Reference nodule transcriptomes for *Melilotus officinalis* and *Medicago sativa* cv. Algonquin

Rui Huang | Wayne A. Snedden | George C. diCenzo

Abstract

Host/symbiont compatibility is a hallmark of the symbiotic nitrogen-fixing interaction between rhizobia and legumes, mediated in part by plant-produced nodule-specific cysteine-rich (NCR) peptides and the bacterial BacA membrane protein that can act as a NCR peptide transporter. In addition, the genetic and metabolic properties supporting symbiotic nitrogen fixation often differ between compatible partners, including those sharing a common partner, highlighting the need for multiple study systems. Here, we report high-quality nodule transcriptome assemblies for *Medicago sativa* cv. Algonquin and *Melilotus officinalis*, two legumes able to form compatible symbioses with *Sinorhizobium meliloti*. The compressed *M. sativa* and *M. officinalis* assemblies consisted of 79,978 and 64,593 contigs, respectively, of which 33,341 and 28,278 were assigned putative annotations, respectively. As expected, the two transcriptomes showed broad similarity at a global level. We were particularly interested in the NCR peptide profiles of these plants, as these peptides drive bacterial differentiation during the symbiosis. A total of 412 and 308 NCR peptides were predicted from the *M. sativa* and *M. officinalis* transcriptomes, respectively, with approximately 9% of the transcriptome of both species consisting of NCR transcripts. Notably, transcripts encoding highly cationic NCR peptides (isoelectric point > 9.5), which are known to have antimicrobial properties, were 2-fold more abundant in *M. sativa* than in *M. officinalis*, and 27-fold more abundant when considering only NCR peptides in the six-cysteine class. We hypothesize that the difference in abundance of highly cationic NCR peptides explains our previous observation that some rhizobial *bacA* alleles which can support symbiosis with *M. officinalis* are unable to support symbiosis with *M. sativa*.

**KEYWORDS**

legumes, NCR peptides, rhizobia, symbiotic nitrogen fixation, transcriptomics

1 | INTRODUCTION

Leguminous plants are able to establish symbiotic relationships with a group of soil bacteria known as rhizobia. During the interaction, the rhizobia are located within a specialized organ known as a nodule where they fix atmospheric nitrogen into ammonia in exchange for reduced carbon from their host. Symbiosis is initiated following an exchange of chemical signals in the rhizosphere between compatible...
partners (Oldroyd, 2013); legumes secrete flavonoids that attract soil rhizobia and induce expression of rhizobial nod genes, leading to rhizobial production of chito-oligosaccharide Nod factors that elicit the nodulation process by legumes. This process involves the curling of root hairs to trap rhizobia and the formation of infection threads within which rhizobia divide and move toward the root cortical layer (Gage, 2004). Rhizobia released from infection threads are internalized by nodule cells, where they develop into mature N₂-fixing bacteroids. In some legumes, such as those belonging to the Inverted Repeat Lacking Clade (IRLC), the rhizobia undergo an irreversible host-induced process known as terminal differentiation that is largely driven by a unique class of legume proteins known as nodule-specific cysteine-rich (NCR) peptides (Van de Velde et al., 2010). Terminal differentiation involves cell enlargement, genome endoreplication, and increased membrane permeability and is thought to increase the efficiency of N₂-fixation (Haag & Mergaert, 2019; Lamouche et al., 2019; Mergaert et al., 2006).

Not all rhizobium/legume pairings are compatible (Pueppke & Broughton, 1999; Wilson, 1939). Partner compatibility is determined by numerous factors impacting both early and late stages of the symbiotic interaction (Walker et al., 2020). The flavonoids secreted by legumes vary, as does the ability of rhizobia to respond to different flavonoids (Kossiak et al., 1987; Maxwell et al., 1989; Pueppke et al., 1998; Recourt et al., 1991). Similarly, the Nod factor produced by rhizobia differ and legume hosts respond only to Nod factors with specific structures (DHaeze & Holsters, 2002). Moreover, legume infection depends on rhizobia producing host-compatible exopolysaccharide molecules (Finan et al., 1985; Leigh et al., 1985), and variations in exopolysaccharide structure can impact specificity at the level of plant ecotype and bacterial strain (Simsek et al., 2007). In addition, some rhizobia secrete effector proteins that induce effector-triggered immune responses in a cultivar-specific manner, thereby influencing host range (Tsukui et al., 2013; Tsurumaru et al., 2015; Yang et al., 2010). Moreover, for IRLC legumes, an effective symbiotic interaction requires compatibility between the host-produced NCR peptides and the rhizobial membrane protein BacA (diCenzo et al., 2017; Wang et al., 2017, 2018; Yang et al., 2017).

NCR peptides are a large class of legume-specific proteins, with ~600 members in Medicago truncatula (Young et al., 2011). These proteins display little conservation in amino acid composition but possess four or six cysteine residues at conserved positions (Mergaert et al., 2003). The length of mature NCR peptides varies from about 20 to 50 amino acids and includes two or three disulfide bridges (Maroti et al., 2015). NCR peptides can be classified as cationic (isoelectric point [pI] ≥ 8), neutral (6 ≤ pI < 8), or anionic (pI < 6) (Maroti et al., 2015). Highly cationic peptides (pI ≥ 9.0) display antimicrobial activity in vitro, likely through disrupting microbial membranes, thereby leading to permeabilization and cell lysis (Maroti et al., 2011; Tiricz et al., 2013). In planta, NCR peptides are required for rhizobium terminal differentiation and an effective symbiosis in IRLC legumes (Van de Velde et al., 2010). Deletion of individual NCR genes has shown that at least two of the ~600 NCR peptides in M. truncatula are essential for N₂-fixation (Horváth et al., 2015; Kim et al., 2015); however, mutation of other NCR genes resulted in N₂-fixation in previously incompatible symbioses (Wang et al., 2017; Yang et al., 2017), demonstrating the role of NCR peptides in partner compatibility.

The ability of rhizobia to establish an effective symbiosis with IRLC legumes requires the membrane protein BacA (Glazerbrook et al., 1993). BacA is a peptide transporter whose deletion results in multiple phenotypes including increased resistance to bleomycin, and gentamicin increased sensitivity to detergents, and altered membrane composition (Ferguson et al., 2004, 2002; Ichige & Walker, 1997; Marlow et al., 2009). In addition, bacA deletion mutants are both unable to import NCR peptides and show increased sensitivity to cationic NCR peptides (Barrière et al., 2017; Guefrachi et al., 2015; Haag et al., 2011). Moreover, rhizobium bacA mutants are unable to fix nitrogen in symbiosis with IRLC legumes; instead, the rhizobia are quickly killed in a NCR peptide-dependent fashion upon release from the infection threads (Glazerbrook et al., 1993; Haag et al., 2011). Intriguingly, BacA appears to be a host-range determinant factor in IRLC legumes. For example, studies have shown that introduction of the bacA or bacA-like genes of Mesorhizobium loti and Bradyrhizobium species into a Sinorhizobium melloti bacA mutant is insufficient to allow N₂-fixation during interaction with IRLC legumes of the genus Medicago (Guefrachi et al., 2015; Maruyu & Saeki, 2010). Similarly, we previously demonstrated that replacement of the S. melloti bacA with the bacA alleles of Sinorhizobium fredii NGR234 or Rhizobium leguminosarum bv. viciae 3841 does not allow for N₂-fixation during symbiosis with Medicago sativa (alfalfa) but does support N₂-fixation on the IRLC legumes Melilotus alba (white sweet clover) and Mellotus officinalis (yellow sweet clover) (diCenzo et al., 2017 and Table S1).

In addition to the above-noted comparison, several symbiotic differences have been observed when S. melloti mutants interact with Medicago versus Melilotus plants (diCenzo et al., 2015; Geddes et al., 2021; Honma & Ausubel, 1987; Zamani et al., 2017), suggesting that Melilotus plants are a valuable secondary model system to study the symbiotic properties of S. melloti. To further develop M. officinalis as a model species for studying symbiosis, here we report a reference nodule transcriptome for M. officinalis. We further compare the characteristics and the expression of NCR genes between M. officinalis and M. sativa to investigate whether the ability of certain bacA alleles to support symbiosis with Melilotus but not Medicago plants is correlated with differences in the NCR peptide profile of these genera.

2 | MATERIALS AND METHODS

2.1 | Plant materials and sample collection

M. sativa cv. Algonquin (alfalfa) and M. officinalis (yellow blossom sweet clover) seeds were purchased from Speare Seeds Limited (Harriston, Ontario, Canada). Seeds were surface sterilized with 95% ethanol for 5 min followed by 2.5% hypochlorite for 20 min and then soaked in sterile double-distilled water (ddH₂O) for 1 h. The sterilized seeds were plated on 1X water agar plates and incubated at room temperature in the dark for 2 days. Five germinated seeds were soaked in sterile double-distilled water (ddH₂O) for 1 h. The sterilized seeds were plated on 1X water agar plates and incubated at room temperature in the dark for 2 days. Five germinated seeds were
placed in autoclaved Leonard Assemblies consisting of two Magenta Jars with a cotton wick extending from the top jar (containing vermiculite mixed with silica sand [1:1 w/v]) into the bottom jar (containing 250 ml of Jensen’s media; Jensen, 1942) and then incubated in a Conviron growth chamber for two nights. Wildtype S. melloti strain Rm2011 was grown overnight at 30°C in LBmc broth (10 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, 2.5 mM CaCl₂, and 2.5 mM MgCl₂), washed with 0.85% NaCl, and diluted to a density of ~1 x 10⁸ CFU ml⁻¹ in sterile ddH₂O. Ten milliliters of cell suspension was then added to each Leonard Assembly. Plants were grown in a Conviron growth chamber with a day (18 h, 21°C, light intensity of 300 μmol m⁻² s⁻¹) and night (6 h, 17°C) cycle. Root nodules were collected 4 weeks postinoculation and immediately flash frozen with liquid N₂ and stored at −80°C until use. All nodules collected from plants grown in the same Leonard Assembly were stored in a single tube and treated as one replicate. The shoots from each pot were dried at 60°C for 2 weeks prior to measuring shoot dry weight (Table S2).

2.2 | RNA extraction and sequencing

Total RNA from three replicates of frozen M. sativa and M. officinalis nodule tissue was extracted using Direct-zol RNA miniprep kits (Zymo Research) according to the manufacturer’s protocol. Total RNA samples were treated with DNase I (New England Biolabs) to degrade any contaminating DNA according to the manufacturer’s protocol and the RNA again purified using Direct-zol RNA miniprep kits. Total RNA samples were run on a MOPS-formaldehyde agarose gel (119 ml MOPS buffer [200 mM MOPS, 80 mM sodium acetate, 10 mM EDTA, pH 7.0, in DEPC-treated ddH₂O], 6 ml formaldehyde, 2.2 μl of 1:1 glycerol mixed with silica sand) into the bottom jar (containing 300 ml of Jensen’s media) and then incubated in a Conviron growth chamber for two nights. Total RNA from three replicates of frozen M. officinalis nodule tissue was checked on an Agilent Bioanalyzer chip following Illumina’s recommended protocol to generate paired-end reads of 125-bases in length.

2.3 | Sequencing read trimming

Preprocessing of the raw reads was performed to ensure only high-quality data were used for de novo transcriptome assembly and differential expression analysis. Read quality was initially evaluated using FastQC version 0.11.9 (Andrews et al., 2015), following which errors in raw reads were identified and corrected by the k-mer-based method of Rcorrector version 1.0.4 (Song & Florea, 2015). The FilterUncorrectablePEfastq.py script (github.com/harvardinformatics/TranscriptomeAssemblyTools/) was used to remove any read pairs where at least one read had an unfixable error identified by Rcorrector. Adaptors sequences, short reads (< 25 bp), and low-quality reads (Q score < 20) were removed using Trim Galore version 0.6.6 (bioinformatics.babraham.ac.uk/projects/trim_galore/), which is a wrapper calling cutadapt version 3.2 (Martin, 2011) and FastQC. The processed reads were further trimmed by Trimmomatic version 0.4.0 (Bolger et al., 2014) included in the Trinity software distribution with the following parameters: SLIDINGWINDOW:5:20 LEADING:3 TRAILING:3 MINLEN:25. The quality and presence of adaptors in the preprocessed reads were then examined using FastQC. Between 38 and 94 million paired-end reads remained per sample following preprocessing (Table S3), with a total of 174,707,055 and 119,333,821 paired-end reads (~43.7 and ~29.8 Gb, respectively) remaining for M. sativa and M. officinalis, respectively (Table S4).

2.4 | Transcriptome de novo assembly and quality control

The nodule transcriptomes of M. sativa and M. officinalis were de novo assembled following the same procedure. First, the preprocessed reads from the triplicate samples were simultaneously provided to Trinity version 2.9.0 for assembly without genome guidance (Grabherr et al., 2011). Then, the assembled contigs were clustered into gene-level clusters using SuperTranscripts (Davidson et al., 2017). Gene isoforms were identified by Corset version 1.09 with the log likelihood ratio threshold set to very high (Davidson & Oshlack, 2014). Based on the Corset clusters, Lacey version 1.14.1 was used to merge the gene isoforms into single long supertranscripts meant to provide a gene-like view of the transcriptome (Davidson et al., 2017).

Multiple methods were used to examine the quality of the Trinity and SuperTranscript assemblies. First, the alignment rates of the preprocessed reads to the assemblies were inspected using STAR version 2.7.8a with the two-pass mode that is more sensitive to alternative splicing (Dobin et al., 2013). Second, assembly statistics such as N50 and number of contigs were calculated using the seqstats software (github.com/clwgg/seqstats). Third, the completeness of the assemblies was evaluated using BUSCO version 5.1.2, run separately using the OrthoDB v10 ‘Fabales’ and ‘Viridiplantae’ reference
databases (Seppey et al., 2019). The assemblies were also compared with the S. meliloti Rm1110 genome (Sallet et al., 2013) using BLASTn version 2.5.0 + (Camacho et al., 2009), which confirmed the absence of contaminating S. meliloti transcripts in the assemblies. Finally, the M. sativa de novo assembly was aligned to a publicly available genome of M. sativa cultivar XinJiangDaYe (Chen et al., 2020) with MUMmer version 4.0+. and 87.3% of transcripts were successfully aligned to the genome.

2.5 Transcriptome annotation

Coding regions within the supertranscripts were predicted by Trans-Decoder version 5.5.0 (github.com/TransDecoder/TransDecoder), using the results of BLASTp searches (E-value cutoff of 1e-5) against the Uniport database as ORF retention criteria (2021 January release) (The UniProt Consortium, 2021). The functional annotation of the predicted coding sequences then proceeded via three steps. First, BLAST bidirectional best hits between the M. truncatula A17 proteome (assembly release r5.0 1.7) (Pecrix et al., 2018) and the longest predicted protein isoform of each contig of our transcriptome assemblies were identified using BLASTp (E-value cutoff of 1e-5, culling limit 1). For all bidirectional best hits, the annotations from M. truncatula A17 were transferred to the corresponding contigs of the M. sativa or M. officinalis transcriptome. Second, all predicted protein isoforms of each contig in each transcriptome assembly were annotated using eggNOG-mapper version 2.1.0 with DIAMOND version 2.0.4 and the Viridiplantae dataset (E-value cutoff of 1e-3) (Buchfink et al., 2015; Huerta-Cepas et al., 2019). Third, for each contig not annotated by BLAST or eggNOG-mapper, the hmssearch function of HMMER version 3.3.2 was used to search all predicted protein isoforms against the complete set of hidden Markov models (HMMs) from the Pfam database and separately against the TIGRFAM version 15.0 HMM database (E-value cutoff of 1e-5) (Eddy, 2011; Haft et al., 2012; Mistry et al., 2021), and results were filtered to remove annotations with a Bit-score < 50. For repetitive annotations from isoforms of a gene, only the consensus annotations were retained. For contigs successfully annotated by more than one of the annotation methods, results from the bidirectional BLAST took priority, followed by the results of eggNOG-mapper, then the Pfam searches, and finally the TIGRFAM searches.

2.6 NCR peptide identification

Considering the high degree of sequence diversity of NCR peptide sequences, the functional annotation methods described above were not sufficiently sensitive to discover genes encoding NCR peptides in the assemblies. Therefore, the SPADA pipeline was used to identify NCR peptides (Zhou et al., 2013). SPADA is specialized to predict cysteine-rich peptides in plant genomes and is distributed with a M. truncatula prediction model. Cysteine-rich peptides in the M. sativa and M. officinalis assemblies were predicted using the SPADA pipeline with the following software: HMMER version 3.0, Augustus version 2.6, GeneWise version 2.2.0, GeneMark.hmm euclaryotic version 3.54, GlimmerHMM version 3.0.1, and GenelID version 1.1 (Birney et al., 2004; Blanco et al., 2007; Lukashin & Borodovsky, 1998; Stanke et al., 2006). The putative NCR peptide sequences were filtered to remove those without a signal peptide and then further filtered based on the E-value (cutoff of 1e-5) and hmm score (cutoff of 50). Filtered sequences were then verified via hmmsearch against the Pfam database, and they were aligned using Clustal Omega version 1.2.4 (Sievers et al., 2011) to ensure the presence of the signature cysteine motif and N terminal signal peptide that are present in bona fide NCR peptides.

2.7 NCR peptides classification and clustering

To predict the lengths of mature NCR peptides, signalP version 4.1g with the notm network was used to predicted cleavage sites and extract mature NCR peptides (Petersen et al., 2011), and the number of cysteine residues in each motif were counted. The pl values of the NCR peptides were predicted using the pIR R package, and the value for each peptide was calculated based on the mean values from all prediction methods excluding the highest and lowest values (Audain et al., 2016). The M. sativa, M. officinalis, and M. truncatula NCR peptides were clustered using CD-HIT-2D with an identity threshold of 0.7 (Li & Godzik, 2006).

2.8 Gene-expression level estimation and differential expression analysis

Gene-expression levels were estimated individually for each library based on transcript abundance estimation using salmon version 0.12.0 in mapping-based mode (library type automatic, validate Mapping) (Patro et al., 2017) and the reference transcriptomes produced as described above. R package deseq2 version 1.32.0 (Love et al., 2014) was used to perform differential expression analysis between M. sativa and M. officinalis, using the raw counts for each replicate from salmon, the length of each gene in each species as an additional parameter during normalization, and limiting the analysis to one-to-one orthologs identified by OrthoFinder version 2.5.2 (Emms & Kelly, 2019). OrthoFinder was run with default settings using the total predicted M. sativa and M. officinalis proteomes including all isoforms, following which orthologs were reduced to one per supertranscript.

2.9 Gene ontology term analysis

The Gene Ontology (GO) terms for M. sativa and M. officinalis were obtained from the M. truncatula A17 proteome (assembly release r5.0 1.7) and annotations from eggNOG-mapper. For transcripts annotated with GO terms from both sources, the consensus GO term
annotations were retained. Then, the GO terms were reduced based on the Generic GO subset (download 10 August 2021).

2.10 | Software information

All analyses were performed in an Ubuntu 20.04.2 LTS (Linux 5.8.0-48-generic) operation system or on the Compute Canada Graham cluster. Custom scripts were written in Python version 3.8.5, and bash. R version 3.6.3 was used during data analysis.

3 | RESULTS AND DISCUSSION

3.1 | Reference nodule transcriptomes for *M. officinalis* and *M. sativa* cv. Algonquin

To establish reference nodule transcriptomes of *M. sativa* cv. Algonquin and *M. officinalis* during symbiosis with *S. meliloti* Rm2011, the poly-A enriched RNA from triplicate samples was sequenced using Illumina technology (2 × 125 bp paired-end reads), generating ∼50 Gb (∼202 million paired-end reads) and ∼35 Gb (∼139 million paired-end reads) of data for *M. sativa* and *M. officinalis*, respectively (see Table S4 for sequencing statistics). The use of a poly-A enrichment step meant that only plant transcripts were represented in the sequencing output. De novo assembly of the *M. sativa* sequencing data resulted in 253,871 contigs, while 192,165 de novo assembled contigs were produced for *M. officinalis*. Contigs expected to represent splice variants of a single gene were merged into so-called “supertranscripts” using the SuperTranscripts program, resulting in compressed assemblies of 79,978 and 64,593 contigs for *M. sativa* and *M. officinalis*, respectively (Table 1). Transcriptomes were annotated as described in the Materials and Methods, resulting in putative annotations for 33,431 *M. sativa* contigs and 28,278 *M. officinalis* contigs (Datasets S1 and S2). Of these, ∼52% (*M. sativa*) and ∼58% (*M. officinalis*) are high confidence annotations as they were transferred from the *M. truncatula* whole genome annotation following identification of putative orthologs using a BLAST bidirectional best hit approach (Table S5). Considering that previous studies have predicted the presence of ∼23,000 long noncoding RNAs (IncRNAs) in *M. truncatula* (Wang et al., 2015) and ∼47,000 IncRNAs in the legume *Pisum sativum* (pea) (Kerr et al., 2017), we hypothesize that the majority of the unannotated *M. sativa* and *M. officinalis* transcripts reflect IncRNAs.

All of the examined assembly summary statistics (mean and median contig length, contig N50) were improved in the compressed assemblies compared with the original de novo assemblies, indicating that the compressed assemblies are of higher structural quality (Table 1). The *M. sativa* transcriptome summary statistics, such as N50 and and transcript length, are consistent with those reported for other *M. sativa* de novo transcriptome assemblies, although the number of transcripts varies likely due to each study examining different tissues (Arshad et al., 2018; Zhang et al., 2015). In addition, the assemblies appear to be robust; greater than 90% of the filtered reads used for transcriptome assembly could be mapped to the corresponding assemblies by STAR (Table 1). Moreover, >92% and >83% of the Viridiplantae and Fabales BUSCO marker genes, respectively, were identified as complete and single-copy in the *M. sativa* and *M. officinalis* compressed assemblies (Figure 1). The structural quality (e.g., high average and median contig length and N50) and BUSCO benchmark scores described here are in line with those reported for other plant de novo transcriptome assemblies (Al-Qurainy et al., 2019; Malovicko et al., 2020; Weisberg et al., 2017). Taken together, these results indicate that our *M. sativa* cv. Algonquin and *M. officinalis* reference nodule transcriptomes are reliable and of high quality.

3.2 | Comparative transcriptome analysis between *M. sativa* and *M. officinalis*

As an initial examination of the *M. sativa* and *M. officinalis* transcriptomes, the annotated functions of the proteins predicted to be encoded by the supertranscripts were summarized using the Generic GO term subset (Figure 2 and Datasets S3 and S4). Approximately 18,363 (26.1%) of the *M. sativa* supertranscripts and 16,674 (28.6%) of the *M. officinalis* supertranscripts were annotated with GO terms.

| Table 1 | Summary statistics from the de novo trinity and compressed (SuperTranscripts) nodule transcriptome assemblies |
|---------|-------------------------------------------------|-------------------------------------------------|------------------|------------------|
|         | **Medicago sativa**                             | **Mellilotus officinalis**                      |
|         | **Trinity assembly**                            | **SuperTranscripts**                           | **Trinity assembly** | **SuperTranscripts** |
| Total number of contigs | 253,871                                      | 79,978                                       | 192,165                                   | 64,593                                   |
| Total number of base pairs (bp) | 239,421,306                                 | 96,211,060                                  | 191,200,637                                | 81,371,534                                |
| Average contig length (bp) | 943                                           | 1,203                                        | 995                                         | 1,260                                      |
| Median contig length (bp) | 618                                           | 734                                          | 646                                         | 780                                         |
| Contig N50 (bp) | 1,466                                       | 1,912                                        | 1,584                                       | 1,968                                       |
| Minimum contig length (bp) | 176                                         | 193                                         | 183                                         | 198                                         |
| Maximum contig length (bp) | 12,655                                      | 29,784                                      | 14,584                                      | 23,941                                      |
| Overall alignment rate (%) | 98.99                                       | 92.14                                       | 99.15                                       | 93.76                                       |
No significant difference in the GO term profiles of the two species was observed, with the five most frequently annotated biological process GO terms being GO:0008150 (biological process), GO:0006950 (response to stress), GO:0006464 (cellular protein modification process), GO:003464 (cellular nitrogen compound metabolic process), and GO:0048856 (anatomical structure development). At this broad scale, the GO term data suggest that there is substantial similarity in the nodule transcriptomes of *M. sativa* and *M. officinalis*.

We next examined the predicted functions of the proteins encoded by the 50 most abundant transcripts in both species (Tables 2 and 3). Not surprisingly, these transcripts were enriched in those predicted to encode nodulins and leghaemoglobin-like proteins. Nodulins refer to diverse proteins expressed specifically in nodule tissue, which play various structural or metabolic roles during symbiotic nitrogen fixation. Among the nodulins are the leghemoglobin proteins that account for up to 40% of the total soluble protein in legume nodules (Nash & Schulman, 1976). Leghemoglobins play an important role in maintaining the low free-oxygen concentration required to protect the oxygen-sensitive nitrogenase enzyme (Ott et al., 2005).

To facilitate further comparison of the *M. sativa* and *M. officinalis* transcriptomes, the proteins predicted to be encoded by the supertranscripts of both species were arranged into orthologous groups using OrthoFinder. A total of 20,237 orthologous groups, accounting for 26,304 *M. sativa* and 24,895 *M. officinalis* supertranscripts, were identified. Interestingly, the abundance of the conserved supertranscripts was significantly higher, on average, than that of the species-specific transcripts (p-value < 2.2e-16; Figure 3). In both plant species, the majority of the most abundant, species-specific annotated transcripts were also nodulins, globin family proteins that are likely species-specific leghaemoglobin isoforms, and some housekeeping genes such as ribonuclease and ribosomal proteins. It is noteworthy that the most abundant *M. sativa*-specific supertranscript is predicted to encode albumin I. Similarly, *M. officinalis* also has a highly expressed albumin I supertranscript. The albumin I peptide family is known to be highly expressed in legume seeds and play roles in seed protection (Rahioui et al., 2014). Expression of albumin I genes has also been observed in *M. truncatula* root nodules, with expression specific to uninfected cells in the nitrogen fixation zone (Limpens et al., 2013).
| Gene ID          | TPM   | Functional prediction                                      |
|-----------------|-------|------------------------------------------------------------|
| Cluster-2.43518 | 30,662| Hypothetical protein (hypothetical leghaemoglobin)         |
| Cluster-2.26078 | 17,769| Putative albumin I                                         |
| Cluster-2.23447 | 12,515| Hypothetical protein (hypothetical leghaemoglobin)         |
| Cluster-2.23176 | 8,065 | Nodulin-25                                                 |
| Cluster-2.22272 | 7,992 | Putative late nodulin                                      |
| Cluster-2.23033 | 6,778 | None                                                       |
| Cluster-2.21873 | 5,919 | Hypothetical protein                                       |
| Cluster-2.22935 | 5,253 | Belongs to the globin family                               |
| Cluster-2.33070 | 5,232 | Putative ribonuclease H-like domain-containing protein     |
| Cluster-2.24197 | 4,245 | Component of the replication protein A complex (RPA)       |
| Cluster-2.22983 | 4,196 | Belongs to the globin family                               |
| Cluster-2.24546 | 3,886 | Predicted NCR peptide (crp1450_Cluster-2.24546_0M_1)     |
| Cluster-2.19511 | 3,344 | None                                                       |
| Cluster-2.26207 | 3,296 | Extensin-like_protein_repeat                               |
| Cluster-2.49512 | 3,173 | Putative blue (type 1) copper binding protein              |
| Cluster-2.21810 | 3,050 | Predicted NCR peptide (crp1160_Cluster-2.21810_0M_1)     |
| Cluster-2.22936 | 3,014 | Belongs to the globin family                               |
| Cluster-2.29430 | 2,789 | Putative late nodulin                                      |
| Cluster-2.22881 | 2,788 | Predicted NCR peptide (crp1430_Cluster-2.22881_0M_1)     |
| Cluster-2.22836 | 2,509 | None                                                       |
| Cluster-2.23729 | 2,271 | Hypothetical protein                                       |
| Cluster-2.22458 | 2,245 | Belongs to the globin family                               |
| Cluster-2.24829 | 2,227 | Hypothetical protein                                       |
| Cluster-2.25168 | 2,209 | None                                                       |
| Cluster-2.24278 | 2,093 | Nodule-specific_GRP_repeat                                 |
| Cluster-2.23928 | 2,038 | Late_nodulin_protein                                      |
| Cluster-2.22042 | 2,029 | None                                                       |
| Cluster-2.23245 | 2,019 | Putative translationally controlled tumor protein          |
| Cluster-2.25794 | 2,018 | Putative protein-synthesizing GTPase                      |
| Cluster-2.31376 | 1,957 | Predicted NCR peptide (crp1190_Cluster-2.31376_0M_1)     |
| Cluster-2.21809 | 1,939 | Predicted NCR peptide (crp1160_Cluster-2.21809_0M_1)     |
| Cluster-2.28083 | 1,854 | Predicted NCR peptide (crp1210_Cluster-2.28083_0M_1)     |
| Cluster-2.23524 | 1,844 | Early nodulin-16                                          |
| Cluster-2.34649 | 1,839 | Putative late nodulin                                      |
| Cluster-2.16993 | 1,808 | Hypothetical protein                                       |
| Cluster-2.21813 | 1,731 | None                                                       |

(Continues)
TABLE 2 (Continued)

| Gene ID   | TPM    | Functional prediction                                      |
|-----------|--------|-----------------------------------------------------------|
| Cluster-2.22457 | 1,705  | Belongs to the globin family                              |
| Cluster-2.28283  | 1,674  | None                                                      |
| Cluster-2.26205  | 1,621  | Predicted NCR peptide (crp1240_Cluster-2.26205_0M_1)     |
| Cluster-2.23310  | 1,607  | Asparagine synthetase                                     |
| Cluster-2.21870  | 1,596  | Nodule-specific_GRP_repeat                                 |
| Cluster-2.19604  | 1,583  | Predicted NCR peptide (crp1420_Cluster-2.19604_0M_1)     |
| Cluster-2.18485  | 1,509  | Hypothetical protein                                      |
| Cluster-2.26876  | 1,458  | Predicted NCR peptide (crp1410_Cluster-2.26876_0M_1)     |
| Cluster-2.21536  | 1,441  | Ubiquitin_family                                          |
| Cluster-2.22785  | 1,408  | None                                                      |
| Cluster-2.23374  | 1,385  | Predicted NCR peptide (crp1420_Cluster-2.23374_0M_1)     |
| Cluster-2.28787  | 1,311  | Putative late nodulin                                     |
| Cluster-2.22402  | 1,292  | Predicted NCR peptide (crp1520_Cluster-2.22402_0M_1)     |
| Cluster-2.30081  | 1,289  | Late_nodulin_protein                                      |

TABLE 3 The 50 most highly abundant transcripts in the *Melilotus officinalis* nodule transcriptome, with the average expression level in transcripts per million (TPM) and the functional annotation

| Gene ID       | TPM     | Functional prediction                                      |
|---------------|---------|-----------------------------------------------------------|
| Cluster-3554.18801 | 38,953  | Belongs to the globin family                              |
| Cluster-3554.18778  | 14,091  | Belongs to the globin family                              |
| Cluster-3554.16063  | 10,412  | Late_nodulin_protein                                     |
| Cluster-3554.15387  | 7,885   | Putative late nodulin                                     |
| Cluster-3554.16088  | 6,146   | Putative late nodulin                                     |
| Cluster-3554.18892  | 6,104   | None                                                      |
| Cluster-3554.15771  | 5,813   | Late_nodulin_protein                                     |
| Cluster-3554.18802  | 5,362   | Belongs to the globin family                              |
| Cluster-3554.18596  | 5,347   | Predicted NCR peptide (crp1430_Cluster-3554.18596_0M_1) |
| Cluster-3554.15456  | 3,912   | Putative translationally controlled tumor protein         |
| Cluster-3554.18808  | 3,864   | Belongs to the globin family                              |
| Cluster-3554.33215  | 3,302   | Hypothetical protein                                      |
| Cluster-3554.18775  | 3,097   | Putative late nodulin                                     |
| Cluster-3554.23000  | 2,990   | Two predicted NCR peptide (crp1180_Cluster-3554.23000_0M_1 and crp1180_Cluster-3554.23000_0M_2) |
| Cluster-3554.36681  | 2,877   | Predicted NCR peptide (crp1500_Cluster-3554.36681_0M_1)  |
| Cluster-3554.21555  | 2,868   | Predicted NCR peptide (crp1430_Cluster-3554.21555_0M_1)  |
| Cluster-3554.16074  | 2,809   | Putative BURP domain-containing protein                    |

(Continues)
These cells are thought to play essential roles in metabolite transport during symbiosis, and albumin I may have a role in protecting some of the nodule cells from rhizobium infection (Limpens et al., 2013). A phylogenetic analysis of *M. truncatula* nodulins and albumin I peptides indicated that the *M. truncula* albumin I clustered with a subset of nodulins, reflecting a close evolutionary relationship between these proteins (Karaki et al., 2016).

We next compared the abundances of supertranscripts conserved in both *M. sativa* and *M. officinalis*, limiting the analysis to the 15,287 one-to-one orthologs detected by OrthoFinder. Despite significant variation in the abundance of orthologous transcripts between *M. sativa* and *M. officinalis*—which may reflect limitations of interspecies transcriptome analysis—a clear correlation in the abundance of orthologous transcripts was detected (residual

| Gene ID               | TPM   | Functional prediction                                                                 |
|----------------------|-------|---------------------------------------------------------------------------------------|
| Cluster-3554.29297   | 2.784 | Hypothetical protein                                                                   |
| Cluster-3554.18196   | 2.723 | None                                                                                   |
| Cluster-3554.18256   | 2.571 | Predicted NCR peptide (crp1440_Cluster-3554.18256_OM_1)                                |
| Cluster-3554.27063   | 2.522 | None                                                                                   |
| Cluster-3554.22577   | 2.449 | Putative blue (type 1) copper binding protein                                          |
| Cluster-3554.15361   | 2.409 | eEF1A                                                                                  |
| Cluster-3554.21311   | 2.384 | Belongs to the globin family                                                           |
| Cluster-3554.18838   | 2.340 | Late_nodulin_protein                                                                  |
| Cluster-3554.18700   | 2.296 | Late_nodulin_protein                                                                  |
| Cluster-3554.11713   | 2.261 | None                                                                                    |
| Cluster-3554.18706   | 2.259 | None                                                                                    |
| Cluster-3554.23445   | 2.214 | None                                                                                    |
| Cluster-3554.17366   | 2.141 | Predicted NCR peptide (crp1430_Cluster-3554.17366_OM_1)                                |
| Cluster-3554.23502   | 2.073 | Hypothetical protein                                                                   |
| Cluster-3554.25153   | 2.009 | Metallothionein-like protein 2                                                         |
| Cluster-3554.21451   | 1.953 | Belongs to the globin family                                                           |
| Cluster-3554.28600   | 1.931 | Hypothetical protein                                                                   |
| Cluster-3554.22971   | 1.849 | Hypothetical protein                                                                   |
| Cluster-3554.18877   | 1.812 | Belongs to the glyceraldehyde-3-phosphate dehydrogenase family                         |
| Cluster-3554.13172   | 1.732 | Predicted NCR peptide (crp1440_Cluster-3554.13172_OM_1)                                |
| Cluster-3554.18195   | 1.704 | Zinc_knuckle                                                                           |
| Cluster-3554.30751   | 1.653 | Putative late nodulin                                                                  |
| Cluster-3554.21774   | 1.635 | Late_nodulin_protein                                                                  |
| Cluster-3554.13784   | 1.629 | Predicted NCR peptide (crp1420_Cluster-3554.13784_OM_1)                                |
| Cluster-3554.24033   | 1.611 | Prolyl isomerase (PPIase)                                                              |
| Cluster-3554.30294   | 1.556 | Metallothionein-like protein                                                           |
| Cluster-3554.24764   | 1.552 | None                                                                                    |
| Cluster-3554.18368   | 1.552 | Nucleoside diphosphate kinase 1                                                       |
| Cluster-3554.25344   | 1.537 | Metallothionein-like protein 1                                                         |
| Cluster-3554.23646   | 1.514 | None                                                                                    |
| Cluster-3554.9874    | 1.512 | Belongs to the universal ribosomal protein ul13 family                                  |
| Cluster-3554.18476   | 1.503 | Late_nodulin_protein                                                                  |
| Cluster-3554.31722   | 1.476 | Putative late nodulin                                                                  |
standard error = .517; Figure 4). Considering the limitations of interspecies differential expression analyses, we restricted our investigation to supertranscripts with absolute log2 fold changes > 5 and a p-value < .05. Using these thresholds, we identified 290 differentially abundant transcripts, 86 of which were more abundant in M. sativa, and 204 of which were more abundant in M. officinalis. It should be noted, however, that only 159 of the differentially abundant transcripts were annotated with the same or similar function in both species, and we focus on these 159 transcripts in the following discussion.

Many of the differentially abundant conserved supertranscripts have annotated functions that suggest that the encoded proteins may impact symbiotic nitrogen fixation. These include 21 supertranscripts annotated as encoding nodulins, which include 16 that are more abundant in M. officinalis and five that are more abundant in M. sativa. In addition, 32 supertranscripts encoding proteins predicted to be associated with transcription and translation activity were differentially abundant, with 24 more abundant in M. sativa and eight more abundant in M. officinalis. We also observed that several supertranscripts encoding proteins predicted to be involved in cell wall synthesis or modification were differentially abundant, with six more highly abundant in M. sativa and one more highly abundant in M. officinalis. Other differentially abundant transcripts included those predicted to encode proteins involved in transport (17 transcripts), fatty acid biosynthesis (three transcripts), flavonoid biosynthesis (three transcripts), and aromatic compound biosynthesis (one transcript). Given that this analysis compares two plant species with differing growth rates (Table S2), we cannot rule out that some of these transcriptomic differences may also reflect variances in nodule maturity and/or host metabolic activity at the time of harvest.

### 3.3 NCR peptide diversity and expression profile

We previously observed that replacing the bacA allele of S. melloti Rm2011 with the bacA alleles of the rhizobia S. fredii NGR234 or R. leguminosarum bv. viciae 3841 resulted in an inability to fix nitrogen with M. sativa, while the ability to fix nitrogen with M. alba and M. officinalis remained (diCenzo et al., 2017 and Table S1). We hypothesized that this was due to differences in the NCR peptide profiles of these species (diCenzo et al., 2017). To test this hypothesis, supertranscripts encoding NCR peptides were identified in the M. sativa and M. officinalis transcriptome assemblies using the SPADA pipeline (Zhou et al., 2013). A total of 412 and 308 supertranscripts encoding NCR peptides were identified in the M. sativa and M. officinalis transcriptomes, respectively, accounting for ~0.5% of all supertranscripts in both assemblies (Datasets S5 and S6). The lower count of NCR transcripts in M. officinalis was offset by a higher median transcript abundance (58.6 transcripts per million [TPM] vs 99.1 TPM; p < .001; Figure 5a), resulting in NCR transcripts accounting for roughly 9% of the total nodule transcriptome in both species.

The number of NCR peptides predicted from the M. sativa transcriptome is similar to the 469 predicted previously based on a M. sativa genome sequence (Montiel et al., 2017). However, only 70% of the NCR peptides showed >70% identity, which could be due to the use of different cultivars and/or the method of NCR peptide prediction. Additionally, 63% of the M. sativa NCR peptides, but only...
23% of M. officinalis NCR peptides, showed >70% sequence identity to NCR peptides from M. truncatula (Dataset S7). Notably, using a 70% identity threshold, only 40 NCR peptides were found in all three species, while two NCR peptides known to be essential for symbiosis in M. truncatula (NCR211 and NCR169) were not identified in M. sativa nor M. officinalis (Horváth et al., 2015; Kim et al., 2015). Overall, these results are consistent with rapid evolution of the NCR gene family in legumes and with the two Medicago species having NCR peptide profiles more similar to each other than to M. officinalis.

**FIGURE 5** Nodule-specific cysteine-rich (NCR) peptide profiles of Medicago sativa and Mellilotus officinalis. NCR peptides were predicted from the M. sativa (orange) and M. officinalis (blue) transcriptome assemblies, and the properties of the NCR peptides are shown in these graphs. (a) Box plots showing the distribution of the abundance (in transcripts per million, TPM) of NCR transcripts, based on triplicate samples. The difference in the distributions for the two species was statistically significant (p-value < .001; pairwise Wilcoxon test). (b) Box plots showing the distribution of the amino acid lengths of mature NCR peptides. No statistically significant difference in the distributions for the two species was detected. (c, d) Histograms showing the distributions of the isoelectric points (pI) for the mature NCR peptides. Histograms are based either on the number of NCR peptides with a given pI value (C) or the total abundance of the transcripts encoding NCR peptides with a given pI value (D). (e, f) Histograms showing distributions of pI for 4-cysteines (e) and 6-cysteines (f) mature NCR peptides based on total abundance of the transcripts encoding NCR peptides with a given pI value.
In both *M. sativa* and *M. officinalis*, NCR peptides had median lengths of 38 residues, with approximately half of the NCR peptides containing between 30 and 40 residues (Figure 5b). Additionally, there was a roughly even number of four- and six-cysteine NCR peptides expressed in both plant species, with the four-cysteine class of NCR peptides accounting for 51%–55% of the NCR transcripts both in terms of number of NCR peptides and expression of NCR transcripts as measured by TPM. The NCR peptides from both hosts also showed broadly similar distributions of pl values between approximately 3 to 11, with one peak around a pl of 4 and another around pl 8 (Figure 5c,d). The pl pattern of the NCR peptides we observed is reminiscent of that reported for other legume species that induce an elongated branched morphology in their microsymbiont, including *M. sativa* and *M. truncatula* (Montiel et al., 2017). Overall, at a global level, the property profiles of NCR peptides for *M. sativa* and *M. officinalis* were very similar, suggesting that the impact of different *bacA* alleles on symbiotic compatibility of *S. melloti* with *M. sativa* is unlikely a consequence of global differences in the NCR peptide profiles of these plants and is more likely due to specific NCR peptides. Identifying which NCR peptides functionally correlate with symbiotic compatibility should be the focus of future studies.

Despite the general similarity in the NCR peptide profiles of *M. sativa* and *M. officinalis*, a key difference emerges when examining the abundance of NCR peptides with extreme pl values; transcripts encoding highly cationic NCR peptides were more abundant in *M. sativa*, while transcripts encoding highly anionic NCR peptides were more abundant in *M. officinalis* (Figure 5d). Previous work has shown that, in general, only cationic NCR peptides with a pl > 9.0 have antimicrobial activity (Lima et al., 2020), with anticanidal activity primarily limited to NCR peptides with a pl > 9.5 (Ördögh et al., 2014). Here, we observed that transcripts encoding highly cationic NCR peptides (pl > 9.0) were ~2.4-fold more abundant in *M. sativa* than *M. officinalis* (Figure 5d). Similarly, transcripts encoding cationic NCR peptides with pl values > 9.5 were ~1.9-fold more abundant in *M. sativa* than *M. officinalis*. Notably, previous work indicated that 4.0% of *M. truncatula* NCR transcripts encode NCR peptides with pl values > 9.5, compared with only 1.8% in the *R. leguminosarum* bv. *viciae* symbiont *P. sativum* (Alves-Carvalho et al., 2015; Montiel et al., 2017; Roux et al., 2014); this compares to 4.7% and 2.3% for *M. sativa* and *M. officinalis*, respectively (Figure 5c). Strikingly, when subdividing the NCR peptides with pl values > 9.5 into those with four or six cysteine residues, we observed that those with six cysteines were ~27-fold more abundant in *M. sativa* than *M. officinalis* (Figure 4e,f). Considering these results, we hypothesize that the ability of the *R. leguminosarum bacA* allele to support symbiosis with *M. officinalis* and *P. sativum*, but not *M. sativa*, is a consequence of the elevated abundance of highly cationic (pl > 9.5) NCR peptides in *Medicago* nodules. Consistent with this, two *M. sativa* NCR peptides were identified that had high sequence identity (83% and 73%) with *M. truncatula* NCR035, which is a cationic NCR peptide with antimicrobial activity against *S. melloti* (Haag et al., 2011), whereas similar sequences were not found in *M. officinalis*. It may be that the BacA proteins of *S. fredii* and *R. leguminosarum* are less capable of transporting highly cationic NCR peptides, and consequently, strains with these BacA proteins may be more sensitive to the antimicrobial activities of these NCR peptides.

## 4. Conclusion

We report high-quality nodule transcriptome assemblies for *M. sativa* cv. Algonquin and *M. officinalis* that we expect will serve as valuable resources for the legume research community. In particular, we expect that the availability of a nodule transcriptome for *M. officinalis* will help establish this plant as a secondary model system for studies of the symbiotic properties of *S. melloti*.

We were particularly interested in using these transcriptomes to compare the properties of the NCR peptides encoded by both species. Despite predicting 33% more NCR peptides in *M. sativa* than *M. officinalis*, NCR transcripts accounted for roughly 9% of the transcriptome (based on TPM values) in both species. In general, the characteristics of the NCR peptides of *M. sativa* and *M. officinalis* were highly similar. However, transcripts encoding cationic NCR peptides with a pl > 9.5 were ~2-fold more abundant in *M. sativa* than in *M. officinalis* and 27-fold more abundant when considering only six-cysteine NCR peptides. These results are consistent with previous observations that transcripts encoding cationic NCR peptides with a pl > 9.5 account for ~2-fold more NCR transcripts in *M. truncatula* compared with *P. sativum*. Cationic, but not neutral or anionic, NCR peptides display antimicrobial activity through disrupting the integrity of microbial membranes (Mikuláš et al., 2016). It has been hypothesized that BacA provides protection against these NCR peptides by importing them into the cytoplasm and thus away from the membrane (Arnold et al., 2017; Nicoud et al., 2021). Considering that the BacA proteins of *S. fredii* and *R. leguminosarum* share less than 60% amino acid identity with the BacA protein of *S. melloti*, it is reasonable to speculate that they have different substrate specificity and may be less capable of transporting cationic NCR peptides (diCenzo et al., 2017). If true, this could explain why the bacA alleles of *S. fredii* and *R. leguminosarum* can support symbiotic nitrogen fixation with *M. officinalis* but not *M. sativa*; the increased production of cationic NCR peptides in *M. sativa*, coupled with lower rates of import into the *S. melloti* cytoplasm, could result in an accumulation of these peptides in the periplasm, resulting in a loss of viability and lack of nitrogen fixation (diCenzo et al., 2017). In future work, it will be interesting to test whether *S. melloti* strains with different bacA alleles display differing sensitivities to these highly cationic NCR peptides or differences in their abilities to transport these peptides.

## Acknowledgments

We thank Karen Ho and Neda Moradin from The Centre for Applied Genomics (Toronto, Canada) for helpful advice in planning the RNA-seq library preparation strategy. This research was enabled, in part, through computational resources provided by Compute Ontario (computeonto.ca) and Compute Canada (computecanada.ca). Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grants to WAS and GCD.
CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT
All custom scripts to perform the analyses described in this study are available through GitHub (github.com/hyhy8181994/Nodule_transcriptome_script). Raw Illumina data are available through the Short Read Archive (SRR15724671, SRR15724670, SRR15724669, SRR15724668, SRR15724667, and SRR15724666) hosted by the National Center for Biotechnology Information (NCBI). The assembled transcriptomes are available through the Transcriptome Shotgun Assembly Sequence Database (GJLW00000000 and GJLK00000000) hosted by the NCBI.

ORCID
Rui Huang https://orcid.org/0000-0002-2144-030X
George C. diCenzo https://orcid.org/0000-0003-3889-6570

REFERENCES
Al-Qurainy, F., Alshami, A., Gaafar, A.-R., Khan, S., Nadeem, M., Alameri, A. A., Tarroum, M., & Ashraf, M. (2019). Comprehensive stress-based de novo transcriptome assembly and annotation of guar (Cyamopsis tetragonoloba (L.) Taub): An important industrial and forage crop. International Journal of Genomics, 2019, 7295859. https://doi.org/10.1155/2019/7295859
Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A. L., Jacquin, F., Klein, A., Martin, C., Boucherot, K., Kreplak, J., Da Silva, C., Moreau, S., Gamas, P., Wincker, P., Gouzy, J., & Burstin, J. (2015). Full-length de novo assembly of RNA-seq data in pea (Pisum sativum L.) provides a gene expression atlas and gives insights into root nodulation in this species. The Plant Journal, 84, 1–19. https://doi.org/10.1111/tjp.12677
Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F., & Wingett, S. (2015). FastQC. A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics. Babraham Institute.
Arnold, M. F. F., Shabab, M., Penterman, J., Boehme, K. L., Griffitts, J. S., & Walker, G. C. (2017). Genome-wide sensitivity analysis of the microsymbiont Sinorhizobium meliloti to symbiotically important, defensin-like host peptides. mBio, 8(4), e01060–e01017. https://doi.org/10.1128/mBio.01060-17
Arshad, M., Gruber, M. Y., & Hannoufa, A. (2018). Transcriptome analysis of microRNA156 overexpression alfalfa roots under drought stress. Scientific Reports, 8, 1–13. https://doi.org/10.1038/s41598-018-27088-8
Audain, E., Ramos, Y., Hermjakob, H., Flower, D. R., & Perez-Riverol, Y. (2016). Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. Bioinformatics, 32(6), 821–827. https://doi.org/10.1093/bioinformatics/btv674
Barrière, Q., Guerfachi, I., Gully, D., Lamouche, F., Pierre, O., Fardoux, J., Chaintreuil, C., Alunni, B., Timchenko, T., Giraud, E., & Mergaert, P. (2017). Integrated roles of BcIA and DD-carboxypeptidase 1 in Bradyrhizobium differentiation within NCR-producing and NCR-lacking root nodules. Scientific Reports, 7, 1–13. https://doi.org/10.1038/s41598-017-08830-0
Binney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genowise. Genome Research, 14(5), 988–995. https://doi.org/10.1101/gr.1865504
Blanco, E., Parra, G., & Guigó, R. (2007). Using geneid to identify genes. Current Protocols in Bioinformatics, 18, 3–4. https://doi.org/10.1002/0471250953.bi0403s18
Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12, 59–60. https://doi.org/10.1038/nmeth.3176
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 1–9. https://doi.org/10.1186/1471-2105-10-421
Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., Zhang, X., Zhang, R., Zhang, Y., Li, Y., Wang, K., He, H., Wang, Z., Fan, G., Yang, H., ... Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. Nature Communications, 11, 2494. https://doi.org/10.1038/s41467-020-1638-x
Davidson, N. M., Hawkins, A. D. K., & Oshlack, A. (2017). SuperTranscripts: A data driven reference for analysis and visualisation of transcriptomes. Genome Biology, 18, 1–10. https://doi.org/10.1186/s13059-014-0410-6
DHaese, W., & Holsters, M. (2002). Nod factor structures, responses, and perception during initiation of nodule development. Glycobiology, 12(6), 79R–105R. https://doi.org/10.1093/glycob/12.6.79R
diCenzo, G. C., Zamani, M., Cowie, A., & Finan, T. M. (2015). Proline auxotrophy in Sinorhizobium meliloti results in a plant-specific symbiotic phenotype. Microbiology, 161(12), 2341–2351. https://doi.org/10.1099/mic.0.000182
diCenzo, G. C., Zamani, M., Ludwig, H. N., & Finan, T. M. (2017). Heterologous complementation reveals a specialized activity for BacA in the Medicago–Sinorhizobium meliloti symbiosis. Molecular Plant-Microbe Interactions, 30(4), 312–324. https://doi.org/10.1094/MPMI-02-17-0030-R
Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635
Eddy, S. R. (2011). Accelerated profile HMM searches. PLoS Computational Biology, 7(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195
Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology, 20, 1–14. https://doi.org/10.1186/s13059-019-1832-y
Ferguson, G. P., Datta, A., Baumgartner, J., Roop, R. M., Carlson, R. W., & Walker, G. C. (2004). Similarity to peroximalomembrane protein family reveals that Sinorhizobium and Brucella BacA affect lipid-a fatty acids. Proceedings of the National Academy of Sciences of the United States of America, 101(14), 5012–5017. https://doi.org/10.1073/pnas.0307137101
Ferguson, G. P., Roop, R. M., & Walker, G. C. (2002). Deficiency of a Sinorhizobium meliloti bacA mutant in alfalfa symbiosis correlates with alteration of the cell envelope. Journal of Bacteriology, 184, 5625–5632. https://doi.org/10.1128/JB.184.20.5625-5632.2002
Finan, T. M., Hirsch, A. M., Leigh, J. A., Johansen, E., Kuldau, G. A., Deegan, S., Walker, G. C., & Signer, E. R. (1985). Symbiotic mutants of Rhizobium meliloti that uncouple plant from bacterial differentiation. Cell, 40, 869–877. https://doi.org/10.1016/0092-8674(85)90346-0
Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiology
Ichige, A., & Walker, G. C. (1997). Genetic analysis of the Rhizobium meliloti bacA gene: Functional interchangeability with the Escherichia coli sbmA gene and phenotypes of mutants. Journal of Bacteriology, 179, 209–216. https://doi.org/10.1128/ JB.179.1.209-216.1997

Jensen, H. L. (1942). Nitrogen fixation in leguminous plants. I. General characters of root-nodule bacteria isolated from species of Medicago and Trifolium in Australia. Proceedings of the Linnean Society of New South Wales, 66, 98–108.

Karaki, L., Silva, P. D., Rizk, F., Chouabe, C., Chantret, N., Eyraud, V., Gressent, F., Sivignon, C., Rahiou, I., Kahn, D., Brochier-Armanet, C., Rabhé, Y., & Royer, C. (2016). Genome-wide analysis identifies gain and loss/change of function within the small multigenic insecticidal albumin 1 family of Medicago truncatula. BMC Plant Biology, 16, 63. https://doi.org/10.1186/s12870-016-0745-0

Kerr, S. C., Gaití, F., Beveridge, C. A., & Tanurdzic, M. (2017). De novo transcriptome assembly reveals high transcriptional complexity in Pisum sativum axillary buds and shows rapid changes in expression of diurnally regulated genes. BMC Genomics, 18, 221. https://doi.org/10.1186/s12864-017-3577-x

Kim, M., Chen, Y., Xi, J., Waters, C., Chen, R., & Wang, D. (2015). An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. Proceedings of the National Academy of Sciences of the United States of America, 112(49), 15238–15243. https://doi.org/10.1073/pnas.1500123112

Kossack, R. M., Bookland, R., Barkei, J., Paaren, H. E., & Appelbaum, E. R. (1987). Induction of Bradyrhizobium japonicum common nod genes by isoflavones isolated from Glycine max. Proceedings of the National Academy of Sciences of the United States of America, 84(21), 7428–7432. https://doi.org/10.1073/pnas.84.21.7428

Lamouche, F., Bonadé-Bottino, N., Mergaert, P., & Alunni, B. (2019). Sym-biotic efficiency of spherical and elongated bacteroids in the Aeschynomene-Bradyrhizobium symbiosis. Frontiers in Plant Science, 10, 377. https://doi.org/10.3389/fpls.2019.00377

Leigh, J. A., Signer, E. R., & Walker, G. C. (1985). Exopolysaccharide-deficient mutants of Rhizobium meliloti that form ineffective nodules. Proceedings of the National Academy of Sciences of the United States of America, 82(18), 6231 LP–6235. https://doi.org/10.1073/pnas.82.18.6231

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Lima, R. M., Kylarová, S., Mergaert, P., & Kondorosi, É. (2020). Unexplored arsenals of legume peptides with potential for their applications in medicine and agriculture. Frontiers in Microbiology, 11, 1307. https://doi.org/10.3389/fmicb.2020.01307

Limpens, E., Moling, S., Hooiveld, G., Pereira, P. B., Bisseling, T., Becker, J. D., & Küster, H. (2013). Cell-and tissue-specific transcriptome analyses of Medicago truncatula root nodules. PLoS ONE, 8(5), e64377. https://doi.org/10.1371/journal.pone.0064377

Love, M., Anders, S., & Huber, W. (2014). Differential analysis of count data—the DESeq2 package. Genome Biology, 15, 10–1186.

Lukashin, A. V., & Borodovsky, M. (1998). GeneMark. Hmm: New solutions for gene finding. Nucleic Acids Research, 26(4), 1107–1115. https://doi.org/10.1093/nar/26.4.1107

Malovichko, Y. V., Shtark, O. Y., Vasileva, E. N., Nizhnikov, A. A., & Antonets, K. S. (2020). Transcriptomic insights into mechanisms of early seed maturation in the garden pea (Pisum sativum L.). Cells, 9(3), 779. https://doi.org/10.3390/cells9030779

Marlow, V. L., Haag, A. F., Kobayashi, H., Fletcher, V., Soccoli, M., Walker, G. C., & Ferguson, G. P. (2009). Essential role for the BacA protein in the uptake of a truncated eukaryotic peptide in Sinorhizobium meliloti. Journal of Bacteriology, 191(5), 1519–1527. https://doi.org/10.1128/JB.01661-08

Maroti, G., Downie, J. A., & Kondorosi, É. (2015). Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules. Current Opinion in Plant Biology, 26, 57–63. https://doi.org/10.1016/j.pbi.2015.05.031

Maroti, G., Kereszt, A., Kondorosi, E., & Mergaert, P. (2011). Natural roles of antimicrobial peptides in microbes, plants and animals. Research in Microbiology, 162(4), 363–374. https://doi.org/10.1016/j.resmic.2011.02.005
Ördögh, L., Vörös, A., Nagy, I., Kondorosi, E., Biondi, E. G., Timchenko, T., Alunni, B., & Mergaert, P. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proceedings of the National Academy of Sciences of the United States of America, 103(13), 5230–5235. https://doi.org/10.1073/pnas.0600912103

Mikulás, K. R., Nagy, K., Bogos, B., Szegletes, Z., Kovács, E., Farkas, A., Váró, G., Kondorosi, É., & Kereszt, A. (2016). Antimicrobial nodules-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. Annals of Clinical Microbiology and Antimicrobials, 15, 43. https://doi.org/10.1186/s12941-016-0159-8

Mistry, J., Shuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardsdon, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Báltí, B., Mergaert, P., Kereszt, A., & Kondorosi, É. (2017). Morphotype of bacteria in different legumes correlates with the number and type of symbiotic NCR peptides. Proceedings of the National Academy of Sciences of the United States of America, 114(19), 5041–5046. https://doi.org/10.1073/pnas.1704217114

Nash, D. T., & Schulman, H. M. (1976). Leghemoglobin and nitrogenase activity during soybean root nodule development. Canadian Journal of Botany, 54(24), 2790–2797. https://doi.org/10.1139/b76-279

Nicoud, Q., Barrière, Q., Busset, N., Dendene, S., Travin, D., Bourge, M., Le Bars, R., Boulouge, C., Lecroë, M., Jenei, S., Kereszt, A., Kondorosi, E., Blondet, E. G., Timchenko, T., Alunni, B., & Mergaert, P. (2021). Sinorhizobium meliloti functions required for resistance to antimicrobial NCR peptides and bacteroid differentiation. mBio, 12, e00895-21. https://doi.org/10.1128/mBio.00895-21

Oldroyd, G. E. D. (2013). Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. Nature Reviews Microbiology, 11(4), 252–263. https://doi.org/10.1038/nrmicro2990

Ördögh, L., Vörös, A., Nagy, I., Kondorosi, É., & Kereszt, A. (2014). Symbiotic plant peptides eliminate Candida albicans both in vitro and in an epithelial infection model and inhibit the proliferation of immortalized human cells. BioMed Research International, 2014, 320796. https://doi.org/10.1155/2014/320796

Ott, T., Dongen, J. T. v., Gu, C., Gunther, C., Krussell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., & Udvardi, M. K. (2005). Symbiotic leghemoglobin is crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Current Biology, 15(6), 531–535. https://doi.org/10.1016/j.cub.2005.01.042

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-ware quantification of transcript expression. Nature Methods, 14, 417–419. https://doi.org/10.1038/nmeth.4197

Pecrix, Y., Staton, S. E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinoud, M. F., Latrasse, D., Zouine, M., Zahm, M., Kreplak, J., Mayjonade, B., Satgé, C., Perez, M., Cauet, S., Marande, W., Chantry-Darmon, C., Lopez-Roques, C., et al. (2018). Whole-genome landscape of Medicago trunculata symbiotic genes. Nature Plants, 4(12), 1017–1025. https://doi.org/10.1038/s41477-018-0286-7

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nature Methods, 8(10), 785–786. https://doi.org/10.1038/nmeth.1701

Pueppke, S. G., Bolanos-Vásquez, M. C., Werner, D., Bec-Férté, M.-P., Promé, J.-C., & Krishnan, H. B. (1998). Release of flavonoids by the soybean cultivars McCall and Peking and their perception as signals by the nitrogen-fixing symbiont Sinorhizobium fredii. Plant Physiology, 117(2), 599–606. https://doi.org/10.1104/pp.117.2.599

Pueppke, S. G., & Broughton, W. J. (1999). Rhizobium sp. strain NGR234 and R. fredii USDA257 share exceptionally broad, nested host ranges. Molecular Plant-Microbe Interactions, 12(4), 293–318. https://doi.org/10.1094/MPMI.1999.12.4.293

Rahioui, I., Eyyraud, V., Karaki, L., Sasse, F., Carre-Piermart, M., Qin, A., Zheng, M. H., Toepper, S., Sivignon, C., Royer, C., Da Silva, P., & Gressent, F. (2014). Host range of the potential biopesticide pea albumin 1b (PA1b) is limited to insects. Toxicol. Annals, 69, 67–76. https://doi.org/10.1111/j.taxon.2014.07.004

Recourt, K., Schildkaste, J., Kjïne, J. W., van Brussel, A. A. N., & Lugtenberg, B. J. J. (1991). Isolation of Vicia sativa subsp. nigra roots with Rhizobium leguminosarum biovar viciae results in release of nod gene activating flavonanes and chalcones. Plant Molecular Biology, 16(5), 841–852. https://doi.org/10.1007/BF00015076

Roux, B., Rodde, N., Jardinoud, M.-F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., Moreau, S., Debélle, F., Capela, D., De Carvalho-Niebel, F., Gouzy, J., Bruand, C., & Gamas, P. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. The Plant Journal, 77(6), 817–837. https://doi.org/10.1111/tpj.12442

Sallet, E., Roux, B., Sauviac, L., Jardinoud, M.-F., Carrère, S., Faraut, T., De Carvalho-Niebel, F., Gouzy, J., Gamas, P., Capela, D., Bruand, C., & Schiex, T. (2013). Next-generation annotation of prokaryotic genomes with EuGene-P: Application to Sinorhizobium meliloti 2011. DNA Research, 20(4), 339–354. https://doi.org/10.1093/dnares/dst014

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In Gene Prediction (Vol. 1962) (pp. 227–245). 10.1007/978-1-4939-9173-0_14. Springer.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Molecular Systems Biology, 7(1), 539. https://doi.org/10.1038/msb.2011.75

Simsek, S., Ojanen-Reuhs, T., Stephens, S. B., & Reuhs, B. L. (2007). Strain-specific human cells. BioMed Research International, 2007, 1–10. https://doi.org/10.1155/2007/436129

Song, L., & Florea, L. (2015). Rcrrector: Efficient and accurate error correction for Illumina RNA-seq reads. GigaScience, 4(1), s13742-s13015. https://doi.org/10.1186/s13742-015-0089-y
Weisberg, A. J., Kim, G., Westwood, J. H., & Jelesko, J. G. (2017). Sequencing and de novo assembly of the Toxicodendron radicans (poison ivy) transcriptome. Genes, 8(11), 317. https://doi.org/10.3390/genes8110317

Wilson, J. K. (1939). Leguminous Plants and Their Associated Organisms (p. 221). Cornell University Agricultural Experiment Station.

Yang, S., Tang, F., Gao, M., Krishnan, H. B., & Zhu, H. (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. Proceedings of the National Academy of Sciences of the United States of America, 107(43), 18735–18740. https://doi.org/10.1073/pnas.1011957107

Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., Price, P. A., Pan, H., Wang, D., Griffitts, J. S., Bisseling, T., & Zhu, H. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in Medicago truncatula. Proceedings of the National Academy of Sciences of the United States of America, 114(26), 6848–6853. https://doi.org/10.1073/pnas.1700715114

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D. R., Meyers, B. C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., ... Roe, B. A. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature, 480(7378), 520–524. https://doi.org/10.1038/nature10625

Zamani, M., diCenzo, G. C., Milunovic, B., & Finan, T. M. (2017). A putative 3-hydroxyisobutyryl-CoA hydratase is required for efficient symbiotic nitrogen fixation in Sinorhizobium melloti and Sinorhizobium fredii. Applied and Environmental Microbiology, 83(7), 1986–1993. https://doi.org/10.1128/AEM.02210-16

How to cite this article: Huang, R., Snedden, W. A., & diCenzo, G. C. (2022). Reference nodule transcriptomes for Melilotus officinalis and Medicago sativa cv. Algonquin. Plant Direct, 6(6), e408. https://doi.org/10.1002/pld3.408

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.