

Article

Probing Biological Nitrogen Fixation in Legumes Using Raman Spectroscopy

Abdolabbas Jafari ^{1,*}, Kritarth Seth ², Armin Werner ¹, Shengjing Shi ², Rainer Hofmann ³
and Valerio Hoyos-Villegas ⁴

¹ Lincoln Agritech, Lincoln University, Lincoln 7647, New Zealand; werner@lincolnagritech.co.nz

² AgResearch, Lincoln 7608, New Zealand; kritarth.seth@agresearch.co.nz (K.S.); shengjing.shi@agresearch.co.nz (S.S.)

³ Plant Biology Department, Lincoln University, Lincoln 7647, New Zealand; rainer.hofmann@lincoln.ac.nz

⁴ Faculty of Agricultural and Environmental Sciences, McGill University, Montreal, QC H3A 0G4, Canada; valerio.hoyos-villegas@mcgill.ca

* Correspondence: jafari@lincolnagritech.co.nz

Abstract: Biological nitrogen fixation (BNF) by symbiotic bacteria plays a vital role in sustainable agriculture. However, current quantification methods are often expensive and impractical. This study explores the potential of Raman spectroscopy, a non-invasive technique, for rapid assessment of BNF activity in soybeans. Raman spectra were obtained from soybean plants grown with and without rhizobia bacteria to identify spectral signatures associated with BNF. δN^{15} isotope ratio mass spectrometry (IRMS) was used to determine actual BNF percentages. Partial least squares regression (PLSR) was employed to develop a model for BNF quantification based on Raman spectra. The model explained 80% of the variation in BNF activity. To enhance the model's specificity for BNF detection regardless of nitrogen availability, a subsequent elastic net (Enet) regularisation strategy was implemented. This approach provided insights into key wavenumbers and biochemicals associated with BNF in soybeans.

Keywords: pattern recognition; chemometrics; biosystems; precision agriculture; non-destructive; rhizobia; symbiosis

Citation: Jafari, A.; Seth, K.; Werner, A.; Shi, S.; Hofmann, R.; Hoyos-Villegas, V. Probing Biological Nitrogen Fixation in Legumes Using Raman Spectroscopy. *Sensors* **2024**, *24*, 4944. <https://doi.org/10.3390/s24154944>

Academic Editors: Simona Violino, Loredana Canfora, Francesca Antonucci, Roberto Ciccoritti

Received: 28 May 2024

Revised: 24 July 2024

Accepted: 27 July 2024

Published: 30 July 2024



Copyright: © 2024 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

Biological nitrogen fixation (BNF) is a critical process in sustainable agriculture. It provides a cost-effective alternative to synthetic nitrogen fertilisers, reducing reliance on non-renewable resources and potential environmental impacts [1,2]. However, accurately measuring BNF activity remains a challenge. Existing methods, such as the acetylene reduction assay, natural abundance method, and δN^{15} isotope dilution method, all have limitations. The natural abundance method, while considered the most accurate, is destructive and time-consuming [2]. Isotope ratio mass spectrometry (IRMS) offers a direct measure of N_2 fixation using $^{15}N_2$ incorporation [2,3], but it is not suitable for high-throughput screening or field applications due to its complexity and cost. Nitrogen balance calculations, based on long-term changes in soil and crop nitrogen content, can estimate potential BNF but lack the precision of direct measurement [2].

In recent years, Raman spectroscopy (RS) has emerged as a promising tool in plant biology and agriculture [4]. This non-destructive technique analyses the vibrational modes of molecules within a sample, generating a characteristic “fingerprint” spectrum [4]. Compared with traditional methods, RS offers several advantages: it is non-destructive, requires minimal sample preparation, and allows simultaneous analysis of multiple metabolites [2,5,6].

Legumes, through symbiosis with bacteria in their root systems, fix atmospheric ni-

trogen (N₂) and convert it into usable forms. This process leads to the production of amino acids, which are primarily synthesised in the roots and transported to the aerial parts of the plant for essential processes like photosynthesis. RS has been successfully applied to analyse various plant metabolites, including amino acids [7], sugars [8,9], and nucleic acids [10]. Additionally, it has been used to study secondary metabolites in medicinal plants, detect plant stress responses, and analyse plant growth regulators [11]. Notably, RS has also been used to analyse cultured bacteria extracted from the rhizosphere, enabling the investigation of plant-microbe interactions [12]. Furthermore, studies have demonstrated the potential of RS gas analysis within closed chambers for the continuous monitoring of plant gas exchange, including N₂ fixation rates [13]. However, macro-scale Raman measurements of plant organs can be challenging due to variations in chemical composition across different tissues. Leaves, stems, and roots have distinct chemical profiles, and even within the same organ, variations can occur depending on location (e.g., distance from the petiole) [14]. Additionally, the orientation of molecules within the tissue can influence the polarised Raman spectra [14]. These variations necessitate targeted measurements from specific regions of the plant or the use of chemometric techniques to account for and analyse such differences.

This study explores the potential of RS as a non-invasive, high-throughput method for measuring BNF activity in soybean plants. We hypothesise that BNF, by altering the plant's metabolic state and potentially affecting energy consumption, could either directly influence the Raman spectra of plant tissues or indirectly lead to changes in the composition of other chemical compounds, ultimately resulting in spectral variations. We aimed to investigate whether these spectral changes can be linked to BNF activity, allowing for its quantification. Specifically, we sought to determine if differences in soybean plants' nitrogen fixation levels could be distinguished through variations in their leaf Raman spectra.

2. Materials and Methods

Previous studies have established that nitrogen concentration in the growth medium reduces nitrogen fixation in legumes [15–18]. Therefore, in this study, nitrogen concentration was used as an agent to provide different levels of nitrogen fixation. Nitrogen fixation is suppressed when there is adequate nitrogen in the rooting medium, as the exchange of nitrogen for carbohydrates between plants and bacteria is halted. This means that legumes no longer provide their carbohydrates to host rhizobia in return for fixed nitrogen, which comes at an energy cost to the plant [19]. Using this phenomenon, we prepared growth media with varying nitrogen concentrations to achieve different nitrogen fixation rates. Specifically, we grew soybean plants in vermiculite using four different nitrogen concentrations (2 mM, 6 mM, 10 mM, and 14 mM solutions of ammonium nitrate) under controlled conditions in a growth chamber, with five replications for each treatment.

Soybean plants (*Glycine max* L. cv. Sonya, non-genetically modified) were grown in a closed, semi-sterile system for six weeks. To ensure precise control over nitrogen availability, the plants were cultivated in 4-litre pails ($n = 56$) filled with vermiculite (grade 2; compressed). Each box was filled with 2.2 L McKnights solution [20,21] and autoclaved (30 min, wet cycle, 121 °C) to eliminate any background rhizobia present. Before autoclaving, the lids of the boxes were drilled to make a 1.5 cm hole in the center for the plant stem to emerge. Soybean seeds were surface sterilised (70% ethanol, 30 s; 1% NaOCl, 2 min; 4X rinses, sterilised deionised water, 30 s) and pre-germinated for 6 days (22 °C; dark) in a 5 cm deep bed of vermiculite and water (3:1; autoclaved) in trays.

Healthy seedlings were then transplanted into the boxes under sterile conditions, carefully placed through the holes on the lids, at one seedling per box. The seedlings were allowed to establish for 2 days in a clean growth room with 600 $\mu\text{mol}/\text{m}^2/\text{s}$ of light for 16 h day at 24 °C and 8 h night at 19 °C. Rhizobia-inoculated plants were inoculated with a pipette at the base of the seedling with 1×10^7 colony forming units (cfu) of *Brad-*

Bradyrhizobium japonicum Strain ICMP3163 (Landcare research, Lincoln, New Zealand) twice, on day 1 (day of inoculation; 8 days from germination) and again on day 8 to give the seedlings the maximum chance of nodule formation. Nitrogen (N) was added as 5% atom ^{15}N ammonium nitrate (Sigma Aldrich, St. Louis, MO, USA) at concentrations of 2, 6, 10, and 14 mM of N. To account for increasing N requirements of the plant while growing, N was not added all at once, but was split into four doses of 10%, 20%, 30%, and 40% added on days 1, 8, 15, and 22, respectively. Plants were watered with sterilised deionised water as and when needed.

To establish a reliable model, we employed isotope ratio mass spectrometry (IRMS) to measure BNF rates, which will serve as the ground truth for subsequent analyses.

2.1. Experimental Design

Achieving varying levels of nitrogen fixation (BNF) for treatment design presented a challenge. Ideally, we sought a manipulation method that would not directly alter plant chemistry, as this could confound spectral data analysis. However, we utilised nitrogen availability as a tool to indirectly control BNF. While this approach influences spectral features, it becomes difficult to disentangle whether observed spectral variations primarily reflect changes in BNF activity or simply the introduced nitrogen (ammonium nitrate) itself.

To distinguish between spectral features arising from BNF activity and those solely due to nitrogen variation, we employed two treatment sets: fixing and non-fixing (Figure 1). Soybean samples in the non-fixing set were grown under completely sterile conditions to ensure that no bacteria for nitrogen fixation existed in the growing medium, while the other set was inoculated with *Bradyrhizobium japonicum* bacteria.

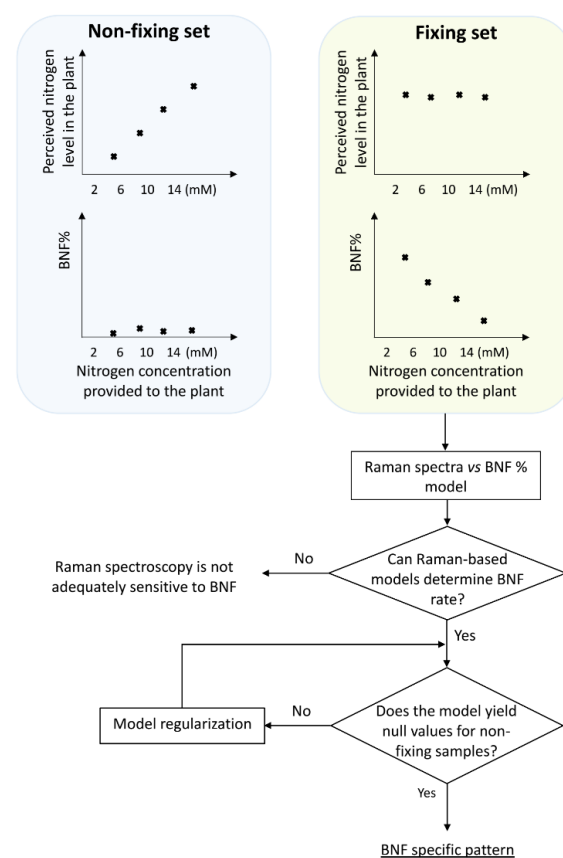


Figure 1. Experimental design and procedure for extracting BNF specific pattern from Raman spectra of the aerial parts of soybean plants. Spectra were collected at six weeks after inoculation, when the total nitrogen content became consistent across all treatments within the nitrogen-fixing set.

We initiated the procedure by developing a model using the Raman data from the fixing set, enabling it to take Raman spectra of the leaves as input and provide nitrogen fixation rates as output. To achieve this, the input data were needed to cover the full range of nitrogen fixation variations. In addition, it is important to note that while we provided the plants with varying nitrogen concentrations, this did not guarantee a uniform distribution of BNF rates across the range. To address this, we conducted δN^{15} measurements to assess BNF variations in the samples and determine the nitrogen accumulated in the plants at the time of measurement.

The schematic illustration in Figure 1 explains the experimental design, including the division of plants into two groups: those inoculated with nitrogen-fixing bacteria (BNF set) and those grown without bacteria (non-fixing set). The figure also depicts the expected observations of the two groups under varying nitrogen treatments. This design acts as a key control experiment to validate the specificity of the models in BNF measurements. We compared the Raman spectral patterns of both groups to assess if the model captures general nitrogen responses or changes specific to nitrogen fixation. If the spectral patterns in both sets followed the same trends with varying nitrogen levels, it would indicate the model might be capturing general nitrogen responses rather than changes specific to nitrogen fixation. This would suggest the method is unsuitable for BNF quantification.

Our δN^{15} isotope ratio mass spectrometry (IRMS) experiments confirmed that at six weeks after germination, soybean plants in the BNF set achieved a mature symbiotic state with the bacteria fully supplying their nitrogen needs. Therefore, it is reasonable to expect that six weeks after inoculation, there should be no discernible differences in N-related traits within the spectra of fixing plants, while the N-related distinctions may persist in the non-fixing set. Nevertheless, it is worth exploring the possibility of residual effects from prior nitrogen deficiency in the spectra of fixing plants. For a robust method specifically measuring BNF (independent of nitrogen variations), evaluation with non-fixing samples is crucial. This ensures the model does not simply detect general nitrogen variation.

2.2. Raman Spectra Acquisition

Raman spectra were collected from the first fully developed leaves of soybean plants using a Horiba iHR320 spectrometer (Horiba, Kyoto, Japan). To optimise signal quality while minimising plant tissue fluorescence, a 785 nm laser source with 350 mW power was utilised for excitation.

The leaves were placed on the acquisition stage of the spectrometer, and the laser was focused on the leaf surface without any sample preparation. To ensure precise positioning of all sampling spots within the focal point of the laser beam, a fixture was used to securely hold the leaves, preventing deformation and local undulation of the leaf surface.

Raman spectra were collected from the samples across the range of 400 cm^{-1} to 1800 cm^{-1} . Each spectrum was acquired with a 20-s acquisition time and three accumulations. For improved signal-to-noise ratio (S/N), three measurements were taken from nearly the same spot on the leaves. To mitigate potential dehydration or burning of the leaf texture due to the laser intensity and acquisition time, the data acquisition spot was slightly shifted by approximately $\pm 2\text{ mm}$ in each iteration relative to the initial spot.

To validate the accuracy of our Raman detection in capturing the chemical composition related to nitrogen fixation, we conducted a comparative analysis using known plant chemicals, including allantoin, β -carotene, pectin, and glycerol. By comparing the experimentally obtained Raman spectra of these compounds with established spectral databases, we were able to confirm the reliability of our Raman system in identifying key functional groups associated with these molecules.

2.3. Data Processing and Analysis

The collected spectra underwent data preprocessing steps, which involved removing cosmic rays and smoothing to eliminate random noise. Various combinations of polynomial orders and kernel sizes were tested using Savitzky–Golay filters. Ultimately, a filter with a polynomial order of 13 and a frame length of 51 demonstrated optimal performance by effectively eliminating random noise while preserving essential Raman peaks. As a result, this specific filter configuration was chosen for subsequent processing steps.

For excitation, a 785 nm laser source was chosen. Shorter wavelengths, like 532 nm, can generate a significantly higher level of fluorescence, especially in some biological materials such as plant leaves. This fluorescence can overwhelm the weaker Raman signal, making it difficult to analyse the desired Raman spectrum. Conversely, longer wavelengths, such as 1064 nm, while minimising fluorescence, often suffer from a significantly weaker Raman scattering effect [22]. Therefore, 785 nm was used as a compromise between minimising fluorescence and maximising the Raman signal intensity for our samples.

Even with the utilisation of a 785 nm laser, the Raman spectra of the soybean samples still showed significant fluorescence intensity. To address this, a baseline detection procedure [23] was employed, with a modification that used an iterative moving average approach instead of polynomial kernels. The modification was justified based on a comparison between the results obtained from the two types of kernels, i.e., moving average versus polynomials. The iterative moving average approach was found to be more effective in reducing fluorescence without distorting the underlying Raman signal.

Initial evaluation of baseline-subtracted spectra revealed that a kernel size of 50 wavenumbers with 50 iterations yielded optimal performance in isolating pure Raman peaks from the fluorescent background using a MATLAB implementation.

To detect BNF-related patterns, partial least squares regression (PLSR) models were developed using Raman data from the leaves of “Fixing set” plants. PLSR is a statistical technique that relates a set of predictor variables (X) to a set of response variables (Y) by extracting latent variables (underlying factors) that explain the maximum covariance (shared variance) between X and Y. Unlike standard regression, which directly models Y based on X, PLSR finds a set of latent variables that capture the most important information in both X and Y. It then uses these latent variables to build a model that predicts Y. It is mostly useful when there are many correlated predictor variables (multicollinearity) which can cause problems in standard regression [24].

The calibration set for developing the model comprised soybean samples from the nitrogen-fixing set. Model performance was assessed using the coefficient of determination (R^2) and root mean square error (RMSE) for both calibration and cross-validation sets. Leave-one-out cross-validation (LOOCV) was employed, where each soybean sample from the nitrogen-fixing set was sequentially excluded from the training set and used as a validation set. This iterative process provided a robust estimate of model performance.

2.4. Model Regularisation

The model’s target BNF estimation for the non-fixing samples was expected to be zero. To refine the BNF information from the Raman spectra of the N-related variations due to nitrogen concentrations applied to the root medium, model regularisation was needed. As illustrated in the final stage of Figure 1, this regularisation procedure was carried out using elastic net (ENet) regularisation.

This procedure involves adjusting the influential vectors of the model to achieve a target null value for the non-fixing set. While this adjustment may reduce the model’s performance on the fixing set, it results in a more specific BNF determination model. We

determined the trade-off point between maximising BNF sensitivity and minimising N sensitivity by applying elastic net regularization.

While this study employs the concept of regularisation, it differs from the typical approach. Here, we do not adjust the model to reduce variance on the validation set. Instead, regularisation is used to force the model to predict the lowest possible values for the non-fixing set, minimising its sensitivity to nitrogen concentration. The resulting performance metrics will then indicate the extent of BNF-representative information captured within the Raman spectra.

Elastic net (EN) regularisation is a hybrid approach that combines L1 (lasso regularisation) and L2 (Tikhonov or ridge regularisation) methods. In all three techniques (L1, L2, and EN), regularisation terms, also known as penalties, are added to the ordinary least squares cost functions. These penalties aim to minimise the model's variance.

The regularisation strength is typically controlled by the hyperparameter λ which determines the amount of shrinkage applied to the coefficients. The model coefficients, β and β_0 are calculated by minimising the cost function $J(\beta)$ [25] as follows:

$$J(\beta) = \frac{1}{2N} \left(\sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right) \quad (1)$$

where

$$P_\alpha(\beta) = \frac{(1 - \alpha)}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 = \sum_{j=1}^P \left(\frac{(1 - \alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right) \quad (2)$$

$P_\alpha(\beta)$: Elastic net's penalty term

We aimed to reduce the sensitivity of the model to factors mostly indicative of nitrogen concentration, ideally achieving the lowest possible value for this plant set. The regularisation process will determine the optimal λ that yields the lowest BNF for the non-fixing plants. An equal balance of L1 and L2 penalties ($\alpha = 0.5$) was incorporated into the cost function. While this regularisation might decrease model performance for the fixing set, it tailors the model specifically for BNF measurement.

3. Results and Discussion

The Raman spectra of both "Fixing" and "Non-fixing" soybean leaves exhibited broadly similar patterns, with the most prominent variations occurring at the peaks around 747 cm^{-1} , 1160 cm^{-1} , and 1525 cm^{-1} . These peaks are associated with molecular vibrations of specific biomolecules including pectin [26] and carotenoids [7].

The average Raman spectra of soybean leaves under different nitrogen treatments are shown in Figure 2 for the two sets of soybean plant samples, i.e., nitrogen-fixing vs. non-fixing sets. Each spectrum represents an average of 15 measurements, including five replicates of three measurement points each.

A comparison between the spectra of the fixing and non-fixing sets reveals that an easily inferable trend in the variations of the Raman intensities is not readily discernible at any specific Raman shift points. This indicates that BNF-related metabolites did not explicitly change the Raman spectrum. Instead, the information on BNF activities is embedded within the Raman data and must be extracted using sophisticated chemometric techniques such as PLSR, as explained in the Methods section.

As illustrated in the first stage of Figure 1, the PLSR models were developed using the Raman data from the "Fixing set". The performance of the models indicated the feasibility of estimating BNF from the leaves' spectra (Figure 3a).

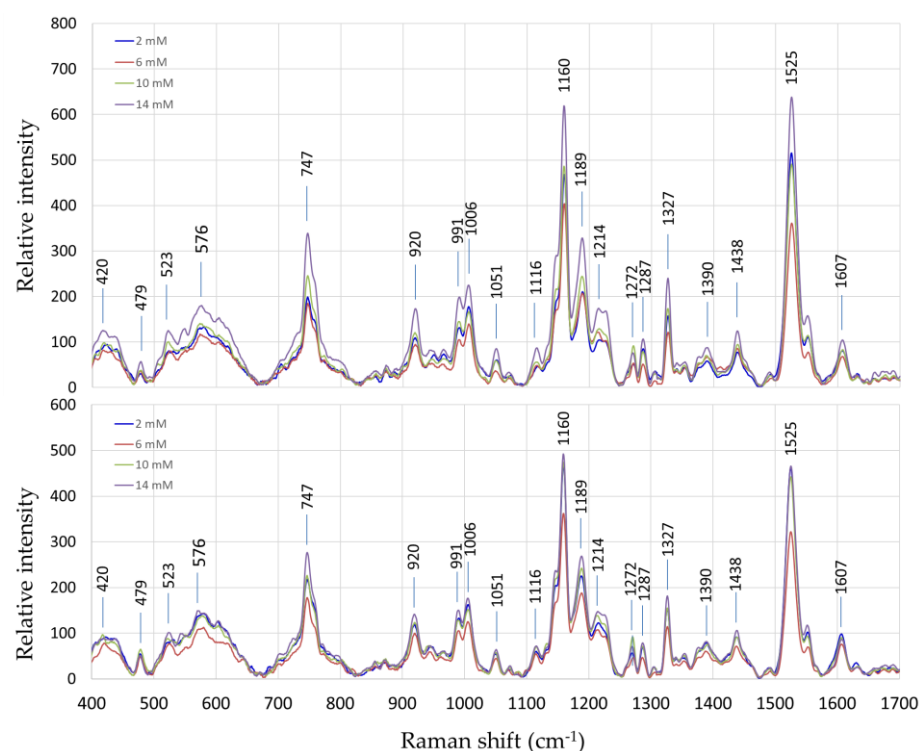


Figure 2. Average Raman spectra of soybean leaves under different nitrogen treatments; nitrogen-fixing samples (**top**) and non-fixing samples (**bottom**).

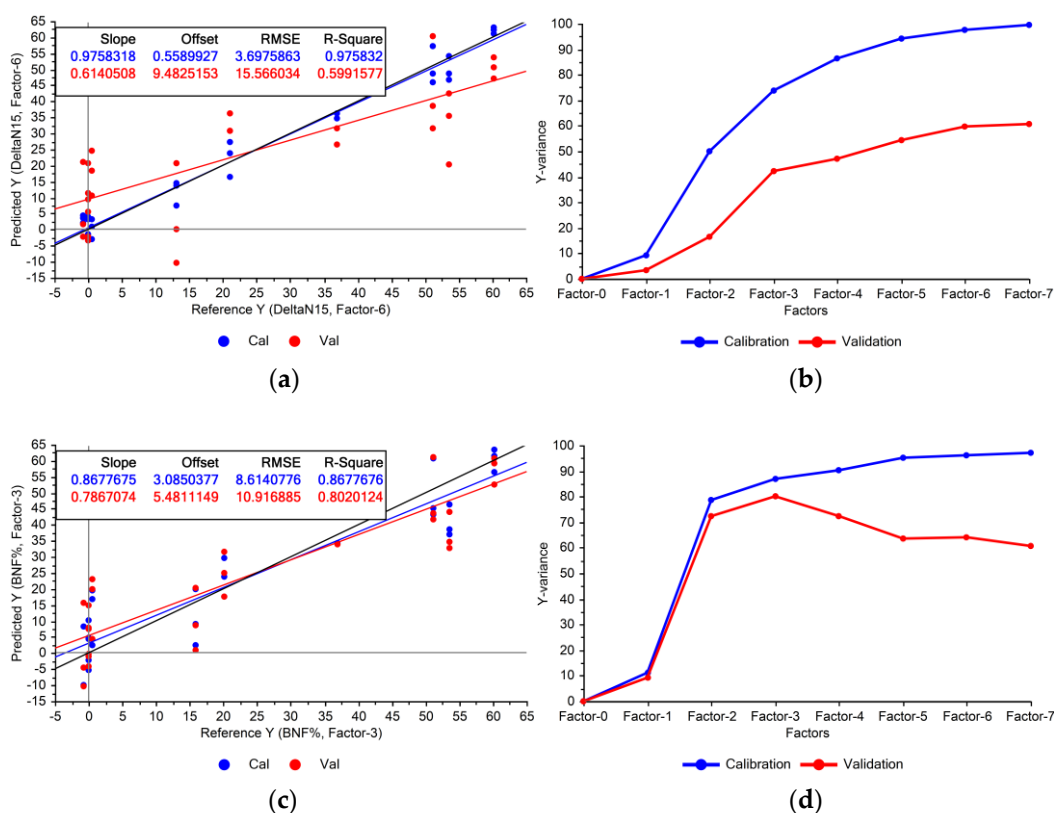


Figure 3. Performance of the BNF estimation models using Raman spectra acquired from fixing plants for calibration (Cal) and validation (Val) datasets: (a) using full Raman spectrum (600-1770 cm⁻¹), (c) using selected Raman peaks. Explained variances by PLSR factors are shown in: (b) full Raman spectrum, (d) selected Raman peaks.

3.1. Model Optimisation

We evaluated various normalisation and standardisation techniques, including area normalisation, standard normal variate (SNV), and multiplicative scatter correction (MSC), on the model's performance. This evaluation showed a reduction rather than an improvement in performance. Consequently, we concluded that the absolute intensity of the Raman data is as crucial as the relative intensity of the Raman shifts. Therefore, normalising or standardising the data would result in the loss of important information.

Figure 3a illustrates that the constructed model exhibited significant variance, indicating notable disparities between calibration and cross-validation accuracies. This suggests overfitting, where the model memorises training data variations rather than capturing underlying patterns, compromising its generalisability to unseen data. Addressing this issue may involve removing features with low importance from the model.

Using the entire Raman spectrum in the modelling process can lead to overfitting due to the numerous features involved, increasing the chance of memorisation. The divergence between the variances explained by the model during training and cross-validation (Figure 3b) is another sign of potential overfitting. While omitting certain features might reduce performance on the calibration set, it could improve validation results and enhance the model's generalisability. Given the already high explained variances on the training set (100% with seven factors), there is ample headroom to reduce the variance in the validation set (Figure 3b).

As the first step towards reducing variables in the model, we used Raman peaks instead of the entire spectrum. Unlike NIR spectroscopy and hyperspectral imaging, Raman peaks directly correspond to specific molecular bonds, indicating specific molecules or compounds. Therefore, a model using selected Raman peaks offers clearer interpretability. We evaluated model performance using only the distinctive peaks listed in Table 1, reducing the input features from 1097 wavenumbers to a smaller set of specific ones.

We used variable importance in projection (VIP) scores to identify the most influential features. Features with a VIP score below 1 were excluded from the PLSR model. Figure 3c shows the model's performance using only the selected Raman peaks for BNF estimation. The explained variances (Figure 3d) suggest improved generalisability compared with the model using the entire spectrum. The model achieved optimal performance with the first three PLS factors, capturing 80% of the variation within the selected Raman peaks. This aligns with expectations, as nitrogen fixation did not occur in the non-fixing set grown under sterile conditions, confirmed by the absence of root nodules.

These findings validate our initial assumption that specific features within the Raman spectra of leaves can reflect nitrogen fixation rates. To further confirm this, we evaluated the model's performance using spectra from the non-fixing sets grown under different nitrogen treatments.

3.2. Regularisation Results

To assess the model's ability to estimate BNF exclusively, irrespective of nitrogen levels, it was validated using Raman spectra from leaves of the non-fixing sets. The model yielded an average value of 13.7% BNF for these plants. While the experiment aimed to eliminate the influence of nitrogen variations, Figure 4 reveals some, albeit minor, correlations between estimated BNF and nitrogen treatment levels. This suggests that some features within the Raman spectra of the leaves still capture information about nitrogen variations in the plant.

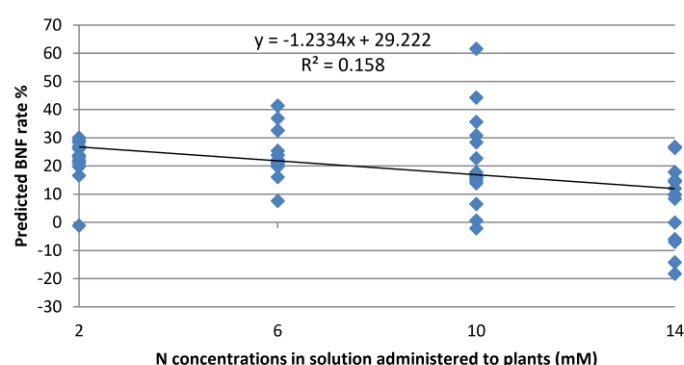


Figure 4. Output of the BNF model on Raman spectra collected from “non-fixing” soybean plants with various levels of nitrogen fixation.

To further refine the BNF-related Raman peaks and potentially identify the associated chemical compounds, regularisation was employed. The aim of regularisation in this study was to decrease the sensitivity of the model to features (wavenumbers) associated with nitrogen concentration rather than nitrogen fixation.

The variation in the model coefficients at various λ values is depicted in Figure 5. Increasing λ increases the penalties on the model, which in turn reduces sensitivity and, in some cases, eliminates certain features depending on the chosen threshold.

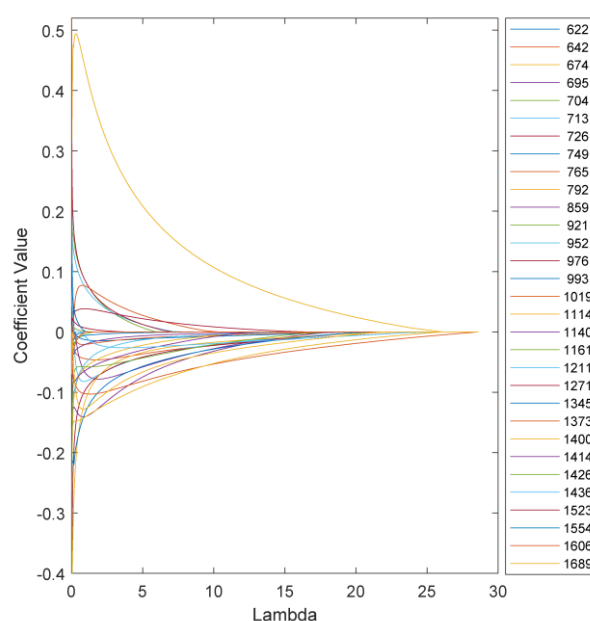


Figure 5. Variation in the ENet model coefficients at various λ values.

The optimal combination of coefficients was determined based on the minimum output of the elastic net (ENet) model at various λ values. Figure 6 depicts the optimal point at $\lambda = 20$, where the non-fixing set achieves a minimum BNF output of 12.6.

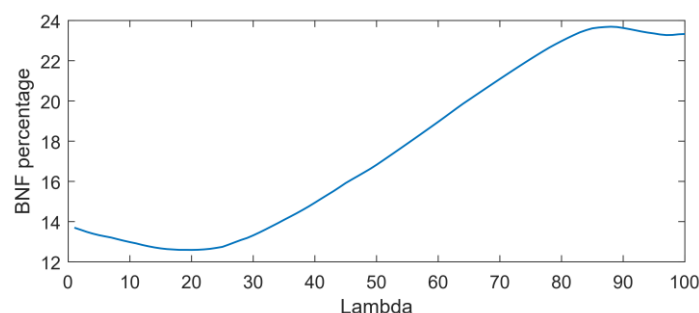


Figure 6. Output of the ENet model for non-fixing soybean samples at various λ values.

At this stage, the coefficients of the features in the model at the optimum λ can be determined. Figure 7 illustrates the coefficients of the model without regularisation ($\lambda = 0$) and at the optimal regularisation ($\lambda = 20$) which corresponds to a BNF% of 12.6% for the non-fixing set.

Once the optimal λ value is determined, the coefficients associated with each feature in the model can be analysed. Figure 7 compares the model coefficients without regularisation ($\lambda = 0$) to those obtained with optimal regularisation ($\lambda = 20$). The latter achieves a minimum BNF value of 12.6% for the non-fixing set.

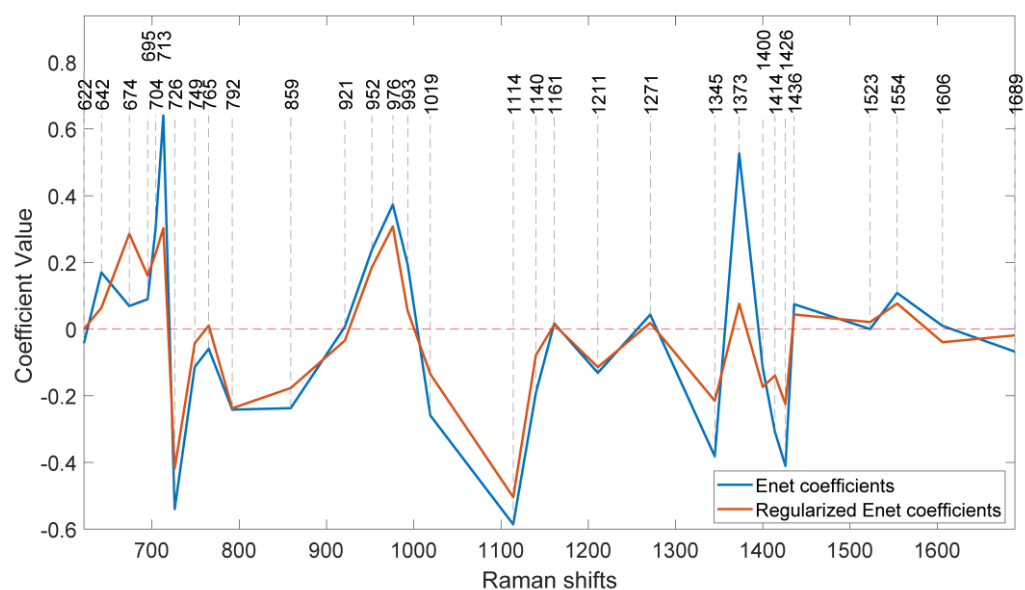


Figure 7. Coefficients of the ENet and regularised ENet models.

It is evident that the effects of regularisation were not uniform across all features. Wavenumbers with significantly reduced coefficient values after regularisation likely correspond to nitrogen concentration, suggesting a weaker contribution to biological nitrogen fixation (BNF). Additionally, the direction of change in coefficients varies across peaks. For example, the regularised model shows a higher coefficient at 674 cm^{-1} compared with the non-regularised model, while the coefficient at 713 cm^{-1} has decreased. Notably, some coefficients were eliminated entirely due to the L1 (lasso) penalty.

To identify the most BNF-related peaks, we can select wavenumbers that meet two criteria: (1) a coefficient greater than 0.1 in the regularised model and (2) a reduction of more than 10% in the coefficient value compared with the non-regularised model. Applying these criteria to the 31 features in Table 1 results in the selection of 15 key wavenumbers.

Interestingly, the two prominent bands at 1160 cm^{-1} and 1525 cm^{-1} , previously associated with carotenoids in Figure 1, do not appear to play significant roles in biological nitrogen fixation (BNF) despite being strong indicators of plant greenness and nitrogen uptake.

Carotenoids typically exhibit two strong Raman bands due to polyene chain stretching vibrations [27]. As changes in nitrogen concentration directly influence leaf chlorophyll content, these peaks were expected to be prominent features in the nitrogen estimation models. However, Table 1 reveals that the key BNF-related peaks are not necessarily the strongest or most distinctive peaks in the leaf spectrum. Instead, the information on BNF appears to be encoded in Raman peaks with relatively lower intensity.

An interesting observation is that while the regularised model's coefficients generally decrease across most peaks, the coefficient at 674 cm^{-1} actually increases. This peak is associated with C–C–O rocking and deformation vibrations in glycerol [28]. This finding is particularly intriguing given the high concentration of glycerol in soybeans [29]. It suggests a potential link between biological nitrogen fixation and glycerol content, warranting further investigation.

Table 1. Assignments of vibrational bands to the influential Raman peaks in BNF specific model developed for soybean plants.

Band (cm^{-1})	Vibrational Mode	Assignment
674	$\nu(\text{C}-\text{C}-\text{O}) + \delta(\text{C}-\text{C}-\text{O})$	Carbohydrates [30], Glycerol [28]
695	$\delta(\text{C}-\text{C}-\text{H})$ twisting	Cellulose Avicel (II) [31]
704	ring deformation	Cytosine [32]
713	C–S–C asymmetric stretch; H ₂ O rocking	Isoleucine, Methionine [33]
726	ring breath	Adenine [34]
859	$\tau(\text{C}-\text{N})$	Arginine, Serine [33]
952	$\delta(\text{CCH})$, $\delta(\text{COH})$	Pectin [26]
976	CH ₂ rocking	Glycerol [28]
1019	C–O–H deformation	Alcoholic hydroxyl [33]
1114	$\nu_s(\text{C}-\text{O}-\text{C})$	Glycosidic ring breathing [35], Cellulose [36]
1211	$\delta(>\text{C}=\text{C}-\text{C}-)$ β -ring	β -Carotene [37–39]
1345	$\nu(\text{C}-\text{O})$; $\delta(\text{C}-\text{O}-\text{H})$	Carbohydrates [30]
1400	ν ; $\delta(\text{C}-\text{OH})\text{COOH}$; $\delta_s(\text{CH}_3)$	Pectin [26], β -carotene [37], Cellulose [36]
1414	$\delta(\text{C}-\text{H})$	Deoxyadenosine triphosphate [28]
1426	$\nu_s(\text{COO}-)$; CH ₂ deformation	Aspartic acid, Glutamic acid, Methionine [28,33]

Soybeans are considered as a complete protein source, containing all nine essential amino acids. The peak at 713 cm^{-1} can be attributed to protein elements and amino acids such as isoleucine and methionine [33]. Isoleucine is essential for nitrogen metabolism, and changes in the source of nitrogen uptake may influence its production. The quantity of isoleucine in plant leaves varies depending on factors such as plant species, growth conditions, and leaf part. While plant leaves are generally not rich sources of isoleucine compared with animal products, soybeans stand out for their good isoleucine content [40]. The peak at 726 cm^{-1} can also be assigned to adenine [34], a fundamental building block of DNA and RNA. Peak positions can shift depending on factors like the pH of the environment. This peak is often attributed to in-plane ring breathing vibrations within the adenine molecule.

Serine and arginine are other essential amino acids for legumes, playing crucial roles in their growth and development, and they exhibit a prominent peak at 859 cm^{-1} [33]. A recent study [41] suggests that arginine serves as a precursor molecule in nitrogen fixa-

tion process, providing the necessary building blocks for the bacteria. This research proposes a metabolic cycle where plant-provided arginine and succinate fuel the nitrogen-fixing bacteria within the nodules.

The peak at 1426 cm^{-1} , the last nitrogen-related peak, may result from overlapping peaks of various amino acids such as aspartic acid, tryptophan, valine, and methionine [33]. Nitrogen, in its structural molecule, contributes to the vibration at 1426 cm^{-1} , where the bond is attributed to the carboxyl group connected to the amino group NH_2 . The accumulation of intensity from the peaks of these amino acids at this wavenumber can compensate for the low concentration of each individual amino acid in soybean plants. Although the combination of wavenumbers selected by our approach demonstrated effective performance in discerning biological nitrogen fixation (BNF), it is not inherently exclusive. The substantial collinearity among the spectral features suggests that alternative Raman wavenumbers could convey similar levels of information. Consequently, insights derived from these data might be transferable through different combinations of Raman shifts. As a result, models with diverse specifications and input variables may achieve comparable performance.

A key advantage of the method presented here is that it eliminates the need for sample preparation. Raman spectroscopy allows for the direct collection of spectra from live leaf samples, significantly reducing measurement costs. This contrasts favourably with existing methods like the ^{15}N isotope dilution and natural abundance techniques. These methods require artificial enrichment of the soil with ^{15}N -labelled fertiliser and the inclusion of reference crops near test plots, both of which can be cumbersome and expensive [42,43]. The Raman spectroscopy-based method explored in this research offers a more cost-effective and practical approach for BNF measurements.

Although the Raman method developed by Jochum et al. [13] provided excellent temporal resolution for tracking nitrogen fixation changes, its limitations include the need for a gas-collecting chamber enclosing the plant and potential interference from soil-released N_2 due to denitrification. These factors hinder its practicality for high-throughput BNF measurement in field settings. Conversely, the optical spectroscopy method developed by Zhang et al. presented a low-cost option with high compatibility with IRMS-based quantification [44]. Our Raman-based method leverages the strengths of optical spectroscopy, such as low cost and rapid analysis, while offering additional advantages. Raman spectroscopy is insensitive to water interference and provides high molecular specificity, enabling us to develop models that require less calibration across diverse environmental conditions.

4. Conclusions

This study demonstrates that Raman spectroscopy can detect changes in the leaf spectra of legumes associated with biological nitrogen fixation (BNF) activity. Analysing these spectral changes allows for the estimation of nitrogen fixation levels in the plants.

The experiments demonstrated that using nitrogen abundance as a control parameter successfully created various levels of nitrogen fixation. However, this approach highlighted the importance of refining the nitrogen fixation model to account for the direct effects of nitrogen abundance on the model's output. Additionally, the use of separate experimental groups for fixing and non-fixing plants proved to be an effective strategy for training and validating a model specifically designed to determine the BNF ratio in soybeans.

The partial least squares regression (PLSR) results for fixing plants supported our initial hypothesis. BNF, likely accompanied by metabolic changes, induced unique spectral signatures in the leaves. These signatures correlated well with the plants' measured BNF rates. Notably, applying the model to non-fixing plants with varying nitrogen levels yielded minimal BNF values. This demonstrates that the model selectively identifies features indicative of BNF, rather than simply reflecting overall plant nitrogen content.

Since plant activity generates a diverse range of metabolites, information related to biological nitrogen fixation (BNF) might also be present in various chemical compounds

and corresponding Raman bands. Consequently, the Raman bands identified in this study for BNF detection might not be entirely unique. However, the novelty lies in incorporating the structural characteristics of BNF-related molecules with multiple Raman peaks into a unified model derived from Raman spectroscopy data. Our findings confirm the presence of specific patterns in the Raman data that act as unique identifiers (fingerprints) of the BNF process. This allows us to leverage these patterns for quantifying nitrogen fixation. However, various factors influence BNF in legumes. Therefore, it is crucial to explore the impact of other parameters on the model's performance to ensure its applicability across various field conditions.

Author Contributions: Conceptualization, A.W., A.J. and V.H.-V.; methodology, A.J., K.S., A.W., V.H.-V., R.H. and S.S.; software, A.J.; validation, A.J. and K.S.; formal analysis, A.J.; investigation, A.J. and K.S.; resources, A.W. and V.H.-V.; data curation, A.J. and K.S.; writing—original draft preparation, A.J.; writing—review and editing, A.J., K.S., A.W., V.H.-V., R.H. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by New Zealand Ministry of Business, Innovation & Employment, to AgResearch, grant number C10X1804.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in FigShare at https://figshare.com/articles/dataset/dx_doi_org_10_6084_m9_figshare_25909780/25909780 (accessed on 28 May 2024).

Conflicts of Interest: Author Kritarth Seth, Shengjing Shi were employed by the company AgResearch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Hardarson, G.; Danso, S. Methods for measuring biological nitrogen fixation in grain legumes. *Plant Soil* **1993**, *152*, 19–23.
- Unkovich, M.; Baldock, J. Measurement of symbiotic N₂ fixation in Australian agriculture. *Soil Biol. Biochem.* **2008**, *40*, 2915–2921.
- Alonso, J.; Gonzalez, P. *Isotope Dilution Mass Spectrometry*; Royal Society of Chemistry: London, UK, 2019.
- Saletnik, A.; Saletnik, B.; Puchalski, C. Overview of popular techniques of Raman spectroscopy and their potential in the study of plant tissues. *Molecules* **2021**, *26*, 1537.
- Zavafer, A.; Ball, M.C. Good vibrations: Raman spectroscopy enables insights into plant biochemical composition. *Funct. Plant Biol.* **2022**, *50*, 1–16.
- Sharma, S.; Bharti, A.; Singh, R.; Uttam, K.N. Non-destructive, label free evaluation of the biochemical profile associated with the growth and ripening process of jamun fruit by confocal micro Raman spectroscopy. *Anal. Lett.* **2022**, *55*, 812–827.
- Schulz, H. and M. Baranska, Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* **2007**, *43*, 13–25.
- De Moliner, F.; Knox, K.; Gordon, D.; Lee, M.; Tipping, W.J.; Geddis, A.; Reinders, A.; Ward, J.M.; Oparka, K.; Vendrell, M. A Palette of Minimally Tagged Sucrose Analogues for Real-Time Raman Imaging of Intracellular Plant Metabolism. *Angew. Chem.* **2021**, *133*, 7715–7720.
- Gersony, J.T.; McClelland, A.; Holbrook, N.M. Raman spectroscopy reveals high phloem sugar content in leaves of canopy red oak trees. *New Phytol.* **2021**, *232*, 418–424.
- Dina, N.E.; Muntean, C.M.; Bratu, I.; Tican, A.; Halmagyi, A.; Purcaru, M.A.P.; Coste, A. Structure and surface dynamics of genomic DNA as probed with surface-enhanced Raman spectroscopy: Trace level sensing of nucleic acids extracted from plants. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *279*, 121477.
- Li, M. Developing Raman Sorting and Imaging Techniques to Characterise and Differentiate Unculturable Microbial Single Cells. Ph.D. Thesis, University of Sheffield, Sheffield, UK, 2012.
- Polisetti, S.; Baig, N.; Bible, A.; Morrell-Falvey, J.; Doktycz, M.; Bohn, P.W. Using Raman spectroscopy and SERS for in situ studies of rhizosphere bacteria. In *Biosensing and Nanomedicine VIII*; SPIE: Bellingham, DC, USA, 2015.
- Jochum, T.; Fastnacht, A.; Trumbore, S.E.; Popp, J.; Frosch, T. Direct Raman spectroscopic measurements of biological nitrogen fixation under natural conditions: An analytical approach for studying nitrogenase activity. *Anal. Chem.* **2017**, *89*, 1117–1122.
- Agarwal, U.P.; Atalla, R.H. In-situ Raman microprobe studies of plant cell walls: Macromolecular organization and compositional variability in the secondary wall of *Picea mariana* (Mill.) BSP. *Planta* **1986**, *169*, 325–332.

15. Saito, S. Field evaluation of the symbiotic fixation abilities among *Rhizobium phaseoli* strains [Rhizobium; Symbiotic fixation of nitrogen]. *Pesqui. Agropecu. Bras.* **1982**, *17*, 999–1006.
16. Farid, M. Symbiotic Nitrogen Fixation in Common Bean. Ph.D. Thesis, University of Guelph, Guelph, ON, Canada, 2015.
17. Akter, Z.; Pageni, B.B.; Lupwayi, N.Z.; Balasubramanian, P.M. Biological nitrogen fixation by irrigated dry bean (*Phaseolus vulgaris* L.) genotypes. *Can. J. Plant Sci.* **2018**, *98*, 1159–1167.
18. Reinprecht, Y.; Schram, L.; Marsolais, F.; Smith, T.H.; Hill, B.; Pauls, K.P. Effects of nitrogen application on nitrogen fixation in common bean production. *Front. Plant Sci.* **2020**, *11*, 1172.
19. Friel, C.A.; Friesen, M.L. Legumes modulate allocation to rhizobial nitrogen fixation in response to factorial light and nitrogen manipulation. *Front. Plant Sci.* **2019**, *10*, 481099.
20. McKnight, T. Efficiency of isolates of *Rhizobium* in the cowpea (*Vigna unguiculata*) group, with proposed additions to this group. *Qld. J. Agric. Sci.* **1949**, *6*, 61–76.
21. Unkovich, M.; Herridge DA VI, 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, Australia, 2008.
22. Kerr, L.T.; Byrne, H.J.; Hennelly, B.M. Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. *Anal. Methods* **2015**, *7*, 5041–5052.
23. Lieber, C.A.; Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* **2003**, *57*, 1363–1367.
24. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
25. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320.
26. Synytsya, A.; Čopíková, J.; Matějka, P.; Machovič, V. Fourier transform Raman and infrared spectroscopy of pectins. *Carbohydr. Polym.* **2003**, *54*, 97–106.
27. Qin, J.; Chao, K.; Kim, M.S. Nondestructive evaluation of internal maturity of tomatoes using spatially offset Raman spectroscopy. *Postharvest Biol. Technol.* **2012**, *71*, 21–31.
28. Pezzotti, G. Raman spectroscopy in cell biology and microbiology. *J. Raman Spectrosc.* **2021**, *52*, 2348–2443.
29. Brentin, R.P. Soy-based chemicals and materials: Growing the value chain. In *Soy-Based Chemicals and Materials*; ACS Publications: Washington, DC, USA, 2014; pp. 1–23.
30. Almeida, M.R.; Alves, R.S.; Nascimbem, L.B.L.R.; Stephani, R.; Poppi, R.J.; de Oliveira, L.F.C. Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis. *Anal. Bioanal. Chem.* **2010**, *397*, 2693–2701.
31. Agarwal, U.P. 1064 nm FT-Raman spectroscopy for investigations of plant cell walls and other biomass materials. *Front. Plant Sci.* **2014**, *5*, 103972.
32. Madzharova, F.; Heiner, Z.; Gühlke, M.; Kneipp, J. Surface-enhanced hyper-Raman spectra of adenine, guanine, cytosine, thymine, and uracil. *J. Phys. Chem. C* **2016**, *120*, 15415–15423.
33. Zhu, G.; Zhu, X.; Fan, Q.; Wan, X. Raman spectra of amino acids and their aqueous solutions. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2011**, *78*, 1187–1195.
34. Kundu, J.; Neumann, O.; Janesko, B.G.; Zhang, D.; Lal, S.; Barhoumi, A.; Scuseria, G.E.; Halas, N.J. Adenine- and adenosine monophosphate (AMP)—Gold binding interactions studied by surface-enhanced Raman and infrared spectroscopies. *J. Phys. Chem. C* **2009**, *113*, 14390–14397.
35. Schenzel, K.; Fischer, S. NIR FT Raman spectroscopy—a rapid analytical tool for detecting the transformation of cellulose polymorphs. *Cellulose* **2001**, *8*, 49–57.
36. Zhbakov, R.; Firsov, S.; Buslov, D.; Nikonenko, N.; Marchewka, M.; Ratajczak, H. Structural physico-chemistry of cellulose macromolecules. Vibrational spectra and structure of cellulose. *J. Mol. Struct.* **2002**, *614*, 117–125.
37. da Silva, K.J.; Paschoal, W., Jr.; Belo, E.A.; Moreira, S.G. Phase transition in *all-trans*- β -carotene crystal: Temperature-dependent Raman spectra. *J. Phys. Chem. A* **2015**, *119*, 9778–9784.
38. Grudzinski, W.; Janik, E.; Bednarska, J.; Welc, R.; Zubik, M.; Sowinski, K.; Luchowski, R.; Gruszecki, W.I. Light-driven reconfiguration of a Xanthophyll Violaxanthin in the photosynthetic pigment–protein complex LHCII: A resonance Raman study. *J. Phys. Chem. B* **2016**, *120*, 4373–4382.
39. Gruszecki, W.I.; Gospodarek, M.; Grudzinski, W.; Mazur, R.; Gieczewska, K.; Garstka, M. Light-induced change of configuration of the LHCII-bound xanthophyll (tentatively assigned to violaxanthin): A resonance Raman study. *J. Phys. Chem. B* **2009**, *113*, 2506–2512.
40. Goldflus, F.; Ceccantini, M.; Santos, W. Amino acid content of soybean samples collected in different Brazilian states: Harvest 2003/2004. *Braz. J. Poult. Sci.* **2006**, *8*, 105–111.
41. Flores-Tinoco, C.E.; Tschan, F.; Fuhrer, T.; Margot, C.; Sauer, U.; Christen, M.; Christen, B. Co-catabolism of arginine and succinate drives symbiotic nitrogen fixation. *Mol. Syst. Biol.* **2020**, *16*, e9419.
42. Jorge, C.J.F. Comparative Analyses of the Nitrogen Fixing Potential of Different Legume Species Grown under Different Agronomic Practices. Ph.D. Thesis, North-West University, Potchefstroom, South Africa, 2017.

43. Voisin, A.S.; Salon, C.; Delfosse, O.; Mary, B. Theoretical evaluation of ^{15}N isotopic methods for measuring symbiotic nitrogen fixation in the field. In Proceedings of the Nitrogen Fixation Research Progress: Proceedings of the 6th International Symposium on Nitrogen Fixation, Corvallis, OR, USA, 4–10 August 1985; Nova Science Publishers Inc.: Hauppauge, NY, USA, 2008.
44. Zhang, H.; Plett, J.M.; Catunda, K.L.M.; Churchill, A.C.; Moore, B.D.; Powell, J.R.; A Power, S.; Yang, J.; Anderson, I.C. Rapid quantification of biological nitrogen fixation using optical spectroscopy. *J. Exp. Bot.* **2024**, *75*, 760–771.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.