene copy was identified as differentially expressed. The Pearson correlation coefficients (r) between the expression profiles of analyzed gene pairs were calculated using Python's Numpy module. First, a cut-off r -value was determined below which two duplicated gene pairs were considered divergent in expression. Then, 10,000 gene pairs were randomly selected and r -values for their expression profiles were calculated. Overall, 95% of the r -values for randomly chosen gene pairs were $r < 0.67$. That means the gene pairs with $r \geq 0.67$ have significantly conserved expression levels at $\alpha = 0.05$. Therefore, the gene pairs with $r < 0.67$ were considered to have diverged in expression.

2.7. Expression Network WGCNA Analysis

Counts of genes were transformed by variance stabilizing transformation using the DESeq2 R package [87]. Data normalization was also part of the transformation. For further WGCNA analysis [107], we filtered out genes having occurred in less than 3 samples with transformed counts greater than or equal to 10. From a total 40,868 genes, 25,873 genes met these criteria and were used for WGCNA analysis. The adjacency matrix is based upon a topological overlap matrix of "signed" type. Modules were identified in one block using soft threshold power = 6, network type "signed hybrid", minModuleSize = 20, mergeCutHeight = 0.45, and deepSplit = 3. A gene co-expression network was drawn using Cytoscape 3.9.0 with edge included by adjacency threshold = 2 [108]. GO enrichment testing was performed by hypergeometric test using the BiNGO 3.0.3 app of Cytoscape [109]. Multiple testing correction was made using Benjamini Hochberg FDR [110]. Considered as enriched were those GO terms having false discovery rate (FDR) p -values of enrichment test less than 0.05. The functional content of GO enrichment terms of correlated groups was summarized through clustering of GO terms of genes within modules using GOMCL, a toolkit to cluster, evaluate, and extract non-redundant associations of GO-based functions [111], according to default parameters.

3. Results

3.1. BNF Efficiency Measurement

Overall, 378 genotypes of the 16 parental plants of two diploid and three tetraploid *T. pratense* varieties were evaluated using ARA. The characteristics of intrapopulation diversity of BNF efficiency among accessions are demonstrated in Figure 1 (partially published in Trněný et al. [13]).

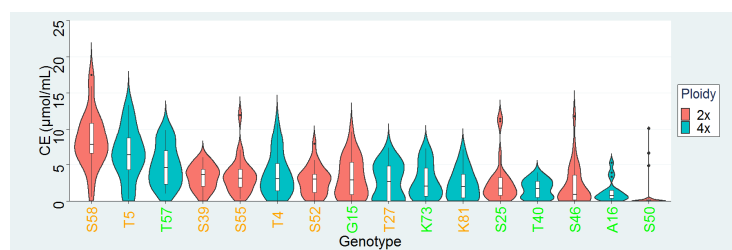


Figure 1. Intrapopulation distribution of BNF efficiency among accessions evaluated prior to RNA sequencing. Varieties indicated by letter: A—*Tatra*, K—*Kvarta*, S—*Start*, T—*Tempus*, G—*Global*. Numerals following the letter indicate number of the parent plant from 2017. Inside the violin plots, median and interquartile ranges indicated by boxplot, minimum and maximum by whiskers, outliers by dots above boxplot. Accessions are ordered by median values of nitrogen fixation. Orange labels—progeny of strong fixators from 2017, green label—progeny of weak fixators from 2017. On the y-axis, measured BNF efficiency is expressed as concentration of ethylene in μmol/mL.

Figure 1 shows that outliers with several times greater BNF efficiency compared to others exist in several populations. The interpopulation variability was high because plants are progeny of either strong- or weak-fixing parental plants and thus influenced by selection. The distribution of intrapopulation fixation level differed slightly between strongly and weakly fixing varieties. In weak-fixing populations, the largest proportion of plants had low fixation efficiency ranging from near zero and up to the mean level, while a smaller part of genotypes had higher efficiency. In strongly fixing populations, the largest proportion of plants had fixation efficiency around the mean level and these proportions decreased to both sides from the mean. Generally, the progeny of plants with high BNF efficiency from 2017 datasets (orange label) had significantly higher BNF efficiency than did the progeny of plants with low nitrogen fixation (green label) (one-tailed Mann–Whitney U test; $p = 2.528 \times 10^{-8}$).

Together with ARA evaluation, analyzed plants also had been measured for their plant fresh mass (Figure 2). The analyzed plants were planted with no exogenously supplied nitrogen, which means that nitrogen availability was one of the crucial factors influencing plant growth and the plants were forced to obtain nitrogen using BNF under these conditions. Figure 2 demonstrates that plants with higher BNF efficiency tended to achieve greater fresh mass and this tendency was clearer in tetraploid plants.

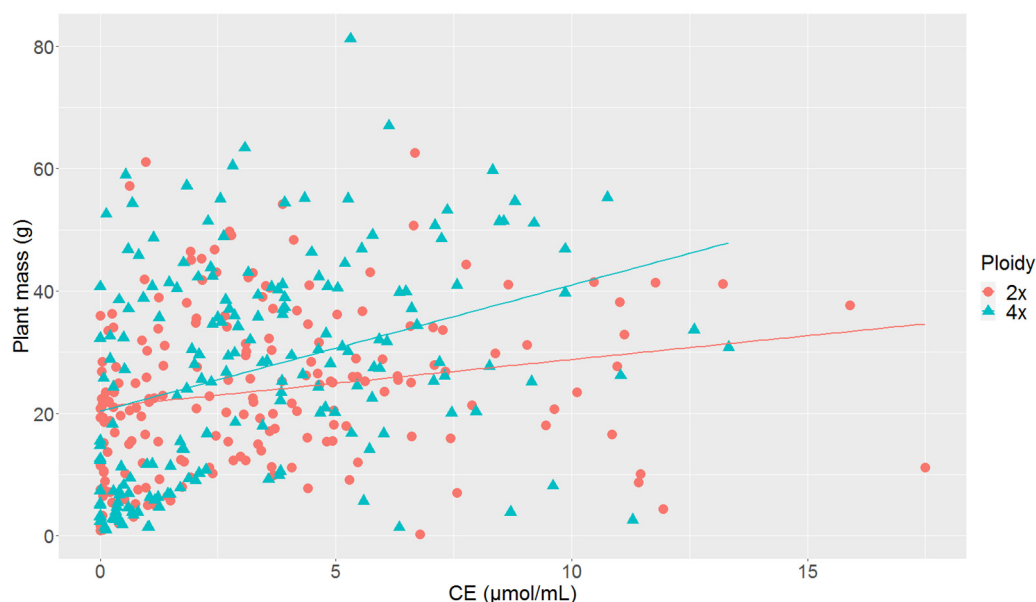


Figure 2. Dependence of measured BNF efficiency expressed as ethylene molar concentration (CE) on plant fresh mass with linear regression lines. Ethylene molar concentration is expressed as ethylene concentration in $\mu\text{mol/mL}$.

3.2. Differential Gene Expression Analysis

As sequenced on the Illumina NextSeq 500 instrument, the transcriptome of 24 samples represented four accessions with high BNF efficiency and four accessions with low BNF efficiency in biological triplicates selected by ARA. Diploid accessions with high/low BNF efficiency were S46/12, S46/7, S55/6, and S25/17. Tetraploid accessions with high/low BNF efficiency were A16/17, A16/21, T57/20 and T57/22. One of the twenty-four samples was poorly sequenced due to an unidentified problem during sequencing, and this sample was omitted from the following analysis. From the total 23 samples analyzed, 428.9 million pair-end reads were assigned to the samples with an average of 18.6 million per sample (Supplementary Figure S1), and the total output was 65.2 Gb. Barcodes were not recognized in 27.1 million reads and these samples were omitted from the analysis.

After mapping sequencing reads to the reference genome, alignment control was performed to evaluate the mapping rate and quality of the sequencing reads. In general, the

number of uniquely mapped reads was about 80%, as shown in Supplementary Figure S1, and gene-body coverage showed a general abundance of reads across transcript bodies to be equal, with a median of around 50 and thus showing no possible 3' or 5' biases.

Prior to DEG analysis, the similarity of gene expression patterns of biological replicates was checked using Pearson correlation coefficient (r), hierarchical clustering, and PCA. Pearson correlation coefficients of rlog normalized biological replicates' r -values were >0.99 in seven out of eight biological replicates (Supplementary Figure S2). A dendrogram of their hierarchical clustering is shown in Supplementary Figure S3. The individual biological replicates are well separated, and the accessions with low and high BNF form separate groups in the dendrogram. Similarly, the PCA plot (Supplementary Figure S4) shows those accessions with different levels of nitrogen fixation to be well separated from one another and that biological replicates are grouped together.

DEG analysis was performed between sequenced contrasting BNF accessions and low BNF was set as the default state. As a result, 37,415 expressed genes were revealed in the sequenced nodules, and 8713 genes (23%) were low counts with average expression < 2 reads. Overall, 491 genes were identified as differentially expressed (\log_2 fold change > 1 , $\text{padj} < 0.05$ were set as thresholds) and 368 genes were overexpressed compared to low-BNF accessions while 123 genes were underexpressed (Figures 3 and 4). A global view of the relationship between expression change and average expression strength is demonstrated by the MA plot in Supplementary Figure S5. The list of DEGs is attached in Supplementary Table S2.

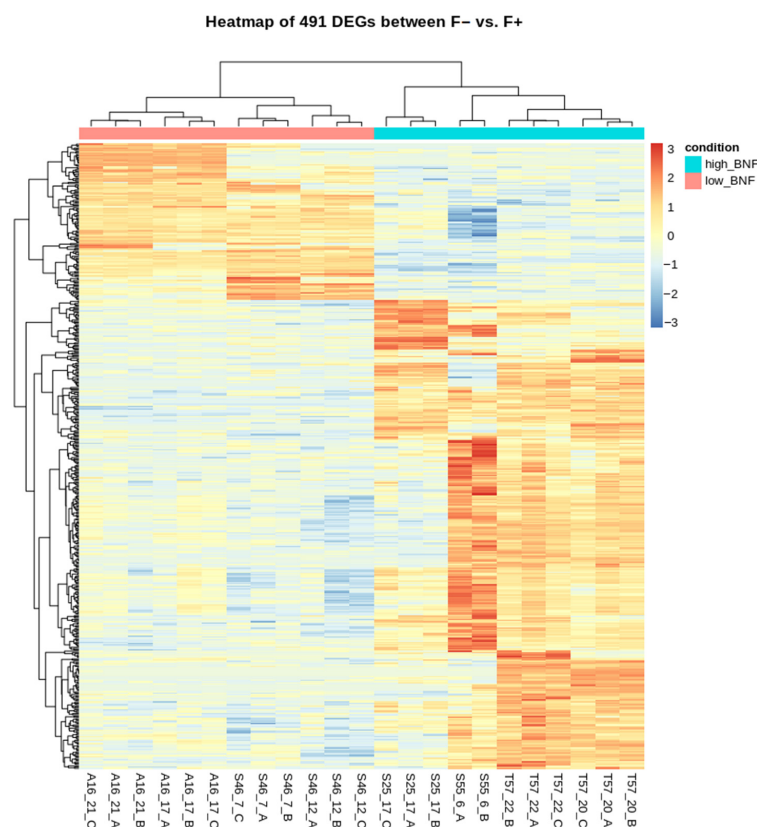


Figure 3. Heatmap of rlog-transformed read counts for 491 differentially expressed genes. Genes are sorted according to hierarchical clustering and the read count values are scaled per row. High-BNF samples are on the right side of the plot (blue line), and low BNF are on the left (red line).

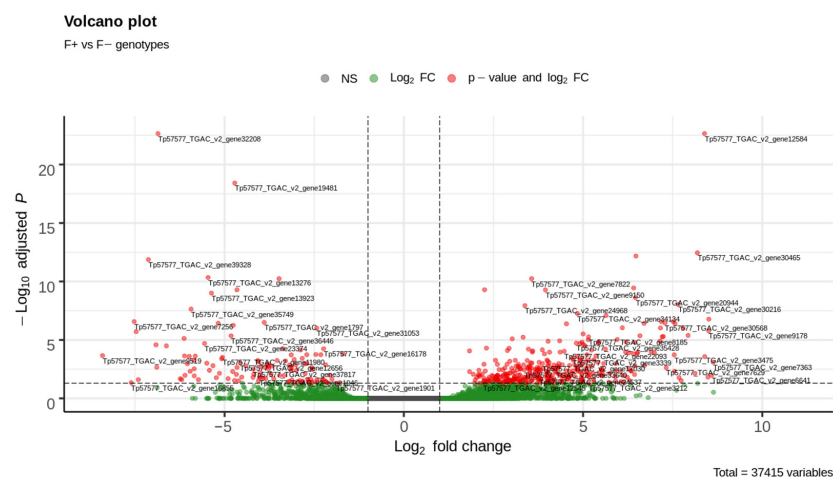


Figure 4. Volcano plot for 491 differentially expressed genes (red dots). The vertical dashed line indicates threshold \log_2 fold change > 1 , the horizontal dashed line indicates threshold $padj > 0.05$.

3.3. Annotation of Differentially Expressed Genes

DEGs were first annotated using an annotation file from sequencing of the *Milvus* variety of *T. pratense* [70]. Of 491 DEGs, 375 (76%) were assigned to at least one category of functional annotation (Pfam, PANTHER, KOG, ec, KO, GO) and the remaining 116 genes were without any hit. Of 116 unannotated genes from the reference annotation file, 71 were partly annotated for at least one of the aforementioned categories using publicly available databases as described in the Methods. Functional annotation was further improved using Blast2Go, and the same pipeline was used for GO enrichment and the inclusion of DEGs in biosynthetic pathways (Table 1). Finally, 446 of 491 DEGs (91%) were successfully annotated for at least one of the functional annotation categories (Pfam, PANTHER, KOG, ec, KO, GO).

Table 1. Top 10 pathways with highest numbers of enzymes encoded by DEGs.

| Pathway | Pathway ID | Enzymes in Pathway |
|---|------------|--------------------|
| Sesquiterpenoid and triterpenoid biosynthesis | map00909 | 12 |
| Pentose and glucuronate interconversions | map00040 | 6 |
| Pantothenate and CoA biosynthesis | map00770 | 5 |
| Steroid hormone biosynthesis | map00140 | 5 |
| Glycerolipid metabolism | map00561 | 5 |
| Galactose metabolism | map00052 | 4 |
| Amino sugar and nucleotide sugar metabolism | map00520 | 4 |
| Cysteine and methionine metabolism | map00270 | 4 |
| Starch and sucrose metabolism | map00500 | 4 |
| Phenylpropanoid biosynthesis | map00940 | 4 |

3.4. Experimental Evidence of Unconfirmed Genes

The original annotation file of *T. pratense* variant *Milvus* [70] used for DEGs contains genes both experimentally confirmed and unconfirmed during assembly and annotation. Unconfirmed genes are those for which the authors could find no transcripts supporting these genes using RNA-seq. During the annotation of 37,415 genes detected in nodules, 863 of these were originally unconfirmed. The expression of the most abundant unconfirmed gene was more than 13,000 mapped reads, 235 genes were considered as low counts with average expression < 2 reads, and 11 of these 863 genes were identified as DE. The expression distribution of the originally unconfirmed genes is demonstrated in Supplementary Table S3.

Of these 863 genes, 596 (69%) were at least partly annotated using GO terms and 690 of them (80%) had at least one hit using InterProScan. The most frequently appearing GO bp terms in the originally unconfirmed genes are listed in Supplementary Table S4. The list of originally unconfirmed genes is attached in Supplementary Table S5.

3.5. Experimental Verification of Sequencing Data

Ten DEGs with detectable expression according to RNA-seq were verified using qPCR on both high and low-nitrogen fixing samples of *T. pratense*. *UBQ* was selected as the reference gene because it showed the most stable expression across different *T. pratense* samples during analogous experiments. For each analyzed gene, primer pairs (Supplementary Table S1) were functional and created a detectable fluorescence signal in at least one of the analyzed samples. The expressions of the genes analyzed using qPCR were in agreement with those from the RNA-seq (Figure 5A–J). Panels A–F and J in Figure 5 show a very good correlation in expression between RNA-seq and qPCR. Panels G and H in Figure 5 show weak expression in qPCR (high Ct value and low fluorescence) that hinders verification while panel I shows that only one genotype had a detectable expression of the analyzed gene for both RNA-seq and qPCR.

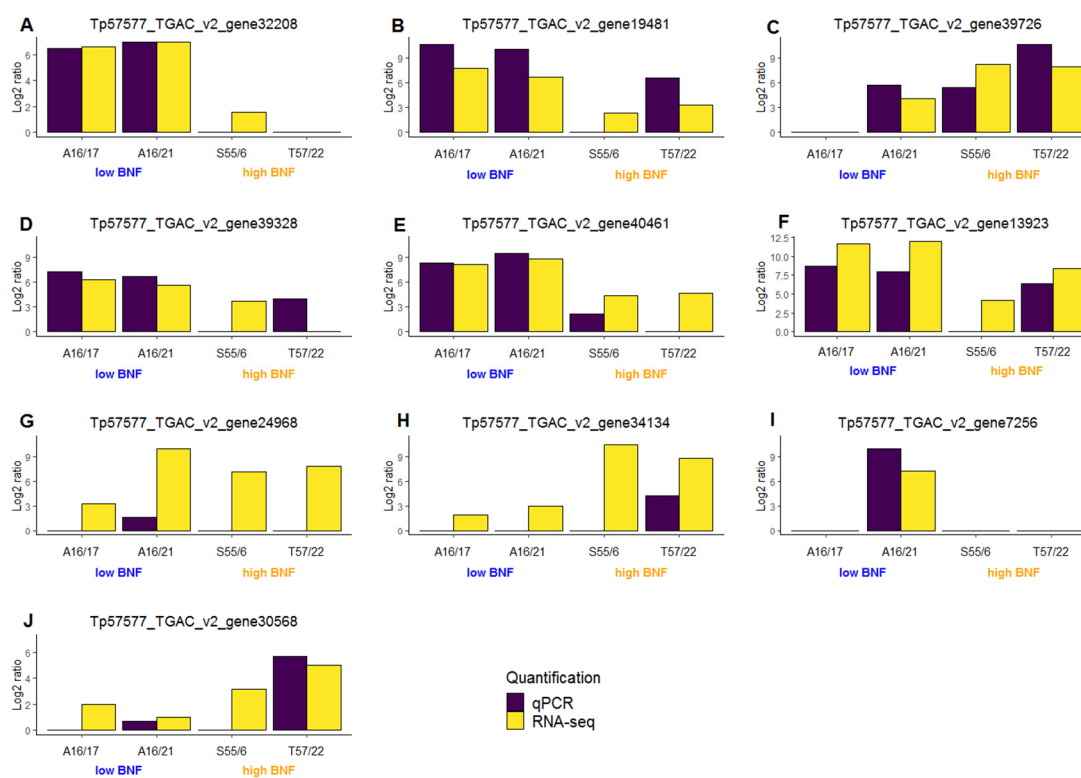


Figure 5. Verification of sequencing data using qPCR for 10 genes (subfigures A–J) in *T. pratense* genotypes with low and high BNF. All samples were analyzed in triplicate and data are presented as means. The x-axis identifies analyzed samples, and the y-axis shows relative expression in the log₂ ratio.

3.6. Identification of NCR Peptides

Protein sequences of 37,415 genes detected in nodules across all samples were inspected for NCR-like structure [98]. This structure, with four or six conserved cysteines, was found in 33 sequences with a length of <150 bp. Four of these were differentially expressed between high and low nitrogen fixing samples while thirteen were marked as low counts and were omitted from the DEG analysis. The sequences were annotated using BLAST and analyzed for signal peptides, subcellular localization, conserved amino acids and domains, and physicochemical parameters to confirm or exclude their inclusion into the NCR peptides. Finally, 4 sequences out of the 33 *T. pratense* sequences detected in nod-

ules with NCR-like structure and length >150 bp were confirmed as NCR peptides (Table 2). The others were either unconfirmed or ruled out as NCR peptides using subsequent in silico analyses.

Table 2. Confirmed NCR peptides detected in nodules of *T. pratense* using RNA sequencing. *P*-value < 0.05 indicates differentially expressed gene.

| Sequence ID | BLAST | Domains | Localization | <i>P</i> -Value |
|---------------------------|-----------------------|---|------------------------------|-----------------------|
| Tp57577_TGAC_v2_gene38999 | Defensin-like protein | Gamma-thionin family, Knot1, Knottin fold | $p = 0.9999$ – Extracellular | 1.33×10^{-8} |
| Tp57577_TGAC_v2_gene36456 | Defensin-like protein | Knot1 | $p = 0.9993$ – Extracellular | 0.133 |
| Tp57577_TGAC_v2_gene7879 | Defensin-like protein | | $p = 1$ – Extracellular | 0.894 |
| Tp57577_TGAC_v2_gene30230 | Defensin-like protein | | $p = 1$ – Extracellular | 0.716 |

3.7. Gene Duplication Analysis within Differentially Expressed Genes

Four hundred ninety-one genes identified as differentially expressed between samples with high and low BNF samples were inspected for possible duplication events and were classified according to a duplication mode (whole genome duplication [WGD], tandem, proximal, transposed, dispersed, non-duplicated). Similarly, the distribution of different modes of gene duplication was inspected across all genes expressed in nodules. The distribution of different modes of gene duplication differs among DEGs and all genes expressed in nodules ($\chi^2 = 24.108$, $df = 5$, p -value = 0.0002): The numbers of genes originated from tandem and dispersed duplication were significantly higher in DEGs while the genes originated from transposed duplication and non-duplicated genes were under-represented in DEGs (Table 3).

Table 3. Distribution of different modes of duplication across differentially expressed genes (DEGs) and genes expressed in nodules (EG) with *p*-values.

| Duplication Mode | DEG (%) | EG (%) | <i>p</i> -Value * |
|------------------|---------|--------|-------------------|
| WGD | 3.96 | 4.03 | 1 |
| Tandem | 14.58 | 9.91 | 0.0012 |
| Proximal | 5.21 | 5.34 | 1 |
| Transposed | 27.71 | 30.89 | 0.1358 |
| Dispersed | 37.08 | 32.75 | 0.0447 |
| Non-duplicated | 11.46 | 17.08 | 0.0008 |
| All | 100.00 | 100.00 | |

WGD—whole genome duplication, * two-tailed Fisher's exact test.

For estimating expression divergence, the Pearson correlation coefficients (r) among the expression profiles of the analyzed genes were calculated for those duplicated pairs within which at least one gene copy was identified as differentially expressed. A cutoff r -value below which two duplicated gene pairs were considered divergent in expression was determined. Inasmuch as 95% of the r -values for 10,000 randomly chosen gene pairs were <0.67, those gene pairs with $r < 0.67$ were considered to have diverged in expression at $\alpha = 0.05$. The duplicated gene pairs were divided according to the mode of duplication and the expression diversity was inspected separately for each duplication mode. Overall, 72–79% of duplicated pairs from dispersed, tandem, transposed, and WGD duplication diverged in expression, and a higher proportion (87%) of duplicated pairs with diverged expression was evaluated for those originated from proximal duplication (Figure 6).

3.8. Weighted Correlation Network Analysis

WGCNA identified 16 modules, each containing genes with similar expression profiles in all samples analyzed. The distribution of genes into modules and their clustering

according to similar expression profiles can be seen in Figure 7. The counts of genes in each module are in Table 4. The two most common modules are Turquoise and Blue, with 9097 and 7178 transcripts, respectively. These two modules are characterized by near mirror-image expression profiles between the two sample groups (samples A16_17, A16_21, S25_17, S46_12 and S46_7 vs. samples S55_6, T57_20 and T57_22). Eigengenes for each module were determined to reflect the common expression trend for genes belonging to that module (Supplementary Figure S6). Eigengenes are defined as the first principal component of each module and represent the module expression profile. A network (Supplementary Figure S7) was created to illustrate the relationships among gene expression profiles and modules. The network shows a pattern of two large groups of genes that are linked together by the Brown module genes. The first group is mainly composed of the Blue, Yellow, and Brown module genes. The second large group is made up of genes of the Green, Turquoise, Red, and Purple modules.

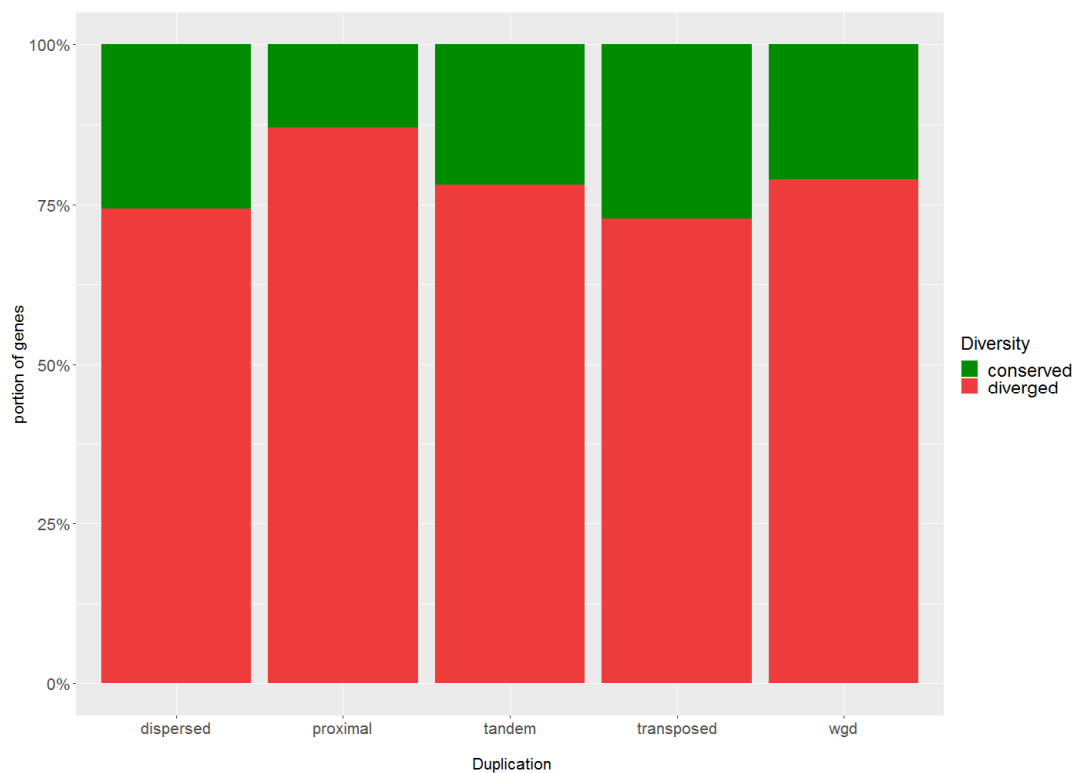


Figure 6. Expression divergence between duplicated gene pairs originated from different modes of duplication. Duplicated pairs were divided according to mode of duplication (x -axis) and proportions of genes with conserved versus diverged expression were calculated (y -axis).

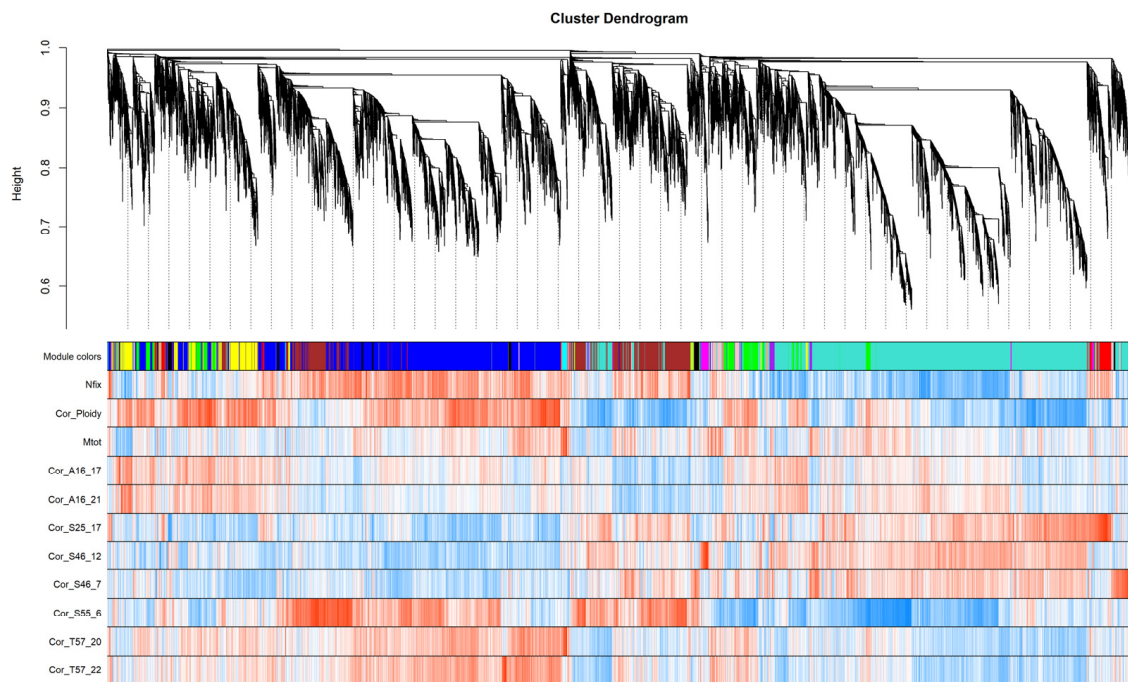


Figure 7. WGCNA analysis: Clustering of genes according to their expression profile similarity. First row assigned a module color to each gene. Rows 2 to 12 are red–blue scale heatmaps of Pearson correlation coefficients between traits and expression levels of the genes (red: 1, blue: -1).

Table 4. Frequency of genes in modules.

| Module Colors | Genes Frequency |
|---------------|-----------------|
| Turquoise | 9097 |
| Blue | 7178 |
| Brown | 3540 |
| Yellow | 1635 |
| Green | 1488 |
| Red | 712 |
| Black | 486 |
| Pink | 486 |
| Magenta | 329 |
| Purple | 210 |
| Greenyellow | 174 |
| Tan | 158 |
| Salmon | 153 |
| Cyan | 146 |
| Grey | 50 |
| Midnightblue | 31 |

The eigenvalues of the modules found were further correlated with the descriptive traits and the correlation coefficients were plotted in heatmaps (Supplementary Figure S8). The correlation criterion was determined by a p -value < 0.01 . The expression profile of a given module is represented by its module eigenvalues which we can correlate with a specific trait. We designed 16 modules, half of which correlate with genotype-specific modules (Midnightblue, Red, Magenta, Tan, Brown, Green, Cyan, Grey). The eigengenes of the Red, Blue, and Salmon modules were most positively correlated with the nitrogen fixation level. Conversely, the Greenyellow, Purple, and Turquoise modules were negatively correlated with nitrogen fixation. The modules Green, Yellow and Blue were negatively correlated, and the Brown module was the most strongly correlated positively with the ploidy level. The Cyan module is correlated the most positively with the sample weight trait

and the Grey module is correlated most negatively with that trait (Supplementary Figure S8). For groups of genes in all modules positively or negatively correlated with nitrogen fixation level, ploidy, and weight, and for groups of DEGs GO annotation, enrichment test and GOMCL summary were performed (Supplementary Table S6).

4. Discussion

The *Rhizobium*–legumes symbiosis has received much attention in recent decades because soil enrichment by nitrogen using BNF has environmental and ecological advantages over the use of synthetic nitrogen fertilizers. Realization of this phenotypic trait, however, is facilitated by the interaction of two genomes (plant \times bacteria) along with an influence of the environment. These conditions, taken together with the involvement of hundreds of genes connected with nodulation and nitrogen fixation, impede research into BNF and the identification of genes with a major influence on BNF efficiency and their utilization for agronomic purposes [14,35]. Therefore, the amount of fixed nitrogen acquired by current nitrogen-fixing plants is far below its potential. It has been estimated that the amounts of fixed nitrogen could be increased by as much as 300% through plant breeding and utilizing genotypes highly efficient in BNF [112]. Moreover, BNF efficiency is a highly variable trait differing not only between species [113] but also among individuals within a given species [13].

Due to the estimated high broad-sense heritability of this trait in relatively stable field conditions (more than 0.8 in *G. max* [114] and 0.9 in inbred lines of *T. incarnatum* [115]), the potential for selecting highly effective BNF genotypes is high, although it has been reported that efficiency of particular genotypes is greatly influenced by both environmental conditions (soil acidity, phosphorus availability) and symbiotic partner [116–118]. That was the reason why we followed up on the conclusions of Trněný et al. [13]; we evaluated the BNF efficiency in the next generation of strong- and weak-fixing red clover genotypes analyzed and evaluated in their publication. Because there is not a consistent opinion regarding the effect of ploidy upon BNF [119,120], diploid and tetraploid red clover genotypes of different red clover varieties were equally included in both contrasting groups (strong and weak fixators) to minimize the effect of ploidy upon BNF efficiency, and all analyzed plants were planted and maintained under the same conditions to reduce the environmental effect.

Among several methods developed for assessing BNF [79], ARA is one of the most widespread and is favored for its high sensitivity, and high throughput potential, especially for comparative purposes in manipulative experiments [121]. Because many factors influence the measured BNF rate, such as temperature [122], light [123], ecosystem successional stage [124] or seasonal/diurnal variations [125,126], ARA is less suitable for obtaining absolute values. As in our case, however, uniform measurement conditions at a specific time enable acquiring relative rates of BNF [127] and thus ARA was a method well suited to our purposes.

RNA-seq and the following differential gene expression analysis were focused upon the discovery of genes differentially expressed to a statistically significant extent within nodules between genotypes with high and low BNF efficiency regardless of ploidy and red clover variety and while controlling for effects of environmental conditions. Nodules served as the target tissue for evaluating nitrogen fixation. The expression profiles obtained reflected the involvement of plentiful genes for processes such as legume–rhizobia interaction and nodule development, and almost 500 DEGs were identified from RNA-Seq data. For the first time, our results report the assessment of genes influencing the efficiency of BNF in red clover.

Because a number of genes were annotated not at all or only in part, annotation of DEGs across genotypes was a necessary step to find their functions and allow their connection to biosynthetic pathways. Insomuch as red clover is not a model genetic plant, its first genome assembly was published only in 2014 [69], nine years later than the draft genome sequences of legumes *M. truncatula* and *L. japonicus* [128]. One year later, another

assembly was published [70] together with the construction of a physical map. Although we used both available annotations to decipher the functions of DEGs, this approach was not sufficient because about a quarter of the DEGs were without any annotation, thereby hindering the disclosure of their functions. Thus, we attempted to improve annotation using recently published annotation files of closely related species. This approach helped to improve annotation and allowed at least one functional annotation category to be assigned to each of more than 90% of DEGs. Even improved annotation, however, is not sufficient to identify the functions of many genes detected as DE, and limited assignment to some functional annotation category may merely suggest rather than reveal a possible function.

As a result of our analyses, DEGs encoded the highest number of enzymes as associated with sesquiterpenoid and triterpenoid synthesis. Terpenoids constitute a highly diverse and widely distributed group of secondary metabolites in plants playing various roles in plant defense, determination of membrane fluidity, or plant growth [129–131]. In the context of BNF and nodulation, it has been demonstrated that terpenoids are able to induce the expression of Nod factors or genes involved in the Nod signaling pathway [132]. Moreover, strigolactones, a group of terpenoid lactones acting as hormones, exhibit various roles in root growth and formation of root nodules in legumes [133,134], and strigolactone genes influence nodulation by inducing the expression of Nod factors of rhizobial bacteria [135]. Among other enriched pathways, several “sugar-related” signaling pathways were found: pentose and glucuronate interconversions (PGI), starch and sucrose metabolism (SSM), galactose metabolism (GM), and amino sugar and nucleotide sugar metabolism (ASNSM). Akbar et al. report the activation of PGI and SSM pathways during salt stress in cotton [136], and those authors hypothesized that the modification of these pathways could lead to significant tolerance to the salt stress. Similarly, shifting concentrations of metabolites within the PGI pathway were found during a study of stress response and host defense against plant herbivores [137]. GM and ASNSM pathways are well-studied in fungal pathogens or pathogen–plant interactions because the metabolites of these pathways are utilized on the wall surfaces as compounds of fungal and/or plant cell walls or virulence factors [138–140]. Among enriched pathways was also phenylpropanoid biosynthesis. Metabolites of this pathway then enter into multiple other pathways, such as lignin and flavonoid biosynthesis, and contribute to the response to both biotic and abiotic stimuli. They are indicators of various stress factors and mediators of particular stress tolerance [141]. They help to invade new habitats [142], or they influence the stability or robustness of plants in relation to mechanical or environmental factors such as drought using phenylpropanoid-based polymers [143]. Flavonoids, secondary metabolites of one of the branches of the phenylpropanoid pathway, are known to have multiple roles during the processes of nodulation and nitrogen fixation. They act as signal molecules during the early phases of the rhizobia and plant interaction [144] or serve as polar auxin transport inhibitors leading to nodule organogenesis [145].

Taken together, the pathways enriched by the representation of DEGs encoding particular enzymes are directly connected with nodulation and BNF (terpenoids, flavonoids), and the metabolites of the others can influence the BNF performance through several possible effects. For instance, metabolites of enriched “sugar-related” pathways are reported to have shifts in concentration under various stress conditions, thus indicating that these compounds could be involved in mechanisms for stress response. Although colonization of symbiotic rhizobia usually does not elicit plant defense mechanisms [146,147], the particular step during nodulation could be a cause of defense response under some circumstances because a plant controls every aspect of the correct nodulation process. In case of any problem or defect, the defense response can occur, and a plant can undergo some sort of stress condition. That means that the enrichment of pathways more or less connected with stress responses between genotypes with high and low fixing efficiency could result from the fact that the process of nodulation has not developed correctly, probably in weak-fixing genotypes, and resulting in plant stress response. Alternatively, some genotypes could have undergone some type of stress conditions (e.g., infection, mechanical damage) before

they were analyzed, although all plants were planted and maintained in the same way, and these stress stimuli could have an effect on nodulation and BNF efficiency. Table 5 summarizes the top 10 enriched GO bp terms, and stress response is one of the most enriched. That supports this hypothesis. Other enriched GO bp terms include several responses to stimulus, developmental processes, or interaction with different organisms, all of which are terms relating to biological processes expected in the context of legume symbiosis and nodulation.

Table 5. Top 10 enriched GO bp terms.

| GO ID | GO Name | p-Value |
|------------|--------------------------------------|------------------------|
| GO:0051704 | multi-organism process | 3.56×10^{-35} |
| GO:0006950 | response to stress | 1.41×10^{-29} |
| GO:0050896 | response to stimulus | 2.52×10^{-29} |
| GO:0009605 | response to external stimulus | 1.74×10^{-27} |
| GO:0001101 | response to acid chemical | 1.79×10^{-26} |
| GO:0048856 | anatomical structure development | 8.30×10^{-26} |
| GO:0042221 | response to chemical | 2.04×10^{-25} |
| GO:0032502 | developmental process | 8.03×10^{-25} |
| GO:0009607 | response to biotic stimulus | 2.16×10^{-24} |
| GO:0043207 | response to external biotic stimulus | 3.40×10^{-23} |

Leghaemoglobin genes were among the genes with the highest expression in the nodule transcriptome (Table 6). The same finding has been proven in *M. truncatula*, where genes for leghaemoglobin were also among the most strongly expressed genes in nodule transcriptome [66], and both species, too, have similar numbers of leghaemoglobin genes [69]. Leghaemoglobin proteins are necessary for the activity of the enzyme nitrogenase [148]. Because nitrogenase is irreversibly inactivated by oxygen [149], leghaemoglobins reduce free oxygen levels inside the bacteroids while allowing ATP production by transporting oxygen for respiratory processes on the bacteroid membrane [150].

Table 6. Top 10 genes with the highest expression in nodules. Mean expression is the number of reads assigned to a particular gene. This number was divided by the length of the sequence for normalization to different sequence lengths.

| Gene ID | Annotation | Mean Expression |
|---------------------------|----------------------------------|-----------------|
| Tp57577_TGAC_v2_gene25864 | Leghaemoglobin | 418.76 |
| Tp57577_TGAC_v2_gene23575 | Nodulin | 210.49 |
| Tp57577_TGAC_v2_gene4062 | Leghaemoglobin | 180.215 |
| Tp57577_TGAC_v2_gene25862 | Leghaemoglobin | 127.54 |
| Tp57577_TGAC_v2_gene25853 | Leghaemoglobin | 60.69 |
| Tp57577_TGAC_v2_gene23582 | IBR protein/transcription factor | 59.54 |
| Tp57577_TGAC_v2_gene39897 | Nodulin/Aquaporin | 55.20 |
| Tp57577_TGAC_v2_gene608 | Embryo-specific protein | 51.20 |
| Tp57577_TGAC_v2_gene29022 | Asparagine synthetase | 43.71 |
| Tp57577_TGAC_v2_gene25837 | Leghaemoglobin | 43.10 |

Inasmuch as red clover has nodules of an indeterminate type whose bacteroids are terminally differentiated, NCR peptides play an important role in nodule development, especially in bacteroid differentiation. Therefore, we strove to identify NCR peptides expressed in nodule transcripts and evaluate their predicted functions using in silico approaches. Ištvanek et al. [69] predicted 542 genes for NCR peptides during the first red clover assembly using tblastx searches against NCR peptides of *M. truncatula*, and that number is comparable with those identified in this model legume [151]. In contrast to this prediction, we were able to identify only 33 genes within the nodule transcriptome that

met the criteria set for the search for genes encoding NCR peptides (structure, conserved cysteines, length). Only 33 out of 37,000 genes detected in nodes had a conserved structure with 4 or 6 cysteines and length <150 bp, and for only 4 out of these 33 sequences were their functions supported by in silico analysis assessing, for example, signal sequence or subcellular location and BLAST searches. These differences in amounts of predicted and detected NCR peptides arose mostly due to our use of different methods. To predict NCR peptides, BLAST searches were performed regardless of structure, length, or other aspects that were considered in identifying NCR peptides in this study. Moreover, not all similar genes need to be really NCR in nature. They can be pseudogenes or can have different functions, such as producing defensins instead of functioning in root nodule symbiosis. As a result, many genes predicted as NCR during red clover assembly lack the typical NCR structure with conserved cysteines. The resulting number of sequences found was significantly lower compared to those in *M. truncatula*, but the abundance of NCR peptides among legume species has been reported to be highly variable [34]. The high number of NCR peptides can be due to: (1) constrained rhizobial growth in nodules, (2) selection against cheaters, (3) control of bacteroid development and metabolism, or (4) a combination of these points. Lower numbers of NCR peptides have been identified in several other legumes, such as 63 in chickpea (*C. arietinum*) [152] and 7 in *Glycyrrhiza uralensis* [32].

Gene duplication is considered to be one of the most important evolutionary mechanisms generating plentiful raw materials for processes such as speciation or neofunctionalization [153]. Gene duplication was realized by several mechanisms to varying degrees that include, among others, single gene duplication and whole genome duplication. Single gene duplication consists of four types: tandem (TD), proximal (PD), transposed (RD), and dispersed duplication (DSD) [106]. In the context of BNF, WGD has been extensively studied in connection with an ancient polyploidy event that occurred in a Papilionoideae lineage of legumes approximately 58 Ma ago [151]. Although it is generally supposed that this event did not precede BNF, it might have facilitated and refined the BNF system using genetic materials provided by this polyploidy event [154]. Here, we classified DEGs into five groups according to the duplication mode. We inspected the distribution of each particular mode among the DEGs, then compared this distribution with those across all genes detected in nodules. While we observed no statistically significant difference between the distribution of WGD, PD, and RD duplicates, TD and DSD duplicates were significantly overrepresented in DEGs and non-duplicated genes were significantly underrepresented in DEGs. The results showed the non-random distribution of a particular mode in DEGs and the preferential representation of duplicated genes connected with BNF efficiency. According to Qiao et al. [106], TD together with PD showed no significant decrease in frequency over time, thus indicating that this mode of duplication offers a continuous supply of genetic material for evolution and important genetic material for rapidly changing environments [155]. Dispersed duplicates are among the most prevalent duplication modes in genomes across different plant species [156]. Expression divergence analysis showed that about 75–80% of duplicated gene pairs diverged from each other in all those duplication modes analyzed, but the answer as to why only TDs and DSDs are overrepresented in DEGs remains unknown.

WGCNA analysis complements DEGs analysis and enables the arrangement of other transcripts. WGCNA analysis is used to classify genes according to their expression profiles. Genes with similar expression patterns may form clusters (modules) [157]. Transcripts in one module have a similar transcription pattern through all RNA-seq samples. In terms of nitrogen fixation, modules Blue, Red, and Salmon are negatively correlated and Greenyellow, Turquoise, and Purple are positively correlated. Of 491 DEGs, 51% belong to the Blue module and 15%, 14%, 8%, and 7%, respectively, to the Turquoise, Brown, Red and Yellow modules. The remaining DEGs are spread across other modules or were filtered prior to WGCNA analysis. Among putative NCR genes, 3 (gene18074, gene23764, and gene38999) of 8 such genes were part of the Blue module and 1 (gene33781) was part of the

Turquoise module. Another 4 identified NCR genes do not fulfill wgcna filter criteria for minimal expression level and expression variance across samples.

An interesting question is of where the known core genes of the root symbiotic nitrogen process appear. To answer this, we borrowed a list of 19 core predisposition genes that were collected in other studies within closely related species, in particular *M. truncatula* [158]. We found their red clover orthologues and then their localization in the WGCNA network and in the DEGs list (Supplementary Table S6). Fifteen core genes are captured in the most common Turquoise module, 2 core genes in the Brown module, and 1 each in the Blue, Green, Greenyellow, and Yellow modules.

Interestingly, no core gene was identified among the DEGs, indicating that the differential phenotype of nitrogen fixation levels is realized not at the level of symbiosis establishment and symbiotic structure formation but rather at the level of fixation regulation. This is supported by the fact that we are comparing not zero fixation level with non-zero but lower with higher fixation levels.

Nitrogen fixation through root nodule symbiosis is an essential process by which diazotrophic organisms make otherwise unavailable nitrogen available for their life needs and, through themselves, make it available to other living organisms. The phenomenon of symbiotic nitrogen fixation has evolved multiple times independently in one evolutionary branch of angiosperms that has been termed the “Nitrogen-fixing clade”. We can assume that, prior to the actual development of the ability to fix nitrogen, plants of this clade must have been predisposed through a support mechanism already in place [158,159]. It is probable that a broad and very complex transcriptomic background allowed nitrogen fixation to evolve while enabling the preservation of transcriptomic diversity in fixing nodules.

In red clover, an important non-model plant and forage crop, we found 491 differentially expressed genes connected with BNF efficiency. Subsequent annotation of genes in nodule transcriptome revealed more than 800 genes not yet experimentally confirmed. We were able to confirm only four nodule-specific cysteine-rich (NCR) peptides in the nodule transcriptome. In addition, we found unequal distribution of different modes of gene duplication in DEGs, with genes originating from tandem and dispersed duplication being significantly overrepresented in DEGs. Finally, using WGCNA we organized expression profiles of the transcripts into 16 modules linked to the analyzed traits, such as nitrogen fixation efficiency or sample-specific modules. Nodule transcriptomics is a rewarding topic. A series of transcriptomic studies have revealed transcripts associated with the root nodule symbiotic process [15,160–173]. The DEGs identified in this study and their analyses allowed a comparison to the nodule transcriptome in genotypes with different BNF efficiency and provided a valuable resource for further investigation of the genetic basis of this trait of interest.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life12121975/s1>, Supplementary Figure S1: Quality control of the mapped reads. Supplementary Figure S2: Example of Pearson correlation coefficient of biological replicates. Supplementary Figure S3: Dendrogram of rlog-transformed read counts. Supplementary Figure S4: PCA plot of rlog-transformed read counts. Supplementary Figure S5: MA plot of genes detected in nodules. Supplementary Figure S6: Eigengenes expression profiles of 16 WGCNA modules. Supplementary Figure S7: WGCNA network. Supplementary Figure S8: Heatmap of relationships between traits and module eigengenes. Supplementary Table S1: Primer pairs used for qPCR verification of sequencing data. Supplementary Table S2: List of differentially expressed genes between genotypes with high and low BNF efficiency. Supplementary Table S3: Expression distribution of originally unconfirmed genes detected in nodules. Supplementary Table S4: Top 10 GO bp terms in originally unconfirmed genes detected in modules. Supplementary Table S5: List of originally unconfirmed genes. Supplementary Table S6: Annotation summary of GO terms for trait-correlated WGCNA modules and DEGs.

Author Contributions: D.V. and J.Ř. conceptualization; D.V. performed most of the experiments and processed sequencing data; O.T. performed WGCNA analysis; D.V. wrote the manuscript; J.Ř. review

and editing. All authors approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education, Youth and Sports of the Czech Republic (project no. MUNI/A/1325/2021). O.T. is supported by institutional funding on long-term conceptual development from the Agricultural Research Ltd. organization founded by the Ministry of Agriculture of the Czech Republic (MZE-RO1722).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and Supplementary Materials.

Acknowledgments: The authors are thankful for and greatly appreciate access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum that was provided under the program “Projects of Large Research, Development, and Innovations Infrastructures” (CESNET LM2015042). In particular, we greatly appreciate access to the CERIT-SC computing and storage facilities provided by the CERIT-SC Center, provided under the program “Projects of Large Research, Development, and Innovations Infrastructures” (CERIT Scientific Cloud LM2015085). Seeds were procured from the GeneBank of Crop Research Institute Ltd., Prague–Ruzyně, Czech Republic.

Conflicts of Interest: O.T. is employed by the company Agricultural Research, Ltd., Troubsko, Czech Republic and he declares no conflict of interests. The remaining authors declare no competing interest.

Abbreviations

| | |
|-------------|---|
| ARA | Acetylene reduction assay |
| ASNSM | Amino sugar and nucleotide sugar metabolism |
| BNF | Biological nitrogen fixation |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| Ct | Cycle threshold |
| DEG | Differentially expressed gene |
| DSD | Dispersed duplication |
| FDR | False discovery rate |
| GM | Galactose metabolism |
| GO | Gene Ontology |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| KO | Kegg orthology |
| KOG | Clusters of Orthologous Groups |
| NCR | Nodule-specific cysteine-rich |
| <i>padj</i> | adjusted <i>p</i> -value |
| PCA | Principal component analysis |
| PD | Proximal duplication |
| PGI | Pentose and glucuronate interconversions |
| RD | Transposed duplication |
| qPCR | quantitative polymerase chain reaction |
| RNAi | RNA interference |
| SSM | Starch and sucrose metabolism |
| TD | Tandem duplication |
| Tp | Trifolium pratense |
| WGCNA | Weighted correlation network analysis |
| WGD | whole genome duplication |

References

1. Lewis, G.P.; Schrire, B.D.; Mackinder, B.A.; Rico, L.; Clark, R. A 2013 Linear Sequence of Legume Genera Set in a Phylogenetic Context—A Tool for Collections Management and Taxon Sampling. *South Afr. J. Bot.* **2013**, *89*, 76–84. [[CrossRef](#)]
2. Azani, N.; Babineau, M.; Bailey, C.D.; Banks, H.; Barbosa, A.R.; Pinto, R.B.; Boatwright, J.S.; Borges, L.M.; Brown, G.K.; Bruneau, A.; et al. A New Subfamily Classification of the Leguminosae Based on a Taxonomically Comprehensive Phylogeny—The Legume Phylogeny Working Group (LPWG). *Taxon* **2017**, *66*, 44–77. [[CrossRef](#)]

3. Bell, C.J. The Medicago Genome Initiative: A Model Legume Database. *Nucleic Acids Res.* **2001**, *29*, 114–117. [[CrossRef](#)] [[PubMed](#)]
4. Udvardi, M.K.; Tabata, S.; Parniske, M.; Stougaard, J. *Lotus Japonicus*: Legume Research in the Fast Lane. *Trends Plant Sci.* **2005**, *10*, 222–228. [[CrossRef](#)]
5. Rose, R.J. *Medicago Truncatula* as a Model for Understanding Plant Interactions with Other Organisms, Plant Development and Stress Biology: Past, Present and Future. *Funct. Plant Biol.* **2008**, *35*, 253. [[CrossRef](#)]
6. Crews, T.E. The Presence of Nitrogen Fixing Legumes in Terrestrial Communities: Evolutionary vs Ecological Considerations. In *New Perspectives on Nitrogen Cycling in the Temperate and Tropical Americas*; Springer: Dordrecht, The Netherlands, 1999; pp. 233–246.
7. Zohary, M.; Heller, D. *The Genus Trifolium.*; Israel Academy of Sciences and Humanities: Jerusalem, Israel, 1984; ISBN 965-208-056-X.
8. Yilmaz, A.; Yeltekin, Y. The Evaluations Of Taxonomic Classifications In The Genus *Trifolium* L. Based On ITS Sequences. *Sak. Univ. J. Sci.* **2022**, *26*, 545–553. [[CrossRef](#)]
9. Kintl, A.; Elbl, J.; Lošák, T.; Vavrková, M.; Nedělník, J. Mixed Intercropping of Wheat and White Clover to Enhance the Sustainability of the Conventional Cropping System: Effects on Biomass Production and Leaching of Mineral Nitrogen. *Sustainability* **2018**, *10*, 3367. [[CrossRef](#)]
10. Hidalgo, L.A.; Chedraui, P.A.; Morocho, N.; Ross, S.; San Miguel, G. The Effect of Red Clover Isoflavones on Menopausal Symptoms, Lipids and Vaginal Cytology in Menopausal Women: A Randomized, Double-Blind, Placebo-Controlled Study. *Gynecol. Endocrinol.* **2005**, *21*, 257–264. [[CrossRef](#)]
11. Occhiuto, F.; Pasquale, R.D.; Guglielmo, G.; Palumbo, D.R.; Zangla, G.; Samperi, S.; Renzo, A.; Circosta, C. Effects of Phytoestrogenic Isoflavones from Red Clover (*Trifolium pratense* L.) on Experimental Osteoporosis. *Phytother. Res.* **2007**, *21*, 130–134. [[CrossRef](#)]
12. Akbaribazm, M.; Khazaei, F.; Naseri, L.; Pazhouhi, M.; Zamanian, M.; Khazaei, M. Pharmacological and Therapeutic Properties of the Red Clover (*Trifolium pratense* L.): An Overview of the New Findings. *J. Tradit. Chin. Med.* **2021**, *41*, 8.
13. Trněný, O.; Vlk, D.; Macková, E.; Matoušková, M.; Řepková, J.; Nedělník, J.; Hofbauer, J.; Vejražka, K.; Jakešová, H.; Jansa, J.; et al. Allelic Variants for Candidate Nitrogen Fixation Genes Revealed by Sequencing in Red Clover (*Trifolium pratense* L.). *Int. J. Mol. Sci.* **2019**, *20*, 5470. [[CrossRef](#)] [[PubMed](#)]
14. Kouchi, H.; Imaizumi-Anraku, H.; Hayashi, M.; Hakoyama, T.; Nakagawa, T.; Umehara, Y.; Suganuma, N.; Kawaguchi, M. How Many Peas in a Pod? Legume Genes Responsible for Mutualistic Symbioses Underground. *Plant Cell Physiol.* **2010**, *51*, 1381–1397. [[CrossRef](#)] [[PubMed](#)]
15. Roux, B.; Rodde, N.; Jardinaud, M.-F.; Timmers, T.; Sauviac, L.; Cottret, L.; Carrère, S.; Sallet, E.; Courcelle, E.; Moreau, S.; et al. An Integrated Analysis of Plant and Bacterial Gene Expression in Symbiotic Root Nodules Using Laser-Capture Microdissection Coupled to RNA Sequencing. *Plant J.* **2014**, *77*, 817–837. [[CrossRef](#)] [[PubMed](#)]
16. Gage, D.J. Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 280–300. [[CrossRef](#)]
17. Masson-Boivin, C.; Giraud, E.; Perret, X.; Batut, J. Establishing Nitrogen-Fixing Symbiosis with Legumes: How Many *Rhizobium* Recipes? *Trends Microbiol.* **2009**, *17*, 458–466. [[CrossRef](#)]
18. Zhang, X.-X.; Turner, S.L.; Guo, X.-W.; Yang, H.-J.; Debelle, F.; Yang, G.-P.; Dénarié, J.; Young, J.P.W.; Li, F.-D. The Common Nodulation Genes of *Astragalus Sinicus* Rhizobia Are Conserved despite Chromosomal Diversity. *Appl. Environ. Microbiol.* **2000**, *66*, 2988–2995. [[CrossRef](#)]
19. Bek, A.S.; Sauer, J.; Thygesen, M.B.; Duus, J.Ø.; Petersen, B.O.; Thirup, S.; James, E.; Jensen, K.J.; Stougaard, J.; Radutoiu, S. Improved Characterization of Nod Factors and Genetically Based Variation in LysM Receptor Domains Identify Amino Acids Expendable for Nod Factor Recognition in *Lotus* Spp. *Mol. Plant. Microbe Interact.* **2010**, *23*, 58–66. [[CrossRef](#)]
20. Riely, B.K.; Ané, J.-M.; Penmetsa, R.V.; Cook, D.R. Genetic and Genomic Analysis in Model Legumes Bring Nod-Factor Signaling to Center Stage. *Curr. Opin. Plant Biol.* **2004**, *7*, 408–413. [[CrossRef](#)]
21. Kassaw, T.; Nowak, S.; Schnabel, E.; Frugoli, J. ROOT DETERMINED NODULATION1 Is Required for *M. Truncatula* CLE12, But Not CLE13, Peptide Signaling through the SUNN Receptor Kinase. *Plant Physiol.* **2017**, *174*, 2445–2456. [[CrossRef](#)]
22. Gleason, C.; Chaudhuri, S.; Yang, T.; Munoz, A.; Poovaiah, B.; Oldroyd, G.E. Nodulation Independent of Rhizobia Induced by a Calcium-Activated Kinase Lacking Autoinhibition. *Nature* **2006**, *441*, 1149–1152. [[CrossRef](#)]
23. Tirichine, L.; Sandal, N.; Madsen, L.H.; Radutoiu, S.; Albrektsen, A.S.; Sato, S.; Asamizu, E.; Tabata, S.; Stougaard, J. A Gain-of-Function Mutation in a Cytokinin Receptor Triggers Spontaneous Root Nodule Organogenesis. *Science* **2007**, *315*, 104–107. [[CrossRef](#)] [[PubMed](#)]
24. Schultze, M.; Kondorosi, A. Regulation of Symbiotic Root Nodule Development. *Annu. Rev. Genet.* **1998**, *32*, 33. [[CrossRef](#)] [[PubMed](#)]
25. Downie, J.A. Legume Nodulation. *Curr. Biol.* **2014**, *24*, R184–R190. [[CrossRef](#)] [[PubMed](#)]
26. Brill, W.J. Biochemical Genetics of Nitrogen Fixation. *Microbiol. Rev.* **1980**, *44*, 449–467. [[CrossRef](#)]
27. Ott, T.; van Dongen, J.T.; Günther, C.; Krusell, L.; Desbrosses, G.; Vigeolas, H.; Bock, V.; Czechowski, T.; Geigenberger, P.; Udvardi, M.K. Symbiotic Leghemoglobins Are Crucial for Nitrogen Fixation in Legume Root Nodules but Not for General Plant Growth and Development. *Curr. Biol.* **2005**, *15*, 531–535. [[CrossRef](#)]
28. Nap, J.-P.; Bisseling, T. Nodulin Function and Nodulin Gene Regulation in Root Nodule Development. In *Molecular Biology of Symbiotic Nitrogen Fixation*; CRC Press: Boca Raton, FL, USA, 2018; pp. 181–229. ISBN 1-351-07474-1.

29. Mergaert, P.; Uchiumi, T.; Alunni, B.; Evanno, G.; Cheron, A.; Catrice, O.; Mausset, A.-E.; Barloy-Hubler, F.; Galibert, F.; Kondorosi, A.; et al. Eukaryotic Control on Bacterial Cell Cycle and Differentiation in the *Rhizobium*–Legume Symbiosis. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5230–5235. [[CrossRef](#)]
30. Haag, A.F.; Arnold, M.F.F.; Myka, K.K.; Kerscher, B.; Dall’Angelo, S.; Zanda, M.; Mergaert, P.; Ferguson, G.P. Molecular Insights into Bacteroid Development during *Rhizobium*–Legume Symbiosis. *FEMS Microbiol. Rev.* **2013**, *37*, 364–383. [[CrossRef](#)]
31. Alunni, B.; Gourion, B. Terminal Bacteroid Differentiation in the Legume–*Rhizobium* Symbiosis: Nodule-Specific Cysteine-Rich Peptides and Beyond. *New Phytol.* **2016**, *211*, 411–417. [[CrossRef](#)]
32. Montiel, J.; Downie, J.A.; Farkas, A.; Bihari, P.; Herczeg, R.; Bálint, B.; Mergaert, P.; Kereszt, A.; Kondorosi, É. Morphotype of Bacteroids in Different Legumes Correlates with the Number and Type of Symbiotic NCR Peptides. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5041–5046. [[CrossRef](#)]
33. Zhou, P.; Silverstein, K.A.; Gao, L.; Walton, J.D.; Nallu, S.; Guhlin, J.; Young, N.D. Detecting Small Plant Peptides Using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinform.* **2013**, *14*, 335. [[CrossRef](#)]
34. Downie, J.A.; Kondorosi, E. Why Should Nodule Cysteine-Rich (NCR) Peptides Be Absent From Nodules of Some Groups of Legumes but Essential for Symbiotic N-Fixation in Others? *Front. Agron.* **2021**, *3*, 654576. [[CrossRef](#)]
35. Roy, S.; Liu, W.; Nandety, R.S.; Crook, A.; Mysore, K.S.; Pislariu, C.I.; Frugoli, J.; Dickstein, R.; Udvardi, M.K. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* **2020**, *32*, 15–41. [[CrossRef](#)] [[PubMed](#)]
36. Sagan, M.; Morandi, D.; Tarengi, E.; Duc, G. Selection of Nodulation and Mycorrhizal Mutants in the Model Plant *Medicago Truncatula* (Gaertn.) after γ -Ray Mutagenesis. *Plant Sci.* **1995**, *111*, 63–71. [[CrossRef](#)]
37. Penmetsa, R.V.; Cook, D.R. Production and Characterization of Diverse Developmental Mutants of *Medicago Truncatula*. *Plant Physiol.* **2000**, *123*, 1387–1398. [[CrossRef](#)]
38. Szczygłowski, K.; Shaw, R.S.; Wopereis, J.; Copeland, S.; Hamburger, D.; Kasiborski, B.; Dazzo, F.B.; de Bruijn, F.J. Nodule Organogenesis and Symbiotic Mutants of the Model Legume *Lotus Japonicus*. *Mol. Plant-Microbe Interact.* **1998**, *11*, 684–697. [[CrossRef](#)]
39. Bolon, Y.-T.; Haun, W.J.; Xu, W.W.; Grant, D.; Stacey, M.G.; Nelson, R.T.; Gerhardt, D.J.; Jeddelloh, J.A.; Stacey, G.; Muehlbauer, G.J.; et al. Phenotypic and Genomic Analyses of a Fast Neutron Mutant Population Resource in Soybean. *Plant Physiol.* **2011**, *156*, 240–253. [[CrossRef](#)]
40. Thykjaer, T.; Stiller, J.; Handberg, K.; Jones, J.; Stougaard, J. The Maize Transposable Element Ac Is Mobile in the Legume *Lotus Japonicus*. *Plant Mol. Biol.* **1995**, *27*, 981–993. [[CrossRef](#)]
41. Schauser, L.; Handberg, K.; Sandal, N.; Stiller, J.; Thykjaer, T.; Pajuelo, E.; Nielsen, A.; Stougaard, J. Symbiotic Mutants Deficient in Nodule Establishment Identified after T-DNA Transformation of *Lotus Japonicus*. *Mol. Gen. Genet. MGG* **1998**, *259*, 414–423. [[CrossRef](#)]
42. Scholte, M.; d’Erfurth, I.; Rippa, S.; Mondy, S.; Cosson, V.; Durand, P.; Breda, C.; Trinh, H.; Rodriguez-Llorente, I.; Kondorosi, E. T-DNA Tagging in the Model Legume *Medicago Truncatula* Allows Efficient Gene Discovery. *Mol. Breed.* **2002**, *10*, 203–215. [[CrossRef](#)]
43. Tadege, M.; Wen, J.; He, J.; Tu, H.; Kwak, Y.; Eschstruth, A.; Cayrel, A.; Endre, G.; Zhao, P.X.; Chabaud, M.; et al. Large-Scale Insertional Mutagenesis Using the Tnt1 Retrotransposon in the Model Legume *Medicago truncatula*. *Plant J.* **2008**, *54*, 335–347. [[CrossRef](#)]
44. Pislariu, C.I.; Murray, J.D.; Wen, J.; Cosson, V.; Muni, R.R.D.; Wang, M.; Benedito, V.A.; Andriankaja, A.; Cheng, X.; Jerez, I.T.; et al. A *Medicago truncatula* Tobacco Retrotransposon Insertion Mutant Collection with Defects in Nodule Development and Symbiotic Nitrogen Fixation. *Plant Physiol.* **2012**, *159*, 1686–1699. [[CrossRef](#)] [[PubMed](#)]
45. Fukai, E.; Stougaard, J.; Hayashi, M. Activation of an Endogenous Retrotransposon Associated with Epigenetic Changes in *Lotus Japonicus*: A Tool for Functional Genomics in Legumes. *Plant Genome* **2013**, *6*, 1–11. [[CrossRef](#)]
46. Małolepszy, A.; Mun, T.; Sandal, N.; Gupta, V.; Dubin, M.; Urbański, D.; Shah, N.; Bachmann, A.; Fukai, E.; Hirakawa, H.; et al. The LORE 1 Insertion Mutant Resource. *Plant J.* **2016**, *88*, 306–317. [[CrossRef](#)] [[PubMed](#)]
47. Cordoba, E.; Shishkova, S.; Vance, C.P.; Hernández, G. Antisense Inhibition of NADH Glutamate Synthase Impairs Carbon/Nitrogen Assimilation in Nodules of Alfalfa (*Medicago sativa* L.): Antisense NADH-GOGAT in Alfalfa Nodules. *Plant J.* **2003**, *33*, 1037–1049. [[CrossRef](#)]
48. Kumagai, H.; Kouchi, H. Gene Silencing by Expression of Hairpin RNA in *Lotus japonicus* Roots and Root Nodules. *Mol. Plant-Microbe Interactions* **2003**, *16*, 663–668. [[CrossRef](#)]
49. Arthikala, M.-K.; Nanjareddy, K.; Lara, M. RNA Interference: A Powerful Functional Analysis Tool for Studying Legume Symbioses. In *RNA Interference*; Avid Science: Berlin, Germany, 2017; pp. 1–45.
50. Jacobs, T.B.; LaFayette, P.R.; Schmitz, R.J.; Parrott, W.A. Targeted Genome Modifications in Soybean with CRISPR/Cas9. *BMC Biotechnol.* **2015**, *15*, 16. [[CrossRef](#)]
51. Wang, L.; Wang, L.; Tan, Q.; Fan, Q.; Zhu, H.; Hong, Z.; Zhang, Z.; Duanmu, D. Efficient Inactivation of Symbiotic Nitrogen Fixation Related Genes in *Lotus japonicus* Using CRISPR-Cas9. *Front. Plant Sci.* **2016**, *7*, 1333. [[CrossRef](#)]
52. Curtin, S.J.; Tiffin, P.; Guhlin, J.; Trujillo, D.I.; Burghardt, L.T.; Atkins, P.; Baltes, N.J.; Denny, R.; Voytas, D.F.; Stupar, R.M.; et al. Validating Genome-Wide Association Candidates Controlling Quantitative Variation in Nodulation. *Plant Physiol.* **2017**, *173*, 921–931. [[CrossRef](#)]

53. Badhan, S.; Ball, A.S.; Mantri, N. First Report of CRISPR/Cas9 Mediated DNA-Free Editing of 4CL and RVE7 Genes in Chickpea Protoplasts. *Int. J. Mol. Sci.* **2021**, *22*, 396. [[CrossRef](#)]
54. Berger, A.; Guinand, S.; Boscardi, A.; Puppo, A.; Brouquisse, R. *Medicago truncatula* Phytohemoglobin 1.1 Controls Symbiotic Nodulation and Nitrogen Fixation via the Regulation of Nitric Oxide Concentration. *New Phytol.* **2020**, *227*, 84–98. [[CrossRef](#)]
55. Burén, S.; López-Torrejón, G.; Rubio, L.M. Extreme Bioengineering to Meet the Nitrogen Challenge. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 8849–8851. [[CrossRef](#)]
56. Yang, J.; Xie, X.; Xiang, N.; Tian, Z.-X.; Dixon, R.; Wang, Y.-P. Polypeptide Strategy for Stoichiometric Assembly of Nitrogen Fixation Components for Synthetic Biology. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E8509–E8517. [[CrossRef](#)] [[PubMed](#)]
57. Battle, A.; Mostafavi, S.; Zhu, X.; Potash, J.B.; Weissman, M.M.; McCormick, C.; Haudenschild, C.D.; Beckman, K.B.; Shi, J.; Mei, R.; et al. Characterizing the Genetic Basis of Transcriptome Diversity through RNA-Sequencing of 922 Individuals. *Genome Res.* **2014**, *24*, 14–24. [[CrossRef](#)] [[PubMed](#)]
58. Sun, Y.; Xiao, H. Identification of Alternative Splicing Events by RNA Sequencing in Early Growth Tomato Fruits. *BMC Genom.* **2015**, *16*, 948. [[CrossRef](#)] [[PubMed](#)]
59. Deng, T.; Pang, C.; Lu, X.; Zhu, P.; Duan, A.; Tan, Z.; Huang, J.; Li, H.; Chen, M.; Liang, X. De Novo Transcriptome Assembly of the Chinese Swamp Buffalo by RNA Sequencing and SSR Marker Discovery. *PLoS ONE* **2016**, *11*, e0147132. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, T.; Huang, L.; Wang, Y.; Wang, W.; Zhao, X.; Zhang, S.; Zhang, J.; Hu, F.; Fu, B.; Li, Z. Differential Transcriptome Profiling of Chilling Stress Response between Shoots and Rhizomes of *Oryza Longistaminata* Using RNA Sequencing. *PLoS ONE* **2017**, *12*, e0188625. [[CrossRef](#)]
61. Jin, H.; Dong, D.; Yang, Q.; Zhu, D. Salt-Responsive Transcriptome Profiling of *Suaeda Glauca* via RNA Sequencing. *PLoS ONE* **2016**, *11*, e0150504. [[CrossRef](#)]
62. Lohani, N.; Singh, M.B.; Bhalla, P.L. RNA-Seq Highlights Molecular Events Associated With Impaired Pollen-Pistil Interactions Following Short-Term Heat Stress in *Brassica napus*. *Front. Plant Sci.* **2021**, *11*, 622748. [[CrossRef](#)]
63. Zhang, C.; Li, X.; Wang, Z.; Zhang, Z.; Wu, Z. Identifying Key Regulatory Genes of Maize Root Growth and Development by RNA Sequencing. *Genomics* **2020**, *112*, 5157–5169. [[CrossRef](#)]
64. Wang, X.; Liu, C.; Tu, B.; Li, Y.; Chen, H.; Zhang, Q.; Liu, X. Characterization on a Novel Rolled Leaves and Short Petioles Soybean Mutant Based on Seq-BSA and RNA-Seq Analysis. *J. Plant Biol.* **2022**, *65*, 261–277. [[CrossRef](#)]
65. Bellés-Sancho, P.; Lardi, M.; Liu, Y.; Eberl, L.; Zamboni, N.; Bailly, A.; Pessi, G. Metabolomics and Dual RNA-Sequencing on Root Nodules Revealed New Cellular Functions Controlled by *Paraburkholderia phymatum* NifA. *Metabolites* **2021**, *11*, 455. [[CrossRef](#)] [[PubMed](#)]
66. Cabeza, R.A.; Liese, R.; Lingner, A.; von Stieglitz, I.; Neumann, J.; Salinas-Riester, G.; Pommerenke, C.; Dittert, K.; Schulze, J. RNA-Seq Transcriptome Profiling Reveals That *Medicago truncatula* Nodules Acclimate N₂ Fixation before Emerging P Deficiency Reaches the Nodules. *J. Exp. Bot.* **2014**, *65*, 6035–6048. [[CrossRef](#)] [[PubMed](#)]
67. Kusakin, P.G.; Serova, T.A.; Gogoleva, N.E.; Gogolev, Y.V.; Tsyganov, V.E. Laser Microdissection of *Pisum Sativum* L. Nodules Followed by RNA-Seq Analysis Revealed Crucial Transcriptomic Changes during Infected Cell Differentiation. *Agronomy* **2021**, *11*, 2504. [[CrossRef](#)]
68. Yuan, S.L.; Li, R.; Chen, H.F.; Zhang, C.J.; Chen, L.M.; Hao, Q.N.; Chen, S.L.; Shan, Z.H.; Yang, Z.L.; Zhang, X.J.; et al. RNA-Seq Analysis of Nodule Development at Five Different Developmental Stages of Soybean (*Glycine Max*) Inoculated with *Bradyrhizobium japonicum* Strain 113-2. *Sci. Rep.* **2017**, *7*, 42248. [[CrossRef](#)]
69. Ištváněk, J.; Jaroš, M.; Křenek, A.; Řepková, J. Genome Assembly and Annotation for Red Clover (*Trifolium pratense*; Fabaceae). *Am. J. Bot.* **2014**, *101*, 327–337. [[CrossRef](#)]
70. De Vega, J.J.; Ayling, S.; Hegarty, M.; Kudrna, D.; Goicoechea, J.L.; Ergon, A.; Rognli, O.A.; Jones, C.; Swain, M.; Geurts, R.; et al. Red Clover (*Trifolium pratense* L.) Draft Genome Provides a Platform for Trait Improvement. *Sci. Rep.* **2015**, *5*, 17394. [[CrossRef](#)]
71. Yates, S.A.; Swain, M.T.; Hegarty, M.J.; Chernukin, I.; Lowe, M.; Allison, G.G.; Ruttink, T.; Abberton, M.T.; Jenkins, G.; Skot, L. De Novo Assembly of Red Clover Transcriptome Based on RNA-Seq Data Provides Insight into Drought Response, Gene Discovery and Marker Identification. *BMC Genom.* **2014**, *15*, 453. [[CrossRef](#)]
72. Herbert, D.B.; Gross, T.; Rupp, O.; Becker, A. Transcriptome Analysis Reveals Major Transcriptional Changes during Regrowth after Mowing of Red Clover (*Trifolium pratense*). *BMC Plant Biol.* **2021**, *21*, 95. [[CrossRef](#)]
73. Chao, Y.; Xie, L.; Yuan, J.; Guo, T.; Li, Y.; Liu, F.; Han, L. Transcriptome Analysis of Leaf Senescence in Red Clover (*Trifolium pratense* L.). *Physiol. Mol. Biol. Plants* **2018**, *24*, 753–765. [[CrossRef](#)]
74. Chao, Y.; Yuan, J.; Li, S.; Jia, S.; Han, L.; Xu, L. Analysis of Transcripts and Splice Isoforms in Red Clover (*Trifolium pratense* L.) by Single-Molecule Long-Read Sequencing. *BMC Plant Biol.* **2018**, *18*, 300. [[CrossRef](#)]
75. Shi, K.; Liu, X.; Pan, X.; Liu, J.; Gong, W.; Gong, P.; Cao, M.; Jia, S.; Wang, Z. Unveiling the Complexity of Red Clover (*Trifolium pratense* L.) Transcriptome and Transcriptional Regulation of Isoflavonoid Biosynthesis Using Integrated Long- and Short-Read RNAseq. *Int. J. Mol. Sci.* **2021**, *22*, 12625. [[CrossRef](#)] [[PubMed](#)]
76. Zhang, J.; Li, J.; Zou, L.; Li, H. Transcriptome Analysis of Air Space-Type Variegation Formation in *Trifolium Pratense*. *Int. J. Mol. Sci.* **2022**, *23*, 7794. [[CrossRef](#)] [[PubMed](#)]
77. Provorov, N.; Tikhonovich, I. Genetic Resources for Improving Nitrogen Fixation in Legume-Rhizobia Symbiosis. *Genet. Resour. Crop Evol.* **2003**, *50*, 89–99. [[CrossRef](#)]

78. Hardy, R.W.F.; Burns, R.C.; Holsten, R.D. Applications of the Acetylene-Ethylene Assay for Measurement of Nitrogen Fixation. *Soil Biol. Biochem.* **1973**, *5*, 47–81. [[CrossRef](#)]
79. Unkovich, M.; Herridge, D.; Peoples, M.; Cadisch, G.; Boddey, B.; Giller, K.; Alves, B.; Chalk, P. *Measuring Plant-Associated Nitrogen Fixation in Agricultural Systems*; Australian Centre for International Agricultural Research (ACIAR): Canberra, ACT, Australia, 2008; ISBN 1-921531-26-6.
80. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 29 October 2020).
81. Bushnell, B. BBLMap: A Fast, Accurate, Splice-Aware Aligner. LBNL Report LBNL-7065E, Lawrence Berkeley National Laboratory. 2014. Available online: <https://escholarship.org/uc/item/1h3515gn> (accessed on 2 November 2020).
82. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
83. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
84. Hartley, S.W.; Mullikin, J.C. QoRTs: A Comprehensive Toolset for Quality Control and Data Processing of RNA-Seq Experiments. *BMC Bioinform.* **2015**, *16*, 224. [[CrossRef](#)]
85. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)]
86. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)]
87. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
88. Team, R.S. RStudio: Integrated Development for R. RStudio, PBC: Boston, MA, USA, 2021; Available online: <http://www.rstudio.com> (accessed on 10 November 2020).
89. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [[CrossRef](#)] [[PubMed](#)]
90. Li, J.; Dai, X.; Liu, T.; Zhao, P.X. LegumeIP: An Integrative Database for Comparative Genomics and Transcriptomics of Model Legumes. *Nucleic Acids Res.* **2012**, *40*, D1221–D1229. [[CrossRef](#)] [[PubMed](#)]
91. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)] [[PubMed](#)]
92. Boeckmann, B. The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [[CrossRef](#)]
93. Kaur, P.; Bayer, P.E.; Milec, Z.; Vrána, J.; Yuan, Y.; Appels, R.; Edwards, D.; Batley, J.; Nichols, P.; Erskine, W.; et al. An Advanced Reference Genome of *Trifolium Subterraneum* L. Reveals Genes Related to Agronomic Performance. *Plant Biotechnol. J.* **2017**, *15*, 1034–1046. [[CrossRef](#)]
94. Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-Throughput Functional Annotation and Data Mining with the Blast2GO Suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [[CrossRef](#)]
95. Bioinformatics, B.; Valencia, S. OmicsBox-Bioinformatics Made Easy. *March* **2019**, *3*, 2019.
96. Koressaar, T.; Remm, M. Enhancements and Modifications of Primer Design Program Primer3. *Bioinformatics* **2007**, *23*, 1289–1291. [[CrossRef](#)]
97. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
98. Maróti, G.; Downie, J.A.; Kondorosi, É. Plant Cysteine-Rich Peptides That Inhibit Pathogen Growth and Control Rhizobial Differentiation in Legume Nodules. *Curr. Opin. Plant Biol.* **2015**, *26*, 57–63. [[CrossRef](#)]
99. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2007**, *35*, D61–D65. [[CrossRef](#)] [[PubMed](#)]
100. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [[CrossRef](#)] [[PubMed](#)]
101. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)] [[PubMed](#)]
102. Almagro Armenteros, J.J.; Sønderby, C.K.; Sønderby, S.K.; Nielsen, H.; Winther, O. DeepLoc: Prediction of Protein Subcellular Localization Using Deep Learning. *Bioinformatics* **2017**, *33*, 3387–3395. [[CrossRef](#)] [[PubMed](#)]
103. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 571–607.
104. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]

105. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496–1503. [[CrossRef](#)] [[PubMed](#)]
106. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene Duplication and Evolution in Recurring Polyploidization–Diploidization Cycles in Plants. *Genome Biol.* **2019**, *20*, 38. [[CrossRef](#)]
107. Langfelder, P.; Horvath, S. WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
108. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
109. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449. [[CrossRef](#)]
110. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [[CrossRef](#)]
111. Wang, G.; Oh, D.-H.; Dassanayake, M. GOMCL: A Toolkit to Cluster, Evaluate, and Extract Non-Redundant Associations of Gene Ontology-Based Functions. *BMC Bioinform.* **2020**, *21*, 1–9. [[CrossRef](#)]
112. Vance, C.P. Legume Symbiotic Nitrogen Fixation: Agronomic Aspects. In *The Rhizobiaceae*; Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 1998; pp. 509–530. ISBN 978-0-7923-5180-1.
113. Sulieman, S.; Schulze, J. The Efficiency of Nitrogen Fixation of the Model Legume *Medicago truncatula* (Jemalong A17) Is Low Compared to *Medicago sativa*. *J. Plant Physiol.* **2010**, *167*, 683–692. [[CrossRef](#)] [[PubMed](#)]
114. Yang, Y.; Zhao, Q.; Li, X.; Ai, W.; Liu, D.; Qi, W.; Zhang, M.; Yang, C.; Liao, H. Characterization of Genetic Basis on Synergistic Interactions between Root Architecture and Biological Nitrogen Fixation in Soybean. *Front. Plant Sci.* **2017**, *8*, 1466. [[CrossRef](#)] [[PubMed](#)]
115. Smith, G.; Knight, W.; Peterson, H. Variation among Inbred Lines of Crimson Clover for N₂ Fixation (C₂H₂) Efficiency 1. *Crop Sci.* **1982**, *22*, 716–719. [[CrossRef](#)]
116. Luna, R.; Planchon, C. Genotype × *Bradyrhizobium japonicum* Strain Interactions in Dinitrogen Fixation and Agronomic Traits of Soybean (*Glycine max* L. Merr.). *Euphytica* **1995**, *86*, 127–134. [[CrossRef](#)]
117. Peoples, M.B.; Baldock, J.A. Nitrogen Dynamics of Pastures: Nitrogen Fixation Inputs, the Impact of Legumes on Soil Nitrogen Fertility, and the Contributions of Fixed Nitrogen to Australian Farming Systems. *Aust. J. Exp. Agric.* **2001**, *41*, 327. [[CrossRef](#)]
118. Yuan, S.; Li, R.; Chen, S.; Chen, H.; Zhang, C.; Chen, L.; Hao, Q.; Shan, Z.; Yang, Z.; Qiu, D.; et al. RNA-Seq Analysis of Differential Gene Expression Responding to Different *Rhizobium* Strains in Soybean (*Glycine max*) Roots. *Front. Plant Sci.* **2016**, *7*, 721. [[CrossRef](#)]
119. Cannon, S.B.; McKain, M.R.; Harkess, A.; Nelson, M.N.; Dash, S.; Deyholos, M.K.; Peng, Y.; Joyce, B.; Stewart, C.N.; Rolf, M.; et al. Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. *Mol. Biol. Evol.* **2015**, *32*, 193–210. [[CrossRef](#)]
120. Thilakarathna, M.S.; Papadopoulos, Y.A.; Grimmett, M.; Fillmore, S.A.E.; Crouse, M.; Prithiviraj, B. Red Clover Varieties Show Nitrogen Fixing Advantage during the Early Stages of Seedling Development. *Can. J. Plant Sci.* **2018**, *98*, 517–526. [[CrossRef](#)]
121. Soper, F.M.; Simon, C.; Jauss, V. Measuring Nitrogen Fixation by the Acetylene Reduction Assay (ARA): Is 3 the Magic Ratio? *Biogeochemistry* **2021**, *152*, 345–351. [[CrossRef](#)]
122. Roughley, R.; Dart, P. Reduction of Acetylene by Nodules of *Trifolium Subterraneum* as Affected by Root Temperature, *Rhizobium* Strain and Host Cultivar. *Arch. Für Mikrobiol.* **1969**, *69*, 171–179. [[CrossRef](#)]
123. Bergersen, F.J. The Quantitative Relationship between Nitrogen Fixation and the Acetylene-Reduction Assay. *Aust. J. Biol. Sci.* **1970**, *23*, 1015–1026. [[CrossRef](#)]
124. Anderson, M.D.; Ruess, R.W.; Uliassi, D.D.; Mitchell, J.S. Estimating N₂ Fixation in Two Species of *Alnus* in Interior Alaska Using Acetylene Reduction and ¹⁵N₂ Uptake. *Ecoscience* **2004**, *11*, 102–112. [[CrossRef](#)]
125. Ayanaba, A.; Lawson, T. Diurnal Changes in Acetylene Reduction in Field-Grown Cowpeas and Soybeans. *Soil Biol. Biochem.* **1977**, *9*, 125–129. [[CrossRef](#)]
126. Zapata, F.; Danso, F.; Hardarson, G.; Fried, M. Nitrogen Fixation and Translocation in Field-Grown Fababean. *Agron. J.* **1987**, *79*, 505–509. [[CrossRef](#)]
127. Vessey, J.K. Measurement of Nitrogenase Activity in Legume Root Nodules: In Defense of the Acetylene Reduction Assay. *Plant Soil* **1994**, *158*, 151–162. [[CrossRef](#)]
128. Young, N.D.; Cannon, S.B.; Sato, S.; Kim, D.; Cook, D.R.; Town, C.D.; Roe, B.A.; Tabata, S. Sequencing the Genespaces of *Medicago truncatula* and *Lotus japonicus*. *Plant Physiol.* **2005**, *137*, 1174–1181. [[CrossRef](#)]
129. Mumm, R.; Posthumus, M.A.; Dicke, M. Significance of Terpenoids in Induced Indirect Plant Defence against Herbivorous Arthropods. *Plant Cell Environ.* **2008**, *31*, 575–585. [[CrossRef](#)]
130. Umehara, M.; Hanada, A.; Yoshida, S.; Akiyama, K.; Arite, T.; Takeda-Kamiya, N.; Magome, H.; Kamiya, Y.; Shirasu, K.; Yoneyama, K.; et al. Inhibition of Shoot Branching by New Terpenoid Plant Hormones. *Nature* **2008**, *455*, 195–200. [[CrossRef](#)]
131. Zhou, Y.; Stuart-Williams, H.; Grice, K.; Kayler, Z.E.; Zavadlav, S.; Vogts, A.; Rommerskirchen, F.; Farquhar, G.D.; Gessler, A. Allocate Carbon for a Reason: Priorities Are Reflected in the ¹³C/¹²C Ratios of Plant Lipids Synthesized via Three Independent Biosynthetic Pathways. *Phytochemistry* **2015**, *111*, 14–20. [[CrossRef](#)] [[PubMed](#)]

132. Ali, M.; Miao, L.; Hou, Q.; Darwish, D.B.; Aldahe, S.S.; Ali, A.; Benedito, V.A.; Tadege, M.; Wang, X.; Zhao, J. Overexpression of Terpenoid Biosynthesis Genes From Garden Sage (*Salvia officinalis*) Modulates Rhizobia Interaction and Nodulation in Soybean. *Front. Plant Sci.* **2021**, *12*, 783269. [[CrossRef](#)] [[PubMed](#)]
133. Foo, E.; Davies, N.W. Strigolactones Promote Nodulation in Pea. *Planta* **2011**, *234*, 1073–1081. [[CrossRef](#)] [[PubMed](#)]
134. Foo, E.; Yoneyama, K.; Hugill, C.J.; Quittenden, L.J.; Reid, J.B. Strigolactones and the Regulation of Pea Symbioses in Response to Nitrate and Phosphate Deficiency. *Mol. Plant* **2013**, *6*, 76–87. [[CrossRef](#)] [[PubMed](#)]
135. ur Rehman, N.; Ali, M.; Ahmad, M.Z.; Liang, G.; Zhao, J. Strigolactones Promote Rhizobia Interaction and Increase Nodulation in Soybean (*Glycine max*). *Microb. Pathog.* **2018**, *114*, 420–430. [[CrossRef](#)]
136. Akbar, A.; Han, B.; Khan, A.H.; Feng, C.; Ullah, A.; Khan, A.S.; He, L.; Yang, X. A Transcriptomic Study Reveals Salt Stress Alleviation in Cotton Plants upon Salt Tolerant PGPR Inoculation. *Environ. Exp. Bot.* **2022**, *200*, 104928. [[CrossRef](#)]
137. Hubbard, C.J.; Li, B.; McMinn, R.; Brock, M.T.; Maignien, L.; Ewers, B.E.; Kliebenstein, D.; Weinig, C. The Effect of Rhizosphere Microbes Outweighs Host Plant Genetics in Reducing Insect Herbivory. *Mol. Ecol.* **2019**, *28*, 1801–1811. [[CrossRef](#)]
138. Seifert, G.J. Nucleotide Sugar Interconversions and Cell Wall Biosynthesis: How to Bring the inside to the Outside. *Curr. Opin. Plant Biol.* **2004**, *7*, 277–284. [[CrossRef](#)]
139. O'Donoghue, E.M.; Somerfield, S.D.; Watson, L.M.; Brummell, D.A.; Hunter, D.A. Galactose Metabolism in Cell Walls of Opening and Senescing Petunia Petals. *Planta* **2009**, *229*, 709–721. [[CrossRef](#)]
140. Lee, D.-K.; Ahn, S.; Cho, H.Y.; Yun, H.Y.; Park, J.H.; Lim, J.; Lee, J.; Kwon, S.W. Metabolic Response Induced by Parasitic Plant-Fungus Interactions Hinder Amino Sugar and Nucleotide Sugar Metabolism in the Host. *Sci. Rep.* **2016**, *6*, 37434. [[CrossRef](#)]
141. La Camera, S.; Gouzerh, G.; Dhondt, S.; Hoffmann, L.; Fritig, B.; Legrand, M.; Heitz, T. Metabolic Reprogramming in Plant Innate Immunity: The Contributions of Phenylpropanoid and Oxylinipin Pathways. *Immunol. Rev.* **2004**, *198*, 267–284. [[CrossRef](#)] [[PubMed](#)]
142. Bais, H.P.; Vepachedu, R.; Gilroy, S.; Callaway, R.M.; Vivanco, J.M. Allelopathy and Exotic Plant Invasion: From Molecules and Genes to Species Interactions. *Science* **2003**, *301*, 1377–1380. [[CrossRef](#)] [[PubMed](#)]
143. Vogt, T. Phenylpropanoid Biosynthesis. *Mol. Plant* **2010**, *3*, 2–20. [[CrossRef](#)] [[PubMed](#)]
144. Peck, M.C.; Fisher, R.F.; Long, S.R. Diverse Flavonoids Stimulate NodD1 Binding to *Nod* Gene Promoters in *Sinorhizobium meliloti*. *J. Bacteriol.* **2006**, *188*, 5417–5427. [[CrossRef](#)]
145. Wasson, A.P.; Pellerone, F.I.; Mathesius, U. Silencing the Flavonoid Pathway in *Medicago truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. *Plant Cell* **2006**, *18*, 1617–1629. [[CrossRef](#)]
146. Mithöfer, A. Suppression of Plant Defence in Rhizobia–Legume Symbiosis. *Trends Plant Sci.* **2002**, *7*, 440–444. [[CrossRef](#)]
147. Staehelin, C.; Krishnan, H.B. Nodulation Outer Proteins: Double-Edged Swords of Symbiotic Rhizobia. *Biochem. J.* **2015**, *470*, 263–274. [[CrossRef](#)]
148. Avenhaus, U.; Cabeza, R.A.; Liese, R.; Lingner, A.; Dittert, K.; Salinas-Riester, G.; Pommerenke, C.; Schulze, J. Short-Term Molecular Acclimation Processes of Legume Nodules to Increased External Oxygen Concentration. *Front. Plant Sci.* **2016**, *6*, 1133. [[CrossRef](#)]
149. Milligan, A.J.; Berman-Frank, I.; Gerchman, Y.; Dismukes, G.C.; Falkowski, P.G. Light-Dependent Oxygen Consumption in Nitrogen-Fixing Cyanobacteria Plays a Key Role in Nitrogenase Protection. *J. Phycol.* **2007**, *43*, 845–852. [[CrossRef](#)]
150. Mylona, P.; Pawlowski, K.; Bisseling, T. Symbiotic Nitrogen Fixation. *Plant Cell.* **1995**, *7*, 869–885. [[CrossRef](#)]
151. Young, N.D.; Debelle, F.; Oldroyd, G.E.D.; Geurts, R.; Cannon, S.B.; Udvardi, M.K.; Benedito, V.A.; Mayer, K.F.X.; Gouzy, J.; Schoof, H.; et al. The *Medicago* Genome Provides Insight into the Evolution of Rhizobial Symbioses. *Nature* **2011**, *480*, 520–524. [[CrossRef](#)] [[PubMed](#)]
152. Mandal, D.; Sinharoy, S. A Toolbox for Nodule Development Studies in Chickpea: A Hairy-Root Transformation Protocol and an Efficient Laboratory Strain of *Mesorhizobium* Sp. *Mol. Plant-Microbe Interact.* **2019**, *32*, 367–378. [[CrossRef](#)] [[PubMed](#)]
153. Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [[CrossRef](#)] [[PubMed](#)]
154. Li, Q.-G.; Zhang, L.; Li, C.; Dunwell, J.M.; Zhang, Y.-M. Comparative Genomics Suggests That an Ancestral Polyploidy Event Leads to Enhanced Root Nodule Symbiosis in the Papilionoideae. *Mol. Biol. Evol.* **2013**, *30*, 2602–2611. [[CrossRef](#)]
155. Hanada, K.; Zou, C.; Lehti-Shiu, M.D.; Shinozaki, K.; Shiu, S.-H. Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli. *Plant Physiol.* **2008**, *148*, 993–1003. [[CrossRef](#)] [[PubMed](#)]
156. Wang, Y.; Ficklin, S.P.; Wang, X.; Feltus, F.A.; Paterson, A.H. Large-Scale Gene Relocations Following an Ancient Genome Triplication Associated with the Diversification of Core Eudicots. *PLoS ONE* **2016**, *11*, e0155637. [[CrossRef](#)]
157. Fuller, T.; Langfelder, P.; Presson, A.; Horvath, S. Review of Weighted Gene Coexpression Network Analysis. In *Handbook of Statistical Bioinformatics*; Lu, H.H.-S., Schölkopf, B., Zhao, H., Eds.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2011; pp. 369–388. ISBN 978-3-642-16344-9.
158. Battenberg, K.; Potter, D.; Tabuloc, C.A.; Chiu, J.C.; Berry, A.M. Comparative Transcriptomic Analysis of Two Actinorhizal Plants and the Legume *Medicago truncatula* Supports the Homology of Root Nodule Symbioses and Is Congruent With a Two-Step Process of Evolution in the Nitrogen-Fixing Clade of Angiosperms. *Front. Plant Sci.* **2018**, *9*, 1256. [[CrossRef](#)]
159. Soltis, D.E.; Soltis, P.S.; Morgan, D.R.; Swensen, S.M.; Mullin, B.C.; Dowd, J.M.; Martin, P.G. Chloroplast Gene Sequence Data Suggest a Single Origin of the Predisposition for Symbiotic Nitrogen Fixation in Angiosperms. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2647–2651. [[CrossRef](#)]

160. Fedorova, M.; van de Mortel, J.; Matsumoto, P.A.; Cho, J.; Town, C.D.; VandenBosch, K.A.; Gantt, J.S.; Vance, C.P. Genome-Wide Identification of Nodule-Specific Transcripts in the Model Legume *Medicago truncatula*. *Plant Physiol.* **2002**, *130*, 519–537. [[CrossRef](#)]
161. El Yahyaoui, F.; Küster, H.; Ben Amor, B.; Hohnjec, N.; Pühler, A.; Becker, A.; Gouzy, J.; Vernié, T.; Gough, C.; Niebel, A.; et al. Expression Profiling in *Medicago truncatula* Identifies More Than 750 Genes Differentially Expressed during Nodulation, Including Many Potential Regulators of the Symbiotic Program. *Plant Physiol.* **2004**, *136*, 3159–3176. [[CrossRef](#)]
162. Lee, H.; Hur, C.-G.; Oh, C.J.; Kim, H.B.; Park, S.-Y.; An, C.S. Analysis of the Root Nodule-Enhanced Transcriptome in Soybean. *Mol. Cells* **2004**, *18*, 53–62. [[PubMed](#)]
163. Liu, J.; Rasing, M.; Zeng, T.; Klein, J.; Kulikova, O.; Bisseling, T. NIN Is Essential for Development of Symbiosomes, Suppression of Defence and Premature Senescence in *Medicago truncatula* Nodules. *New Phytol.* **2021**, *230*, 290–303. [[CrossRef](#)] [[PubMed](#)]
164. Benedito, V.A.; Torres-Jerez, I.; Murray, J.D.; Andriankaja, A.; Allen, S.; Kakar, K.; Wandrey, M.; Verdier, J.; Zuber, H.; Ott, T.; et al. A Gene Expression Atlas of the Model Legume *Medicago truncatula*. *Plant J.* **2008**, *55*, 504–513. [[CrossRef](#)]
165. Høgslund, N.; Radutoiu, S.; Krusell, L.; Voroshilova, V.; Hannah, M.A.; Goffard, N.; Sanchez, D.H.; Lippold, F.; Ott, T.; Sato, S.; et al. Dissection of Symbiosis and Organ Development by Integrated Transcriptome Analysis of *Lotus japonicus* Mutant and Wild-Type Plants. *PLoS ONE* **2009**, *4*, e6556. [[CrossRef](#)] [[PubMed](#)]
166. Libault, M.; Farmer, A.; Joshi, T.; Takahashi, K.; Langley, R.J.; Franklin, L.D.; He, J.; Xu, D.; May, G.; Stacey, G. An Integrated Transcriptome Atlas of the Crop Model *Glycine max*, and Its Use in Comparative Analyses in Plants: Soybean Transcriptome Atlas. *Plant J.* **2010**, *63*, 86–99. [[CrossRef](#)]
167. Maunoury, N.; Redondo-Nieto, M.; Bourcy, M.; Van de Velde, W.; Alunni, B.; Laporte, P.; Durand, P.; Agier, N.; Marisa, L.; Vaubert, D.; et al. Differentiation of Symbiotic Cells and Endosymbionts in *Medicago truncatula* Nodulation Are Coupled to Two Transcriptome-Switches. *PLoS ONE* **2010**, *5*, e9519. [[CrossRef](#)]
168. Breakspear, A.; Liu, C.; Roy, S.; Stacey, N.; Rogers, C.; Trick, M.; Morieri, G.; Mysore, K.S.; Wen, J.; Oldroyd, G.E.D.; et al. The Root Hair “Infectome” of *Medicago truncatula* Uncovers Changes in Cell Cycle Genes and Reveals a Requirement for Auxin Signaling in Rhizobial Infection. *Plant Cell* **2014**, *26*, 4680–4701. [[CrossRef](#)]
169. Larrainzar, E.; Riely, B.K.; Kim, S.C.; Carrasquilla-Garcia, N.; Yu, H.-J.; Hwang, H.-J.; Oh, M.; Kim, G.B.; Surendrarao, A.K.; Chasman, D.; et al. Deep Sequencing of the *Medicago truncatula* Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol.* **2015**, *169*, 233–265. [[CrossRef](#)]
170. Damiani, I.; Drain, A.; Guichard, M.; Balzergue, S.; Boscari, A.; Boyer, J.-C.; Brunaud, V.; Cottaz, S.; Rancurel, C.; Da Rocha, M.; et al. Nod Factor Effects on Root Hair-Specific Transcriptome of *Medicago truncatula*: Focus on Plasma Membrane Transport Systems and Reactive Oxygen Species Networks. *Front. Plant Sci.* **2016**, *7*, 794. [[CrossRef](#)]
171. Jardinaud, M.-F.; Boivin, S.; Rodde, N.; Catrice, O.; Kisiala, A.; Lepage, A.; Moreau, S.; Roux, B.; Cottret, L.; Sallet, E.; et al. A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the *Medicago truncatula* Root Epidermis. *Plant Physiol.* **2016**, *171*, 2256–2276. [[CrossRef](#)]
172. Schiessl, K.; Lilley, J.L.S.; Lee, T.; Tamvakis, I.; Kohlen, W.; Bailey, P.C.; Thomas, A.; Luptak, J.; Ramakrishnan, K.; Carpenter, M.D.; et al. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in *Medicago truncatula*. *Curr. Biol.* **2019**, *29*, 3657–3668. [[CrossRef](#)] [[PubMed](#)]
173. Rudaya, E.S.; Kozyulina, P.Y.; Pavlova, O.A.; Dolgikh, A.V.; Ivanova, A.N.; Dolgikh, E.A. Regulation of the Later Stages of Nodulation Stimulated by IPD3/CYCLOPS Transcription Factor and Cytokinin in Pea *Pisum sativum* L. *Plants* **2021**, *11*, 56. [[CrossRef](#)] [[PubMed](#)]