

CAPITAL UNIVERSITY OF SCIENCE AND  
TECHNOLOGY, ISLAMABAD



Genome-Wide Identification and  
Characterization of NLP Genes  
in *Trifolium pratense* Red Clover

by

Hafsa

A thesis submitted in partial fulfillment for the  
degree of Master of Science

in the

Faculty of Health and Life Sciences

Department of Bioinformatics and Biosciences

2025

Copyright © 2025 by Hafsa

All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, by any information storage and retrieval system without the prior written permission of the author.

*I dedicate this work with deepest love and gratitude to my beloved parents whose unwavering support, sacrifices, and endless encouragement have been my guiding light.*

*Your boundless faith in me, even in moments of doubt, gave me the strength to persevere. This achievement is as much yours as it is mine.*

*Thank you for being my foundation, my inspiration, and my greatest blessings.*

*"A parent's love is the fuel that enables a normal human being to do the impossible."*



## CERTIFICATE OF APPROVAL

### Genome-Wide Identification and Characterization of NLP Genes in *Trifolium pratense* Red Clover

by

Hafsa

(MBS233014)

### THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Syed Afzal Shah	NUMS, Rawalpindi
(b)	Internal Examiner	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad
(c)	Supervisor	Dr. Sami Ullah Jan	CUST, Islamabad

---

Dr. Sami Ullah Jan

Thesis Supervisor

September, 2025

---

Dr. Syeda Marriam Bakhtiar  
Head  
Dept. of Bioinfo. & Biosciences  
September, 2025

---

Dr. Sahar Fazal  
Dean  
Faculty of Health & Life Sciences  
September, 2025

## *Author's Declaration*

I, **Hafsa** hereby state that my MS thesis titled “**Genome-Wide Identification and Characterization of NLP Genes in *Trifolium pratense* Red Clover**” is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.



**(Hafsa)**

Registration No: MBS233014

---

## *Plagiarism Undertaking*

I solemnly declare that research work presented in this thesis titled “**Genome-Wide Identification and Characterization of NLP Genes in *Trifolium pratense* Red Clover**” is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.



(Hafsa)

Registration No: MBS233014

## *Acknowledgement*

All the praises and thanks be to Allah, Who has guided me to this, and never could have been found guidance, where it not but Allah had guided me. In times of darkness, You, Allah are my guiding light. Thank you for illuminating my path.

I would like to express my deepest gratitude to the following individuals without whom this thesis would not have been possible: My supervisor, Sir Sami Ullah Jan, for their unwavering guidance, invaluable insights, and continuous support throughout the entire research process. Their expertise and encouragement have been instrumental in shaping the direction of this study. I am grateful for their valuable feedback and constructive criticism, which greatly contributed to the improvement of this thesis. I am thankful for his assistance and resources, which played a crucial role in the successful completion of this research.

I also acknowledge the support of Dean Dr. Sahar Fazal and Dr. S. Marriam Bakhtia Head of Department, for their leadership and for providing the excellent research facilities that enabled this bioinformatics study.

I extend my heartfelt thanks to **CUST, Islamabad** for providing me with the necessary resources and support. I am also grateful to my professors for their constructive suggestions and motivation.

Special thanks to my family for their unwavering support, understanding, and encouragement during challenging times. Their belief in me kept me motivated.

Lastly, I acknowledge all the individuals who directly or indirectly contributed to this endeavor. This accomplishment would not have been possible without their help.

(Hafsa)

---

## Abstract

Plants utilize nitrogen as a vital macronutrient precursor for producing proteins and chlorophyll. Nitrogen is externally applied to agricultural soil in order to achieve maximum productivity. However, plants with low Nitrogen Use Efficiency (NUE) loses the applied nitrogen making economic losses and environmental pollution. Therefore, plant NUE must be improved for sustainable agricultural practices with reduced economic losses and environmental pollution. The most feasible way to improve NUE is through genetic engineering in plants. By identifying and manipulating the NUE response genes and transcription factors is promising in improving plant NUE. This study aimed to identify NODULE INCEPTION-LIKE PROTEIN (NLP) gene family in *Trifolium pratense*, or red clover through genome wide analysis. A total of six NLP genes (*TpNLP1–TpNLP6*) were found having conserved RWP-RK and PB1 domains that are important for DNA binding and protein interactions. Phylogenetic study showed that *TpNLPs* and NLP genes from different plants, like *Arabidopsis thaliana*, rice, and maize, are the homologues originated from common ancestor. Physicochemical parametric analysis demonstrated that *TpNLPs* are similar to their *Arabidopsis* counterparts in terms of molecular weight, isoelectric point, and localization. Predictions of protein-protein interactions showed that *TpNLPs* interact with genes that help with nitrogen uptake, transport, and assimilation, which supports their significance in NUE. Also, promoter analysis found cis-regulatory elements that are related to phytohormones, stress responses, and plant growth. Overall, based on the similarities found in structural and functional attributes of both *AtNLPs* and *TpNLPs*, it is proposed that *TpNLPs* can be effectively utilized as potential candidate genes in genetic engineering for improving NUE within red clover and beyond the species. However, future wet-lab studies should be conducted in order to verify the findings from the current *In Silico* analysis. Improving NUE in red clover could make farmers less reliant on fertilizers and make farming more sustainable, which would be beneficial for both crop production and the environment.

**Keywords:** Nitrogen, Nitrogen Use Efficiency (NUE), NIN-LIKE PROTEIN (NLP), Transcription Factor, *Trifolium pratense*, Genome-Wide Analysis.

# Contents

<b>Author’s Declaration</b>	<b>iv</b>
<b>Plagiarism Undertaking</b>	<b>v</b>
<b>Acknowledgement</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xii</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Nitrogen Use Efficiency (NUE) . . . . .	2
1.3 Plant Transcription Factors (TFs) . . . . .	3
1.4 Problem Statement . . . . .	6
1.5 Aim and Objectives . . . . .	6
1.6 Scope of Study . . . . .	7
<b>2 Literature Review</b>	<b>8</b>
2.1 <i>Trifolium pratense</i> . . . . .	8
2.1.1 Morphology . . . . .	8
2.2 Distribution . . . . .	9
2.3 The Significance of <i>Trifolium pratense</i> . . . . .	10
2.3.1 Medicinal Value of <i>Trifolium pratense</i> . . . . .	10
2.3.1.1 Phytoestrogen Activity . . . . .	10
2.3.1.2 Anti-inflammatory Support . . . . .	11
2.3.1.3 Respiratory Comfort . . . . .	11
2.3.1.4 Dermatological Benefits . . . . .	11
2.3.1.5 Heart Health Support . . . . .	12
2.3.1.6 Boosts Immune System Function . . . . .	12
2.3.1.7 Reduces Stress and Anxiety . . . . .	12

---

2.3.1.8	Encourages Proper Digestion . . . . .	12
2.3.1.9	Supports the Body's Natural Detoxification Processes . . . . .	13
2.3.1.10	Bone Health . . . . .	13
2.3.1.11	Potential Anti-Cancer Effects . . . . .	13
2.3.2	Agriculture Value of <i>Trifolium pratense</i> . . . . .	14
2.3.3	Nutritional Value of <i>Trifolium pratense</i> . . . . .	14
2.3.4	Ecological Value of <i>Trifolium pratense</i> . . . . .	15
2.4	Role of Nitrogen in Plant Growth and Development . . . . .	16
2.4.1	Nitrogen Helps Build Proteins and Genetic Material . . . . .	16
2.4.2	Nitrogen Fuels Photosynthesis by Supporting Chlorophyll Production . . . . .	17
2.4.3	Nitrogen in Enzyme Activation and Cofactor Synthesis . . . . .	17
2.4.4	Nitrogen and Secondary Metabolite Production . . . . .	17
2.4.5	Nitrogen Drives Plant Growth and Development . . . . .	18
2.4.6	Nitrogen Helps Plants Cope with Stress . . . . .	18
2.4.7	Nitrogen Improves Nutrient Uptake Through Healthy Roots . . . . .	18
2.4.8	Plants Depend on Nitrogen from Soil or Symbiotic Relationships . . . . .	19
2.4.9	Nitrogen Deficiency can Severely Impact Plants . . . . .	19
2.5	Nitrogen Regulation in Plants . . . . .	19
2.6	Role of Transporters & Transcription Factors in Nitrogen Uptake by Roots . . . . .	23
2.7	Impact of Nitrogen Use Efficiency (NUE) . . . . .	24
2.8	Enhancement of Nitrogen Use Efficiency . . . . .	25
2.8.1	Precise Farming . . . . .	25
2.8.2	Optimizing Fertilizer Management . . . . .	25
2.8.3	Crop Rotation and Cover Crops . . . . .	26
2.8.4	Nitrogen-Fixing Plants . . . . .	26
2.8.5	Genetic Approaches . . . . .	26
2.8.6	Soil Health Improvement . . . . .	26
2.8.7	Precision Irrigation . . . . .	26
2.9	Transcription Factors . . . . .	27
2.10	Role of Transcription Factors . . . . .	27
2.11	Genes as Transcription Factor in Regulating Nitrogen Response . . . . .	29
2.12	<i>NLP</i> Gene Role in Nitrogen Signaling . . . . .	29
2.13	Study of the <i>NLP</i> Gene Family in <i>Arabidopsis thaliana</i> . . . . .	31
2.14	The Investigation of the <i>NLP</i> Gene family in <i>B. napus</i> . . . . .	32
2.15	The Investigation of the <i>NLP</i> Gene Family in Tomatoes . . . . .	32
2.16	Key Findings on the <i>NLP</i> Gene Family in Rice . . . . .	33
2.17	Study of <i>NLP</i> Genes in Maize . . . . .	33
2.18	Key Findings on the <i>NLP</i> Gene Family in Wheat . . . . .	34
2.19	Key Findings on the <i>NLP</i> Gene Family in Tea plant . . . . .	35
2.20	Key Findings on the <i>NLP</i> Gene Family in Alfalfa . . . . .	36

---

<b>3</b>	<b>Methodology</b>	<b>37</b>
3.1	Retrieval of Reference NLP Sequences from Arabidopsis . . . . .	38
3.2	Identification of NLP Genes in <i>Trifolium pratense</i> . . . . .	38
3.3	Redundancy Elimination and Sequence Curation . . . . .	38
3.4	Labelling & Sequence Similarity Assessment . . . . .	39
3.5	Physicochemical and Domain Characterization . . . . .	39
3.6	Conserved Motif Discovery . . . . .	39
3.7	Phylogenetic Analysis . . . . .	40
3.8	Gene Structure and Motif Mapping . . . . .	40
3.9	Protein-Protein Interaction Prediction . . . . .	40
3.10	Putative cis-Regulatory Elements Identification in <i>TpNLPs</i> Homologues . . . . .	41
<b>4</b>	<b>Results</b>	<b>42</b>
4.1	Identification of <i>NLP</i> genes in <i>Trifolium pratense</i> . . . . .	42
4.2	Conserved Domain Identification . . . . .	43
4.3	Physiochemical Properties . . . . .	44
4.4	Conserved Motif Discovery . . . . .	46
4.5	Gene Structure . . . . .	47
4.6	Protein-Protein Interaction . . . . .	48
4.7	The <i>TpNLPs</i> Gene Family's Evolutionary Relationships and Sequence Alignment . . . . .	61
4.8	Cis-regulatory elements . . . . .	63
<b>5</b>	<b>Discussion</b>	<b>65</b>
<b>6</b>	<b>Conclusion and Future Work</b>	<b>68</b>
6.1	Conclusion . . . . .	68
6.2	Future Direction . . . . .	69
	<b>Bibliography</b>	<b>71</b>

# List of Figures

2.1	<i>Trifolium pratense</i> (Red clover) plant . . . . .	9
2.2	Nitrogen's function in plant development and growth . . . . .	16
2.3	Nitrogen uptake is regulated by several processes, and the picture shows the main root nitrogen transporters and the metabolic interactions between them. Although it is not the usual in reality, for the purpose of simplicity it is stated that receptors, H <sup>+</sup> -ATPase, transporters, and aquaporin's are all present in the same root cell . . . . .	22
3.1	Research Methodology . . . . .	37
4.1	Logos of identified motifs . . . . .	46
4.2	Consensus motifs of <i>AtNLPs</i> and <i>TpNLPs</i> . . . . .	47
4.3	Display the result of Gene structure determination . . . . .	48
4.4	Protein Protein interaction of <i>AtNLP1</i> . . . . .	48
4.5	Protein Protein interaction of <i>AtNLP2</i> . . . . .	49
4.6	Protein Protein interaction of <i>AtNLP3</i> . . . . .	50
4.7	Protein Protein interaction of <i>AtNLP4</i> . . . . .	51
4.8	Protein Protein interaction of <i>AtNLP5</i> . . . . .	52
4.9	Protein Protein interaction of <i>AtNLP6</i> . . . . .	52
4.10	Protein Protein interaction of <i>AtNLP7</i> . . . . .	53
4.11	Protein Protein interaction of <i>AtNLP8</i> . . . . .	54
4.12	Protein Protein interaction of <i>AtNLP9</i> . . . . .	55
4.13	Protein Protein interaction of <i>TpNLP1</i> . . . . .	56
4.14	Protein Protein interaction of <i>TpNLP2</i> . . . . .	57
4.15	Protein Protein interaction of <i>TpNLP3</i> . . . . .	58
4.16	Protein Protein interaction of <i>TpNLP4</i> . . . . .	58
4.17	Protein Protein interaction of <i>TpNLP5</i> . . . . .	59
4.18	Protein Protein interaction of <i>TpNLP6</i> . . . . .	60
4.19	Sequence similarity between <i>Trifolium pratense</i> and <i>Arabidopsis thaliana</i> NLP proteins (similarity greater than 60% is indicated with red font color). . . . .	62
4.20	Phylogenetic tree . . . . .	62

# List of Tables

2.1	Taxonomy of <i>Trifolium pratense</i> (Red clover) . . . . .	9
4.1	Conserved domains of NLP gene families in <i>Arabidopsis thaliana</i> and <i>Trifolium pratense</i> . . . . .	43
4.2	Physical and chemical properties of NLP gene families of <i>Arabidopsis thaliana</i> and <i>Trifolium pratense</i> . . . . .	45
4.3	Protein Protein interaction of <i>AtNLP1</i> . . . . .	49
4.4	Protein Protein interaction of <i>AtNLP2</i> . . . . .	49
4.5	Protein Protein interaction of <i>AtNLP3</i> . . . . .	50
4.6	Protein–Protein interaction of <i>AtNLP4</i> . . . . .	51
4.7	Protein Protein interaction of <i>AtNLP5</i> . . . . .	52
4.8	Protein Protein interaction of <i>AtNLP6</i> . . . . .	53
4.9	Protein Protein interaction of <i>AtNLP7</i> . . . . .	54
4.10	Protein Protein interaction of <i>AtNLP8</i> . . . . .	55
4.11	Protein Protein interaction of <i>AtNLP9</i> . . . . .	56
4.12	Protein Protein interaction of <i>TpNLP1</i> . . . . .	56
4.13	Protein–Protein interaction of <i>TpNLP2</i> . . . . .	57
4.14	Protein Protein interaction of <i>TpNLP3</i> . . . . .	58
4.15	Protein Protein interaction of <i>TpNLP4</i> . . . . .	59
4.16	Protein Protein interaction of <i>TpNLP5</i> . . . . .	59
4.17	Protein Protein interaction of <i>TpNLP6</i> . . . . .	60
4.18	Number of cis-regulatory elements identified in promoter regions of AtNLPs and TpNLPs Gene Families. . . . .	64

# Abbreviations

<b>AMT</b>	Ammonium Transporter
<b>CDD</b>	Conserved Domain Database
<b>GRAVY</b>	Grand Average of Hydropathicity
<b>GS/GOGAT</b>	Glutamine Synthetase/Glutamate Synthase
<b>MEME</b>	Multiple Em for Motif Elicitation
<b>NCBI</b>	National Center for Biotechnology Information
<b>NLP</b>	Nodule Inception-Like Protein
<b>NRE</b>	Nitrate Responsive cis-Element
<b>NRT</b>	Nitrate Transporter
<b>NUE</b>	Nitrogen Use Efficiency
<b>PB1</b>	Phox and Bem1 domain
<b>RWP-RK</b>	Arginine-Tryptophan-Proline-rich motif followed by a protein kinase
<b>STRING</b>	Search Tool for Recurring Instances of Neighboring Genes
<b>TAIR</b>	The Arabidopsis Information Resource
<b>TF</b>	Transcription Factor

# Chapter 1

## Introduction

### 1.1 Background

Nitrogen is an important nutrient for plants that helps them grow and develop. It is a necessary constituent for amino acids, which serve as the building blocks of proteins needed to make enzymes that fuel metabolic processes and plant tissues [1]. Nucleic acids, like DNA and RNA, are made up of nitrogen and are needed for cell division and passing on genetic information. Its creation of chlorophyll, the green pigment that drives photosynthesis and allows plants to transform sunlight into chemical energy, is one of its most obvious functions. Furthermore, nitrogen helps plants transfer energy through substances like ATP (adenosine triphosphate). Plants usually show signs of not having enough nitrogen, like older leaves turning yellow, slow growth, and less fruit or vegetable production. In general, plants need nitrogen to have healthy green growth, a healthy metabolism, and a lot of food production [1].

The atmosphere of Earth is 78% nitrogen. Nitrogen is used by plants as nitrate and ammonium from the soil, even though it is abundant in the atmosphere. Only 30–40% of nitrogen provided is actually used by crops; the remainder is lost to environmental processes such as runoff, volatilization, and leaching. In addition to lowering crop yields, this inefficiency adds to pollution in the environment [2].

Therefore, improving nitrogen usage efficiency (NUE) is a major goal in contemporary agriculture. Particularly in well-aerated soil, plants have evolved efficient methods for absorbing and metabolizing nitrogen. Plants primarily take nitrogen in the form of nitrate, which is primarily carried by nitrate transporters. Nitrate reductase and nitrite reductase convert the nitrate that cells absorb into ammonium. Glutamate synthetase and glutamine synthetase then aid in the absorption of ammonium into amino acids [3].

Legumes and rhizobial bacteria, which are housed in root nodules, have a symbiotic interaction that allows them to fix atmospheric nitrogen. This mechanism converts inert N<sub>2</sub> gas into usable nitrogen, ammonia, which helps them reduce their reliance on synthetic fertilizers and soil nitrogen [4]. NIN is one of the key transcription factors that regulates this symbiotic connection (Nodule Inception). Because the fixed nitrogen helps plants grow and store seed protein, and the retained nitrogen makes the soil better for future crops, legumes are crucial to sustainable agriculture [5].

## 1.2 Nitrogen Use Efficiency (NUE)

The Nitrogen Use Efficiency (NUE) is a measurement that indicates how efficiently plants make use of the nitrogen that they receive. Because of this, the amount of food that can be cultivated and the ecological health of the area are both directly influenced [6]. It is the amount of crop yield that is produced for each unit of nitrogen that is utilized in the production of the crop. Specifically, it is important to consider how well it is able to absorb nitrogen as well as how well it is able to use nitrogen. A plant's ability to absorb nitrogen from the soil is referred to as its uptake efficiency, while its utilization efficiency refers to how effectively the plant is able to use the nitrogen it has taken in to grow and produce food [7]. During the agricultural process, nitrogen is transferred through the soil, the crops, and the animals; however, this natural process is frequently disrupted by factors such as leaching or the removal of crops during harvest [8]. In order to compensate for this, farmers utilize synthetic fertilizers; hence, it is essential for both economic

and environmental reasons to make optimal use of nitrogen. Improving NUE is a good way to cut down on nitrogen waste and its effects on the environment [9]. A little increase in efficiency of just 1% could save the globe about \$1.1 billion a year and increase food yield.

The usual way to figure out NUE is:

$$NUE(\%) = \frac{(Total\ N\ uptake\ in\ fertilized\ plot - Total\ N\ uptake\ in\ control\ plot)}{Amount\ of\ N\ applied} \times 100 [6]$$

Increasing NUE is beneficial to agriculture in a number of different ways, including increased yields, decreased costs for fertilizer, and a reduced impact on the environment [7]. By increasing the efficiency with which plants use nitrogen, they reduce the amount of nitrogen that is lost to the environment or waterways, hence reducing emissions of greenhouse gases and preventing water contamination [9]. The optimization of net utilization efficiency (NUE) allows farmers to achieve sustainable output, reduce the costs of inputs, and contribute to the improvement of ecosystems.

### 1.3 Plant Transcription Factors (TFs)

Several transcription factors are very important for controlling the expression of genes that are involved in taking up nitrates, using them, and other metabolic processes that are important for plants to send nitrogen signals quickly [42]. These transcription factors act as key regulators, enabling plants to fine-tune gene expression in response to fluctuating soil nitrate levels [44]. By orchestrating this adaptive response, they help plants optimize nitrogen absorption and utilization, ensuring better growth and resource efficiency under varying environmental conditions [46]. This sophisticated regulatory mechanism highlights how plants dynamically adjust their nitrogen metabolism to maximize nutrient uptake while minimizing waste [48].

The NLP (NIN-like protein) family is very important for controlling genes that are involved in nitrogen metabolism and signaling nitrate in plants [46]. Researchers

first found these transcription factors, called NIN-like proteins (NLPs), because they helped legume nodules form [4]. Later, they found that they also had a role in nitrate signaling pathways [47]. NLPs have a conserved RWP-RK DNA-binding domain that lets them bind to certain promoter regions of target genes [50]. Because NLPs can bind to things, they can control gene expression very accurately in response to the availability of nitrate and other environmental signals [52]. NLPs directly affect the expression of genes that are responsible for taking in, moving, and using nitrates through this method.

The molecular basis of nitrogen uptake, transport, and metabolism in plants involves the coordinated action of multiple transcription factors and gene families [37]. Nitrate transporters such as NRT1 and NRT2, along with associated channels including chloride channels and slow anion-channel-related homologues, mediate the absorption of inorganic nitrate ( $\text{NO}_3^-$ ) from soil and its transport into plant cells [39]. Inside the plant, nitrate is changed into ammonium ( $\text{NH}_4^+$ ) by the successive actions of nitrate reductases (NIA1, NIA2) and nitrite reductase (NiR) [28]. The ammonium is then incorporated into organic amino acids like glutamate and glutamine via enzymes including GS and GOGAT [2]. These amino acids serve as essential nitrogen donors for the synthesis of critical macromolecules such as nucleic acids, proteins, and chlorophyll.

Beyond their metabolic functions, nitrate and amino acids also act as signaling molecules that influence various physiological processes through their interaction with specific transcription factors [34]. These nitrogen-responsive regulatory networks ultimately determine plant structure, function, and overall nitrogen use efficiency (NUE) [38]. Interestingly, studies in *Chlamydomonas reinhardtii* have revealed that nitrogen limitation triggers the MID (minus dominant) protein, which controls the differentiation of vegetative cells into gametes, demonstrating the broad significance of nitrogen-responsive regulatory mechanisms across different organisms [45]. This protein either activates or deactivates the minus or plus gametes differentiation pathway in response to nitrogen cues [45]. After the leucine zipper motif, MID has a conserved sequence termed RW-PYRK. This sequence was first discovered to be the first member of a new TF family called the RWP-RK

gene family [46]. Plants include all vascular plants, green algae, slime molds, and RWP-RK gene family members [47]. Later, the leguminous plant *Lotus japonicas* was found to harbor the first nodule inception gene (NIN), which has a RWP-RK domain that regulates nitrogen-mediated symbiotic root-nodule development [4]. Additional NIN proteins and Researchers have detected NLPs (Nodule-Inducing Proteins) in other plants that aren't legumes, like *Arabidopsis*, rice, wheat, and maize. Animals, on the other hand, don't have this property [47].

After extensive investigation, two sub-families of the RWP-RK gene family were discovered: (a) RKD, or the RWP-RK domain-containing gene family, and (b) NLP, or the RKD with an extra domain at the C-terminus known as Phox and Beml (PB1) [46]. Proteins that have been shown to share structural characteristics with NIN genes are known as NIN-Like Proteins (NLPs). Leguminous and non-leguminous plants both contain NLPs, but only legumes have NIN [4]. NLPs are transcription factors particular to plants that play a key role in controlling the nitrogen response. The N-terminal region of NLPs is responsible for transcriptional activation of genes, while the PB1 domain (PFOOS64) mediates protein-protein interactions and RWP-RK (PF02042) makes DNA binding [50]. The binding of NLPs to the nitrate-responsive cis-element (NRE) in the promoter region activates the transcription of nitrate-regulated genes [53].

The NLP gene family is a reliable regulator of N-responsive genes, therefore NUE might get better [49]. Thus far, genome-wide research has identified that *Arabidopsis thaliana* contain nine NLP genes, rice contain six, maize nine, thirty-one in *Brassica napus*, eighteen in wheat, and nine in *Arabidopsis thaliana* [50], [52]. However, no comparable studies on NLPs in red clover have been published to far. Researchers have been studying vascular plants to fully understand the structure and function of NLP genes that can boost NUE since they were first discovered [46].

The vascular plant known as Red clover (*Trifolium pratense*), a perennial herbaceous legume that is extensively found in temperate climates, is the subject of our study [16]. Trifoliate leaves with three oval or heart-shaped leaflets, frequently adorned with a red crescent, are the hallmark of low-growing, creeping red clover

plants that form dense mats [17]. Large areas can be colonized by the plant because of its alternating leaf arrangement on the stem and stolon, or above-ground runners. Pollinators like bees are drawn to the tiny, fragrant, red or purple-tinged flowers of red clover, which are grouped in spherical inflorescences (flower heads). [24]

Red clover leaves and petals have a mild, somewhat sweet flavor and are edible and nutritious. They can be eaten fresh, boiled, or brewed into tea [21]. It promotes soil health and sustainable diets since it is high in calcium, magnesium, fiber, protein, and vitamin C [19]. Because of its resilience and capacity to fix nitrogen, it is perfect for permaculture, organic farming, and erosion management. White clover is a multipurpose plant in agriculture, ecology, and nutrition that may help with digestion, heart health, and antioxidant consumption [22].

Legumes are unique in that they can fix atmospheric nitrogen by forming symbiotic relationships with rhizobia bacteria. This procedure lessens reliance on artificial fertilizers while also improving soil fertility [5].

## 1.4 Problem Statement

Although the NLP gene family has been thoroughly explored in a number of plants, little is known about how it functions in *Trifolium pratense*, or red clover. Through the inspecting and characterization of the NLP gene family in red clover, our study seeks to close this gap.

## 1.5 Aim and Objectives

The goal of this study is to find and describe the NLP gene family in *Trifolium pratense* utilizing computer-based methods.

- To identify NLP genes in red clover (*Trifolium pratense*).

- To assess the evolutionary relationships of *TpNLP* genes with other plants.
- To compare the structural and functional attributes of *TpNLPs* with *AtNLPs*.

## 1.6 Scope of Study

Red clover (*Trifolium pratense*) presents a remarkably broad scope of study spanning agricultural, ecological, nutritional, and environmental disciplines [16], with researchers investigating its nitrogen-fixing capabilities that enhance soil fertility and reduce fertilizer dependence in sustainable farming systems, its value as high-protein forage for livestock, and its role in supporting pollinator populations through its nectar-rich flowers, while ecologists examine its effectiveness in erosion control, weed suppression, and land restoration, nutritionists explore its edible leaves and flowers as potential functional foods rich in protein, vitamins, and antioxidants with possible medicinal applications [21], [22], urban planners consider its use in low-maintenance, drought-resistant landscaping, and geneticists work to improve its stress tolerance and productivity, making this humble plant a critical focus for advancing sustainable agriculture, biodiversity conservation, resilience of the environment to climate change and food security [7], [19].

# Chapter 2

## Literature Review

### 2.1 *Trifolium pratense*

#### 2.1.1 Morphology

*Trifolium pratense* L, sometimes known as red clover, is a perennial plant that looks nice and has a unique, structured shape. The plant's roots include a deep taproot that is both wide and strong [17]. These roots create special nodules that add important nutrients to the soil by working together with nitrogen-fixing bacteria. This quality not only makes red clover a pretty plant, but it also helps farmers [24].

Red clover stems are robust, but they have a slight hairiness and can grow either straight up or with a delicate lean. They can grow to a height of 8 to 31 inches and frequently have multiple branches that originate from the base. Its leaves are likely the most distinctive feature, consisting of three oval leaflets that are arranged alternately along the stem [17]. A subtle, pale marking that resembles a crescent moon is often present on each leaflet, which contributes to its delicate uniqueness [19]. The delicate yet practical character of the foliage is further emphasized by the small, thin stipules that encircle the stem at the base of each leaf [17].

The gorgeous, spherical flower heads that red clover produces when it blooms are bursting with pinkish-purple hues [21]. There are numerous tiny, pea-shaped flowers packed closely together in each fluffy cluster [22]. These flowers have a significant role in traditional herbal medicine in addition to being a popular attraction for bees and butterflies. The flowers turn into tiny pods after pollination, each of which contains a single seed, ensuring the plant species' survival [26].



FIGURE 2.1: *Trifolium pratense* (Red clover) plant [15]

TABLE 2.1: Taxonomy of *Trifolium pratense* (Red clover) [16]

Classification	
Kingdom:	<i>Plantae</i>
Phylum:	<i>Tracheophyta</i>
Class:	<i>Magnoliopsida</i>
Order:	<i>Fabales</i>
Family:	<i>Fabaceae</i>
Genus:	<i>Trifolium</i>
Species:	<i>pratense</i>

## 2.2 Distribution

Although the red clover plant is native to Europe and Central Asia, it has spread around the world due to its adaptability, which has led to its extensive distribution. It is able to thrive in a variety of situations, including partially shaded regions and sunny lawns, and it is also able to survive in soil that is not ideal [19].

Particularly in Pakistan's colder northern regions, red clover (*Trifolium pratense*) has established a modest but significant role in the country's agricultural landscape. Although this hardy plant is not indigenous to this region, it has been successfully planted in climate-suitable places and is mostly used as wholesome animal feed . It is primarily seen flourishing in Khyber Pakhtunkhwa's foggy highlands, such as Swat, Murree, and Mansehra, where the soil drains effectively and the temperatures remain comparatively low. While some elevated regions of Baluchistan, such as Quetta, May occasionally grow it, it is less prevalent there. The Pothohar Plateau, which surrounds Islamabad and Rawalpindi, also supports some growth [18].

## 2.3 The Significance of *Trifolium pratense*

A little but potent leguminous plant, red clover (*Trifolium pratense*) is extremely important in many fields, especially agriculture, medicine, and human nutrition. It is a useful tool in initiatives for sustainability and health promotion because of its versatility, advantages for the environment, and rich phytochemical profile.

### 2.3.1 Medicinal Value of *Trifolium pratense*

Traditional medicine has used *Trifolium pratense* for a long time. It is known that the plant has a lot of bioactive chemicals in it, including as flavonoids, phenolic acids, tannins, isoflavones, and saponins. These phytochemicals make the plant very good at fighting inflammation and free radicals, which helps the body stay healthy [19].

#### 2.3.1.1 Phytoestrogen Activity

As a result of isoflavones including genistein, daidzein, biochanin A, and formononetin, red clover is especially well-known for its phytoestrogen qualities. These substances can attach to estrogen receptors in the body and structurally

resemble human estrogen. Because of this, red clover is frequently used to treat menopausal symptoms like mood swings, hot flashes, and night sweats. Its estrogen-like properties may also lower the risk of malignancies that are hormone-dependent, such as prostate and breast cancers [21].

### **2.3.1.2 Anti-inflammatory Support**

Red clover contains compounds that are believed to have anti-inflammatory properties. As a result, it could be a sensible option for encouraging inflammatory conditions. Whether used topically as a poultice or consumed as an herbal tea, it may help lessen pain related to conditions like arthritis, joint discomfort, and muscular soreness. Its potential analgesic properties offer a gentle alternative to conventional analgesics [22].

### **2.3.1.3 Respiratory Comfort**

Throughout history, red clover has been utilized to improve several aspects of respiratory health. Those who feel it can assist alleviate the symptoms of bronchitis, colds, and coughs are using it. Because of its expectorant properties, it may be able to break up mucus and phlegm, which will make it easier to exit the body. In addition to reinforcing the body's defenses, its reported antibacterial capabilities may also be beneficial in the fight against respiratory infections [22].

### **2.3.1.4 Dermatological Benefits**

Red clover has been used for a long time to cure several skin problems. It can be administered topically or taken as a tea to treat skin ulcers, psoriasis, acne, and eczema symptoms. It is a common solution in natural skincare techniques because of its anti-inflammatory and detoxifying properties, which encourage skin regeneration and healing [22].

### **2.3.1.5 Heart Health Support**

Red clover has historically been associated with cardiovascular health support and blood purification. It is claimed to have the capacity to raise blood circulation, lower cholesterol, and enhance heart health generally. It is believed that regular consumption of white clover teas or tinctures can promote cardiovascular health [23].

### **2.3.1.6 Boosts Immune System Function**

Red Clover is also known for its immune-boosting effects because it has a lot of antioxidants and other good chemicals. Antioxidants help get rid of free radicals, which are unstable molecules that can hurt cells. This lowers oxidative stress and boosts the immune system. Also, Red Clover can help the immune system work well by getting rid of toxins and metabolic waste that can be weighing down the body's defenses. Adding Red Clover to your overall wellness strategy can help boost and protect your immune system, which will make you healthier and more resilient [23].

### **2.3.1.7 Reduces Stress and Anxiety**

Red clover has long been used to encourage serenity and comfort. It is a fantastic option for lowering tension, stress, and anxiety because of its mild sedative qualities. Red Clover is an excellent complement to an herbal stress-reduction program since it helps to regulate the body's natural rhythms, which can encourage calmness and relaxation [23].

### **2.3.1.8 Encourages Proper Digestion**

Red clover is also good for the digestive tract. Constipation can be lessened and regular bowel motions can be encouraged by its modest laxative action. Furthermore, it is a friendly ally for preserving gut health because of its inherent

anti-inflammatory qualities, which help ease stomach discomfort. Red clover can enhance digestion and general nutrient absorption by supporting a healthy gut environment, which will benefit the body as a whole [23].

#### **2.3.1.9 Supports the Body's Natural Detoxification Processes**

This is one of Red Clover's most notable health advantages. Red clover, a natural blood purifier, supports the liver and lymphatic system, which aids in the body's removal of waste and impurities. The body may operate more effectively as a result of this cleansing action, enhancing general wellbeing.

Maintaining health requires detoxification, and Red Clover's mild yet potent cleansing qualities provide a natural means of assisting the body in remaining balanced and energetic.

#### **2.3.1.10 Bone Health**

Red clover's estrogenic properties may assist maintain bone mineral density, especially in postmenopausal women who are more susceptible to osteoporosis. In promoting skeletal health, it is a natural substitute for hormone replacement treatment due to its capacity to slow down bone loss [23].

#### **2.3.1.11 Potential Anti-Cancer Effects**

The results of preliminary scientific research suggest that red clover may have the ability to avoid the effects of chemotherapy. In addition to their potential to modify cell signaling pathways, suppress the proliferation of cancer cells, and cause apoptosis (also known as programmed cell death), isoflavones found in the plant have also been shown to have these capabilities. These findings are encouraging; however, in order to demonstrate the efficacy of red clover in the prevention or treatment of cancer, additional clinical trials that are significantly more comprehensive are required [23].

### 2.3.2 Agriculture Value of *Trifolium pratense*

In the agricultural sector, red clover possesses a substantial and varied value. In addition to being a superb feed legume, it also significantly boosts the productivity of animals [24]. Because of its high protein content and digestibility, it is beneficial to ruminants like as cattle, sheep, and other animals that produce milk.

This results in increased weight gain and milk production. Because of its ability to fix nitrogen from the atmosphere, red clover is a crucial component of environmentally responsible agricultural systems. This is in addition to the fact that it can be used as feed. The symbiotic relationship between Rhizobium bacteria and the soil results in the enrichment of the soil with natural nitrogen, which in turn reduces the requirement for synthetic fertilizers.

This, in turn, results in cheaper input costs for farmers and a reduction in pollution in the environment. Additionally, the broad root system of this plant adds to the long-term productivity and general health of the soil by improving the structure of the soil, increasing the amount of organic matter, and lowering the rate of erosion.

In addition to being an effective cover crop, red clover is beneficial to crops in the surrounding area because it helps reduce the number of weeds and provides a habitat and honey to pollinators such as bees. Additionally, when it is used into crop rotations, it has the potential to disrupt disease cycles and boost the yields of subsequent crops, making it a very valuable component of resilient and diverse agricultural systems [24].

### 2.3.3 Nutritional Value of *Trifolium pratense*

Red clover (*Trifolium pratense*) is a nutritional powerhouse, offering a wealth of benefits whether it's feeding livestock or, in its sprout and extract forms, potentially supporting human health. For farm animals, it's a top-tier forage, brimming with protein, often 15% to 27% of its dry weight, which is fantastic for boosting milk production and promoting healthy weight gain. Its fiber is highly digestible, meaning animals can easily absorb its goodness, and it's packed with essential

minerals like calcium, phosphorus, and potassium, alongside vital trace elements. When we look at red clover for human consumption, especially the sprouts or extracts, the star of the show is its isoflavones. These plant compounds, including biochanin A and formononetin, are fascinating because they can act a bit like human estrogen, leading to research into their potential to ease menopausal symptoms. Beyond that, red clover also provides other polyphenols and flavonoids that act as antioxidants, along with useful vitamins such as C and B vitamins, and even beta-carotene. Its sprouts are also a good source of dietary fiber, and the seeds contain beneficial unsaturated fatty acids and tocopherols (a form of Vitamin E). So, whether it's contributing to robust livestock or offering unique compounds for our well-being, red clover truly stands out for its diverse nutritional contributions[25].

#### **2.3.4 Ecological Value of *Trifolium pratense***

Red clover has incredibly high ecological value, making it essential to both sustainable agriculture and thriving ecosystems. Its remarkable capacity to fix atmospheric nitrogen by collaborating with specialist microorganisms in its root nodules is its most well-known contribution. By enriching the soil and drastically lowering the demand for synthetic nitrogen fertilizers, this natural process essentially takes atmospheric nitrogen and changes it into a form that plants can utilize, saving farmers money and lessening their influence on the environment. In addition to providing nitrogen, red clover's deep and wide root system breaks up compacted soil, improves water infiltration, and builds organic matter—all of which help to enhance soil structure and lessen erosion [17], [20]. Additionally, because of its bright, nectar-rich blooms, red clover is an essential resource for pollinators, including bumblebees, honeybees, and different butterfly species [24], [26]. This helps pollinate other crops and wild plants and maintains a variety of insect populations. As a green manure or cover crop, it aggressively suppresses weeds, outcompeting unwanted plants and reducing the need for herbicides. When added to crop rotations, it also helps break the cycles of pests and diseases, resulting in a more robust and balanced agricultural system. Red clover is essentially a

multifunctional ecological agent that supports biodiversity, improves soil health, and encourages a more sustainable method of land management [26].

## 2.4 Role of Nitrogen in Plant Growth and Development

One of the most important macronutrients for plant life, nitrogen is necessary for plant growth and development [27]. Due to the fact that nitrogen is necessary for the majority of the biological processes that plants carry out, it is impossible for them to thrive without it. This section provides an in-depth explanation of the numerous ways in which nitrogen is beneficial to plants.

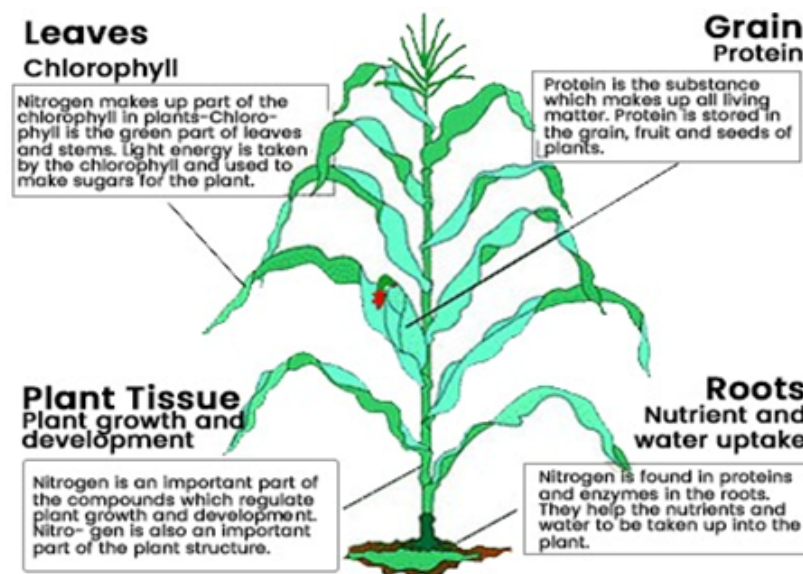


FIGURE 2.2: Nitrogen's function in plant development and growth [27].

### 2.4.1 Nitrogen Helps Build Proteins and Genetic Material

For plants to grow and develop, they need nitrogen, which is one of the most important macronutrients for plant life. Nitrogen is essential for plant growth and development. Since nitrogen is required for the bulk of the biological activities that plants carry out, it is impossible for them to survive without it. This is because nitrogen is essential for the functions that plants do. In the following section, an

in-depth description of the different ways in which nitrogen is beneficial to plants is provided [28].

### **2.4.2 Nitrogen Fuels Photosynthesis by Supporting Chlorophyll Production**

Chlorophyll, the green pigment necessary for photosynthesis, is primarily composed of nitrogen. Light energy is transformed by this process into chemical energy that is stored in sugars, which are the main source of energy for plants. Chlorophyll production declines when nitrogen levels fall, resulting in chlorosis (pale or yellowing leaves). In addition to decreasing the plant's aesthetic attractiveness, this condition hinders photosynthesis, which eventually slows growth [27], [36]. Therefore, maintaining both photosynthetic efficiency and healthy plant development depends on having enough nitrogen [28].

### **2.4.3 Nitrogen in Enzyme Activation and Cofactor Synthesis**

Cofactors that include nitrogen are essential for plants' enzymes to operate properly [27], [28]. FAD and NAD<sup>+</sup>, which promote energy transfer and cellular respiration, are important examples. Coenzyme A, which is essential for metabolic functions, can also be synthesized with nitrogen. The need of nitrogen for sustaining basic plant metabolism and energy production systems is illustrated by these nitrogen-dependent cofactors [27].

### **2.4.4 Nitrogen and Secondary Metabolite Production**

Nitrogen is necessary for primary metabolism as well as the production of secondary metabolites that provide protection in plants. It promotes the synthesis of protective compounds like glucosinolates, caffeine, nicotine alkaloids, and cyanogenic glycosides [29]. These metabolites of nitrogen aid plants in fending off

pests, fending off diseases, and surviving environmental stressors. Plants get vital ecological adaptation and survival benefits in harsh settings by directing nitrogen into these defense mechanisms while preserving growth processes [30].

#### **2.4.5 Nitrogen Drives Plant Growth and Development**

Strong growth is made possible by nitrogen, which is an essential building ingredient for the formation of plant tissue. It promotes elongation and cell division, which causes stems, roots, leaves, and reproductive organs to grow [1], [36]. Nitrogen, a crucial component of amino acids, promotes protein synthesis, which is essential for grain nutrient storage and seed development [28]. The output potential and seed quality of major crops like corn, rice, and wheat are strongly influenced by this essential vitamin. Nitrogen availability ultimately controls crop nutritional value and agricultural output by sustaining these fundamental physiological functions [27], [40].

#### **2.4.6 Nitrogen Helps Plants Cope with Stress**

Nitrogen also aids plants in resisting environmental stressors such salt, drought, and nutritional shortages. It uses signaling molecules like nitric oxide to control how cells react to stress [30], [44]. By improving root function to access deeper soil moisture, nitrogen increases water intake during drought [17], [38]. Additionally, nitrogen triggers defense-related metabolic processes that provide pathogen protection [30], [42]. Nitrogen is essential for maintaining plant health in harsh climatic conditions because it supports these stress adaption mechanisms [30], [38].

#### **2.4.7 Nitrogen Improves Nutrient Uptake Through Healthy Roots**

Plants need robust root systems in order to efficiently absorb water and nutrients from the soil [17], [39]. The creation of proteins and enzymes that improve root

function is facilitated by nitrogen, which aids in root growth. Plants that have well-developed roots that are fed nitrogen may absorb more nutrients from the soil, which enhances their general well-being and yield [39].

### **2.4.8 Plants Depend on Nitrogen from Soil or Symbiotic Relationships**

Through the pores of their roots, plants predominantly receive nitrogen as nitrate ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) [39]. Glutamate and other enzymes convert these ions into organic compounds. In order to convert atmospheric nitrogen into forms that may be utilized, certain plants build symbiotic relationships with nitrogen-fixing bacteria, such as *Rhizobium* in legumes. This natural process provides essential nutrients and improves soil fertility [39].

### **2.4.9 Nitrogen Deficiency can Severely Impact Plants**

Plants lacking in nitrogen exhibit yellow leaves, slowed growth, and lower yields as a result of decreased protein and chlorophyll synthesis [33]. Chlorosis, or the yellowing of leaves caused by a decrease in chlorophyll levels, is the most noticeable symptom [2]. Plants that are deficient in nitrogen also grow more slowly because they cannot produce enough proteins for cell division or tissue formation. Crops like corn or wheat may take longer to mature and produce significantly less if they receive insufficient nitrogen. [7], [40]

## **2.5 Nitrogen Regulation in Plants**

The majority of the nitrogen (N), the most important nutrient required by plants to finish their life cycles, is taken up by the roots [1]. Nitrogen (N) is a macronutrient that plants receive from their soil through their roots throughout their life cycle. The types and amounts of N have a big influence on how plants grow and

interact with their environment [28], [34]. One important factor that usually limits crop productivity in agricultural soils is the amount of nitrogen (N). This leads to the widespread usage of nitrogen fertilizers throughout the world, despite their high cost and detrimental effects on ecosystems. Consequently, it is essential that we advance our current knowledge of N feeding in plants in order to mitigate the worldwide effects of all these human activities [6]. For the most part, free amino acids, urea, and short peptides are the forms of organic and inorganic nitrogen found in soil [39].

Examples of inorganic forms of nitrogen are nitrate ( $\text{NO}_3^-$ ) and ammonium ( $\text{NH}_4^+$ ) [39]. Variations in soil composition and varying microbial conversions, which are both impacted by agronomic methods and environmental factors, cause these nutrients to be considerably less accessible to roots throughout time and space [32]. While the amounts of  $\text{NH}_4^+$  usually vary from 20 to 200 M, the most prevalent form in aerobic soils is  $\text{NO}_3^-$ , which can range from 1 to 5 mM. [39].

Additionally, it has been demonstrated that the availability of amino acids can influence the structure of the roots and the development of the plant, even when the quantities are quite low. Similar to the previous example, plants consume nitrogen directly, as evidenced by the identification of urea transporters in the root cell's plasma membrane. Formulations based on urea account for more than half of the nitrogen fertilizers that are utilized in agricultural production [40], [45].

$\text{NO}_3^-$ , on the other hand, leaches quickly, while  $\text{NH}_4^+$  is slowly released after being absorbed by soil particles. But urea and free amino acids, which are often found in concentrations between 1 and 150 M to 70 M [39], make up only a small part of the nutrients that crops can easily get. There is some disagreement about how important organic nitrogen is for crop nutrition. People haven't paid much attention to crop fields, even though there is evidence that plants can take in proteins, peptides, and amino acids from the soil. However, more investigations in recent years have brought up a number of issues [45]. The early results indicating plants grow faster when they have only amino acids led to the discovery of ways for roots to move amino acids around so that they may be absorbed. The molecular

bases of these systems were confirmed in crops, with *A. thaliana* being the key one that was made clearer.

The early results indicating plants grow faster when they have only amino acids led to the discovery of ways for roots to move amino acids around so that they may be absorbed. The molecular bases of these systems were confirmed in crops, with *A. thaliana* being the key one that was made clearer [39].

Additionally, it has been demonstrated that the availability of amino acids, even in trace levels, can influence the growth and root structure of plants [43]. The direct utilization of this nutrient by plants has also been demonstrated by the discovery of urea transporters in the plasma membrane of root cells. Urea-based formulations make up over half of nitrogen fertilizers used in agriculture [40], [45].

The kinds of nitrogen that are delivered affect hormonal balance, plant development, root and leaf function, seed germination, and the overall amount of nitrogen that is available [28]. Nitrogen nutrition works as a whole system in which different kinds of nitrogen interact with each other in a dynamic way. Studies demonstrate that nitrate, ammonium, and organic nitrogen each have different impacts on metabolism, but they also affect how the others work. These nitrogen sources work together to control important physiological processes such hormone signaling, growth rates, root structure, leaf growth, and seed germination.

Studies demonstrate that nitrate, ammonium, and organic nitrogen each have different impacts on metabolism, but they also affect how the others work. These nitrogen sources work together to control important physiological processes such hormone signaling, growth rates, root structure, leaf growth, and seed germination [37].

The results show that nitrogen nutrition occurs through interconnected metabolic networks. This means that both the contributions of each nitrogen form and their interactions with each other need to be studied [43]. This point of view helps us better understand the many ways nitrogen affects plant growth and physiology, giving us important information about how plants use nutrients [34], [38].

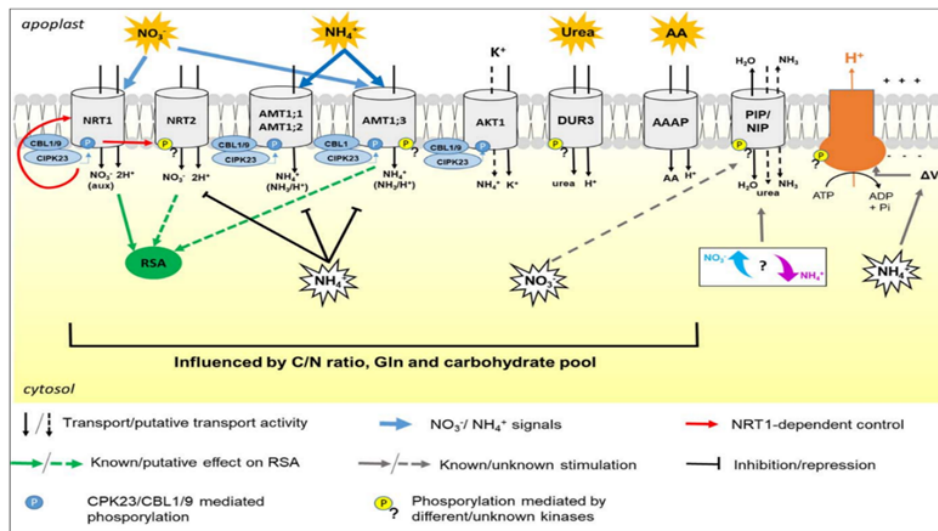


FIGURE 2.3: Nitrogen uptake is regulated by several processes, and the picture shows the main root nitrogen transporters and the metabolic interactions between them. Although it is not the usual in reality, for the purpose of simplicity it is stated that receptors, H<sup>+</sup>-ATPase, transporters, and aquaporin's are all present in the same root cell [39].

Nitrogen forms are transported and detected by proteins found in the PM of root cells. Our knowledge of the signaling components and transcriptional processes that regulate absorption has grown substantially in the last several years [11], [44].

It was emphasized that post-translational modifications (PTMs) of transporters, including protein complex assembly and phosphorylation, are essential for rapid adaptation to unforeseen fluctuations in nitrogen availability. From a physiological perspective, the electrochemical proton gradient and water homeostasis are two critical plasma membrane processes associated with nitrogen absorption. A multitude of research have been undertaken to elucidate the proteins implicated in nitrogen uptake, with the most comprehensive molecular analysis performed on Arabidopsis [11], [44].

We chose to begin with Arabidopsis even though additional crop parallelisms were found. However, due to their many unique characteristics, our work does not cover the description of analogous traits in actinorhizal crops, rice (*Oryza sativa*), and legumes (*Fabaceae* spp.) [13].

## 2.6 Role of Transporters & Transcription Factors in Nitrogen Uptake by Roots

The biological function of these proteins is closely regulated by the accessible amount of nitrate as well as the nutritional condition of the plant for N. When  $\text{NO}_3^-$  becomes available once more, nitrogen-absorbing enzymes and  $\text{NO}_3^-$  transporters are quickly stimulated. This is the usual  $\text{NO}_3^-$  primary response in N-starved plants [44]. The buildup of  $\text{NO}_3^-$  and its byproducts, including glutamine, is associated with further downregulation of absorption after this adaptation. The significance of PTMs was emphasized, and it was shown that  $\text{NO}_3^-$  transporters' transcriptional regulation in Arabidopsis and crops plays a significant role [20]. When it comes to  $\text{NO}_3$  absorption, the PM of Arabidopsis root cells contains three members of the NPF family. The  $\text{NO}_3$  efflux to the external medium is mediated by NAXT1 (AtNPF2.7), whereas the constitutive inflow in the low-affinity range ( $>0.25$  mM) is mostly caused by NRT1.1 (AtNPF6.3) and NRT1.2 (AtNPF4.6) [11], [14], [44].

The relationships of three key protein types transmitters, receptors, and transporters carefully control and impact the intake of N nutrients. Both the plant's nutritional status for N and the availability of  $\text{NO}_3^-$  strictly control the action of these proteins. N-starved plants experience the typical  $\text{NO}_3^-$  primary response (NPR) upon the restoration of  $\text{NO}_3^-$  access, which includes the rapid activation of nitrogen-absorbing enzymes and  $\text{NO}_3^-$  transporters. The buildup of  $\text{NO}_3^-$  and its byproducts, including glutamine, is associated with further downregulation of absorption after this adaptation. In crops and Arabidopsis, it was discovered that the transcriptional control of  $\text{NO}_3^-$  transporters was important. The latter was primarily in charge of constitutive inflow in the low-affinity range ( $>0.25$  mM), while PTMs were crucial for  $\text{NO}_3$  inflow.  $\text{NO}_3$  efflux is mediated to the extracellular medium by NAXT1 (AtNPF2.7) [44].

Ammonium is a crucial nitrogen fertilizer that plants may easily absorb and absorb. It is known as a signaling molecule that affects both RSA and plant growth. The nutritional state of a plant has a significant impact on its capacity to absorb and distribute  $\text{NH}_4^+$  throughout its organs. The estimated  $\text{NH}_4^+$  concentrations

in the cytosol and vacuole at the cellular level are 1–10 mM and 1–45 mM, respectively. It's interesting to note that with both high and low  $\text{NH}_4^+$  supply, the aplastic  $\text{NH}_4^+$  content in roots is buffered at 1 to 2 mM. [28]. Plants take ammonia ions ( $\text{NH}_4^+$ ) from the soil through their roots. Transport proteins in the root cell membranes enable this absorption [45].

## 2.7 Impact of Nitrogen Use Efficiency (NUE)

The most important ingredient that limits agricultural output is nitrogen, which is also necessary for the synthesis of DNA and proteins [1], [27]. Therefore, most farmers employ nitrogen fertilizers to increase agricultural income and production. Nitrogen (N) can be lost through a variety of processes, including nitrous oxide ( $\text{N}_2\text{O}$ ) emissions, ammonia ( $\text{NH}_3$ ) volatilization, nitrate ( $\text{NO}_3^-$ ) leaching, denitrification losses as di-nitrogen ( $\text{N}_2$ ) gas emissions, and others. These processes can worsen climate change and environmental degradation. Recent research indicates that a significant portion of the nitrogen loss in the environment is caused by the agriculture industry [9], [40]. Unless the entire food production-consumption chain is significantly changed and more efficient nitrogen management strategies are developed, future population growth and rising food prices would increase the demand for mineral fertilizer, potentially resulting in additional nitrogen losses [6], [9].

The term nitrogen utilization efficiency refers to the complex processes of nitrogen intake, translocation, absorption, and remobilization [6], [7]. Additionally, the amount of cotton lint and seed produced has increased as a result of the nitrogen application. Cotton's NUE is determined by its nitrogen utilization efficiency (UtE) and nitrogen uptake efficiency (UpE). After nitrogen was applied, which served as a substitute for cotton NUE, lint yielded the collected information [41]. The total N absorbed by plants following N treatment is known as UpE, and the cotton yield ratio is divided by the total plant N to determine UtE. In times of nitrogen scarcity, nitrogen UpE in plants is more significant than nitrogen UtE.

Like other nutrients, nitrogen is a mobile nutrient that is more likely to escape from soil [40].

## 2.8 Enhancement of Nitrogen Use Efficiency

Better irrigation techniques, improved fertilizer administration that takes the 4Rs into account, and the use of hybrids with higher potential yields and lower N inputs are all associated with increases in nitrogen use efficiency (NUE). Through efficient management, or N source, rates, time, and placement, NUE is increased in agricultural systems. NUE is generally improved by nitrogen inhibitors, split nitrogen treatments, irrigation scheduling, and fertilizer placement techniques that take the crop variety and soil into account. It has been demonstrated that late N treatment improves N accumulation during the boll-setting phase. NUE is also thought to be an essential part of fertilizer inputs to any agricultural system since it keeps the N balance between inputs and outputs without having an adverse effect on the economy or environment. Applying nitrogen fertilizer to rose plants during the first bloom stage is another way to increase NUE since they use nitrogen more effectively for reproduction [41].

Several strategies can improve NUE:

### 2.8.1 Precise Farming

By accurately applying fertilizers according to plant needs, technology can prevent over-application and reduce nitrogen runoff and leaching [41].

### 2.8.2 Optimizing Fertilizer Management

By accurately applying fertilizers according to plant needs, technology can prevent over-application and reduce nitrogen runoff and leaching [41].

### **2.8.3 Crop Rotation and Cover Crops**

NUE may be increased by crop rotation or the use of cover crops, which improve nutrient uptake, reduce nitrogen leaching, and diversify root structures [41].

### **2.8.4 Nitrogen-Fixing Plants**

By utilizing nitrogen-fixing bacteria in the soil or introducing leguminous crops, accessible nitrogen can be enhanced without only relying on synthetic fertilizers [41].

### **2.8.5 Genetic Approaches**

Crops bred for improved nitrogen absorption efficiency or resilience to low nitrogen situations can have their NUE raised [41].

### **2.8.6 Soil Health Improvement**

Plants are better able to absorb nitrogen when soil structure, microbial activity, and nutrient availability are improved through practices including mulching, composting, and no-till farming [41].

### **2.8.7 Precision Irrigation**

Managing water effectively can reduce nitrogen losses by preventing excessive leaching caused by overwatering [41].

Raising NUE not only boosts crop output and reduces input costs for farmers, but it also aids in addressing environmental issues such as water pollution from nitrogen runoff and greenhouse gas emissions from nitrogen-based fertilizers. Achieving a balance between agricultural productivity and environmental sustainability requires optimizing nitrogen usage [41].

## 2.9 Transcription Factors

Plant transcription factors typically have a number of structural components, including an oligomerization site, a transcription-regulation domain, a DNA-binding region, and a nuclear localization signal, with a few exceptions. Some transcription factors may have two different types of DNA-binding and oligomerization domains, but most only exhibit one type, often in several copies. Crucial elements of transcription factor activity are regulated by the combined tertiary structure of DNA-binding regions and oligomerization sites, which usually overlap or are close to one another. Several transcription factors feature pairs of nuclear localization signals, and basic amino acid residues are crucial to their function, just like DNA-binding domains [10], [12].

Members of multigene families, which encode transcription factors, can be found scattered over the genome or grouped on a single chromosome. Distribution and sequence studies suggest that exon capture, gene duplication and mutation, and translocation may have had a role in the formation of transcription factor families. In plants, the activities of transcription factor genes are controlled both during and after transcription, whereas the activities of their protein by-products are controlled throughout the post-translational process. This intricate regulation demonstrates how challenging it is to regulate plant gene expression [12], [44].

## 2.10 Role of Transcription Factors

To initiate transcription, regulatory proteins known as TFs bind to particular DNA regions. Because it necessitates the coordinated actions of multiple proteins, the regulation of gene expression by transcription factors in eukaryotes is considered combinatorial. A significant number of housekeeping genes, which appear constitutive or uncontrolled, are required by nearly all cell types. However, tissue-specific control of gene expression is the primary driver of cellular differentiation, as seen by the vast diversity of cell types found in many species. The same genome produces all cell kinds, each with a unique function—muscle cells

contract and expand, immune system cells recognize invaders, and red blood cells exchange oxygen. To enable a cell to carry out its assigned role, genes governing cell identity are activated under highly particular temporal, geographical, and environmental circumstances [44]. The main method for regulating gene expression in eukaryotes appears to be transcriptional regulation.

Through the activation or repression of transcription, transcription factors regulate the expression of genes. The main method for regulating gene expression in eukaryotes appears to be transcriptional regulation. Through the activation or repression of transcription, transcription factors regulate the expression of genes. These activators and repressors function through chromatin structural change, protein-protein interactions, and DNA-protein interactions. Additionally, they are all governed by various regulations. Through a number of unique mechanisms, a single transcription factor can affect the transcription of several target genes [12], [45].

TFs regulate plant growth and development, impacting processes such as cell differentiation, organ formation, blooming, and fruit production. Different TFs control different developmental stages in a variety of plant tissues [44]. Plants continually adapt to a variety of environmental factors, including light, temperature, drought, and diseases. TFs contribute significantly to these responses by activating genes that help plants adapt and survive in a variety of environments.

In many plant processes, transcription factors (TFs) are important regulators. They affect pathogen defense mechanisms and drought adaptation by regulating the expression of genes that respond to stress. Additionally, TFs control the formation of secondary metabolites, such as flavonoids, pigments, and protective substances. In addition to these tasks, they control vital physiological processes include hormone signaling pathways and nutrition absorption [34], [37].

In order to create complex signaling networks, TFs frequently interact with other regulatory proteins. In response to developmental and environmental stimuli, these interactions provide dynamic regulatory networks that precisely coordinate

gene expression. Plants are able to incorporate various information and mount suitable physiological responses thanks to these intricate networks [34], [42].

## 2.11 Genes as Transcription Factor in Regulating Nitrogen Response

*RWP-RK* transcription factors are particular to plants and affect gametogenesis, nodulation, and nitrate sensitivity. The phylogeny, regulation, and genome-wide characterization of *RWP-RK* genes in nitrogen-fixing plant species that nodulate and those that do not, however, are little understood (NFC). As a result, 292 *RWP-RKs* were discovered, comprising 278 from 25 NFC species and 14 from the outgroup *Arabidopsis thaliana*. Of the 292 *RWP-RKs* that we classified, two subfamilies are the *RWP-RK* domain proteins and NIN-like proteins (NLPs) [46], [47] (RKDs). The NLP genes were only marginally more increased in nitrate response and nodulation than the RKD genes, per phylogeny and transcriptome analyses of the *RWP-RKs*. Additionally, the nodule-specific NLP genes in certain nodulating NFC species and the *AtNLP* genes in *A.thaliana* most likely originated from (OG0002084). The expression of *AtNLP* genes and symbiotic genes during N-starvation is more strongly correlated, based on *A.thaliana* co-expression networks under N-supplementation and -starvation circumstances [47], [49]. By controlling the expression of *PvNLP2*, which is closely related to *AtNLP6* and *AtNLP9* but has a different common ancestor, it was demonstrated that N-starvation encouraged nodulation in *P. vulgaris* (OG0004041). Since it was shown that the *NLP* genes causing the N-starvation response and the particular expression of nodulation had different origins, nodulation would have originated in NFC plant species [46].

## 2.12 *NLP* Gene Role in Nitrogen Signaling

NIN-like proteins are crucial transcription factors for nitrate signaling (*NLPs*). Soil *NLP* genes respond to nitrogen. As sensors of nitrogen availability, these genes

activate or deactivate in response to either an excess or a deficiency of nitrogen. Transcription factors that are a component of *NLP* cause the expression of genes that are involved in the uptake of nitrogen from the soil and its incorporation into organic compounds inside the plant. In nitrogen metabolism, they control the expression of regulatory proteins, enzymes, and transporters [46], [49].

According to reports, the nutrient- $\text{Ca}^{2+}$ -*NLP* regulation pathway integrates transport, transcription, systemic growth activities, and metabolism in plants and is essential for nitrate signaling [46].

The Arabidopsis nitrate transporter 1.1 (NPF6.3/NRT1.1) has been identified as a plasma membrane nitrate sensor [11], [14]. While nitrate activates the important nitrate response genes, calcium-dependent protein kinases 10/30/32 (CK10/30/32) mediate  $\text{Ca}^{2+}$  signals by nitrate and phosphorylate *NLP6/7* to preserve their position in the nucleus [14], [44]. The NIN protein, which controls the development of symbiotic root nodules, originated in the legume *Lotus japonicus*. Compared to other non-leguminous plants as Arabidopsis, rice, wheat, and maize, mammals lacked as many NIN proteins or *NLPs*. *NLPs* have an extra PB1 domain for protein-protein interactions, and NIN proteins and *NLPs* both feature the RWP-RK domain for DNA binding. NRG2, PCF (TCP)-domain family protein 20, and nitrate-inducible GARP-type transcriptional repressor 1 (NIGT1) are other transcription factors that have been shown to interact with *NLPs*. By encouraging lateral roots and root hairs, *NLP* genes affect root growth by increasing the surface area available for soil nitrogen uptake.

By controlling important genes involved in nitrogen use, they improve the efficiency of nitrogen use. These genes enhance plants' capacity to efficiently use scarce nitrogen resources by assisting them in adapting to fluctuating nitrogen levels. *NLP* genes improve plants' ability to adapt to shifting nitrogen conditions by regulating how they react to signals of nitrogen availability, whether there is an excess or a shortage [47].

Beyond nitrate signaling, *NLPs* are involved in a number of processes, including the control of root cap cell detachment, nitrate-dependent nodule formation,

nitrogen-phosphate crosstalk, nitrogen shortage response, and nitrate-enhanced seed germination. Their various functions demonstrate how crucial they are to the overall control of plant nutrients [46], [47].

## 2.13 Study of the *NLP* Gene Family in *Arabidopsis thaliana*

The transcription factor NIN-LIKE PROTEIN 7 (*NLP7*) is essential to the nitrate response network of *Arabidopsis*. It coordinates the transcriptional response that is controlled by nitrate. Since all nine of the *NLPs* can bind the nitrate responsive cis- element (NRE) in vitro, it is believed that they all play a role in the nitrate response. These *NLPs* are encoded in the *Arabidopsis* genome [46], [47]. Through a nuclear retention mechanism that causes calcium-dependent protein kinases to phosphorylate a conserved S205 residue and activates early nitrate-responsive genes, nitrate controls the nucleus accumulation of *Arabidopsis NLP7* and *NLP6* (CDPKs) [44], [47].

Despite being crucial for nitrate-promoted seed germination, *NLP8*'s function is not controlled by nitrate-dependent nuclear retention because it is kept in the nucleus independent of nitrate exposure [47], [48]. Recent research indicates that both *Arabidopsis NLP2* and *NLP7* are essential for vegetative growth because they significantly affect rosette biomass when grown in nitrate. As a result, *NLP2* and *NLP7* share characteristics like nitrate-dependent activation but are members of distinct evolutionary clades. The proportionate role and impact of *NLP2* within the nitrate-related GRN, as well as its effect on nitrate-dependent growth, are still unknown. In reaction to nitrate, *NLP2* plays a key function in controlling the metabolism of carbon and energy [44], [47].

The nitrate-dependent coordination of N metabolism and other related activities, some of which depend on the molecular interactions between *NLP2* and *NLP7* [44], [47].

## 2.14 The Investigation of the *NLP* Gene family in *B. napus*

Plant responses to nitrogen deficiency depend on conserved transcription factors called NODULE-INCEPTION-like proteins. Of the 31 *NLP* genes present in *B. napus*, 15 are located in the C sub genome and 16 in the A sub genome. Most *BnaNLP* proteins are located in the nucleus, according to subcellular localization predictions. The *NLP* gene family can be categorized into three groups based on phylogenetic analysis, and both monocot and dicot progenitors had at least three ancient copies of the gene prior to their divergence. The Group III *NLP* genes may have arisen from various duplication events in the Brassicaceae species. A three-dimensional structural analysis revealed that 14 amino acids of the *BnaNLP7-1* protein were involved in DNA binding, despite the fact that no binding sites were discovered [48].

## 2.15 The Investigation of the *NLP* Gene Family in Tomatoes

It is evident that nitrate is the preferred inorganic nitrogen source for the tomato (*Solanum lycopersicum*), one of the most important horticultural crops. An analysis of tomato *NLP* genes in detail. *NLP* genes in tomatoes are well conserved. The development of the *SINLP* gene was mostly triggered by segmental duplication. The functional divergence of gene families during evolution through positive selection may be attributed to specific *SINLP* genes. The patterns of expression of the *SINLP* genes confer several physiological functions, such as controlling nitrate uptake in tomato growth and development. Further functional investigation is needed to understand each *SINLP*'s regulatory responsibilities, especially those of *SINLP3* and *SINLP6*. It is anticipated that in order to determine the molecular basis of nitrogen use and enhance tomato nitrogen usage efficiency, a

thorough investigation of *SINLP*'s involvement in diverse nutritional situations will be required [49].

## 2.16 Key Findings on the *NLP* Gene Family in Rice

Recent research has provided insight into how rice plants use a unique protein known as *OsNLP1* to regulate their nitrogen requirements. This protein helps rice react to nitrogen availability in its surroundings by functioning as a master switch. Plants swiftly increase their production of *OsNLP1* in response to nitrogen scarcity. *OsNLP1*'s capacity to increase rice productivity is what makes it so significant. Regardless of whether nitrogen is abundant or scarce, researchers discovered that rice plants with additional *OsNLP1* develop better and yield more grains. However, rice that is deficient in this protein suffers, particularly in situations when nitrogen is scarce, leading to reduced yields. *OsNLP1* functions by directing a network of genes related to the intake and utilization of nitrogen. When necessary, it activates these genes by directly interacting with their regulatory switches. This finding is intriguing because it raises the possibility that we could create rice cultivars with improved nitrogen utilization, which could lower fertilizer requirements without sacrificing yields [50].

Farmers may be able to produce more food with fewer resources thanks to these results, which could result in more sustainable rice growing methods. The study demonstrates how a fundamental understanding of plant biology can result in workable answers to actual agricultural problems [50].

## 2.17 Study of *NLP* Genes in Maize

By examining the conserved domains and gene structures of nine *NLP* genes (*ZmNLPs*) in maize, researchers were able to identify them. At different phases of growth, these genes displayed distinct patterns of expression in eight different

tissues. Strong overexpression of two genes, *ZmNLP4* and *ZmNLP5*, suggested that they play a part in the early nitrogen response. This was corroborated by the fact that within 30 minutes of nitrogen treatment, the known nitrogen-responsive gene *ZmNRT2.2* increased more than fivefold [51].

Four *ZmNLPs* (*ZmNLP4*, 5, 6, and 8) increased their expression when nitrogen was accessible, according to additional qRT-PCR experiments. This demonstrates that these genes aid in maize's reaction to nitrogen. Additionally, the study discovered that *NLP* proteins bind to certain DNA sequences (NREs) that regulate genes linked to nitrate. Better crop growth with less fertilizer may result from our understanding of how maize regulates nitrogen utilization [51].

## 2.18 Key Findings on the *NLP* Gene Family in Wheat

The genome-wide study of *NLP* (NIN-like protein) transcription factors in wheat has greatly improved our understanding of how nitrogen signals in this important crop. Researchers found 33 *NLP* genes in the hexaploid genome of wheat using a variety of genomic and transcriptomic methods. This is a huge increase compared to diploid plants because of polyploidization processes. The study showed that a number of *TaNLP* genes, namely *TaNLP4* and *TaNLP7*, are turned on a lot more when there isn't enough nitrogen. These genes control how much nitrogen the plant takes in and uses.

It looks like these transcription factors work together to control the production of nitrate transporters and metabolic enzymes, which helps wheat use nitrogen more efficiently. In addition to their main job of responding to nutrients, the study found that some *TaNLPs* also respond to abiotic stress, which means they may be involved in larger mechanisms that help organisms adapt to their environment. Phylogenetic analysis gave us information about how plants evolved by putting wheat *NLPs* into different clades with functions that are either the same as or different from those of other plants.

Finding *TaNLP7* as a major regulator opens up intriguing possibilities for molecular breeding efforts that want to create wheat types that use nitrogen more efficiently and use less fertilizer. This research not only helps us understand wheat's nitrogen response network better, but it also gives us useful genetic targets for making wheat better in a way that is good for the environment and meets the growing need for more food. The results show that *NLP* transcription factors are important for wheat's ability to adjust to changing nitrogen levels and environmental challenges [52].

## 2.19 Key Findings on the *NLP* Gene Family in Tea plant

A new research of the entire genome has found and described NIN-like protein (*NLP*) transcription factors in tea plants. These proteins are very important for how the plants respond to nitrates and how they use nitrogen. Scientist's uncovered 7 different *CsNLP* genes in the tea plant genome, but they were not equally spread out across chromosomes. Phylogenetic analysis put these *CsNLPs* into three main groups. They were more closely related to *NLP* genes from dicotyledonous plants like poplar and Arabidopsis than to monocots. The study indicated that distinct *CsNLP* members have diverse expression patterns in different tea plant tissues. For example, numerous genes showed considerable upregulation in roots when they were treated with nitrate. *CsNLP2* and *CsNLP7*, in particular, showed a high response to nitrate induction, which suggests that they are very important for using nitrogen. Structural study showed that these *NLP* proteins have conserved RWP-RK and PB1 domains, but they also have distinct sequence changes that may help tea plants perform different things. The study also found a lot of cis-elements in *CsNLP* promoters that respond to stress and hormones, which suggests that they might help connect nitrogen signaling with responses to environmental stress. These results are the first systematic study of *NLP* genes in tea plants. They also lay the groundwork for figuring out how these genes work at

the molecular level to improve nitrogen use efficiency, which could help improve tea quality and output in the future by better managing nitrogen [53].

## 2.20 Key Findings on the *NLP* Gene Family in Alfalfa

Researchers found 24 *MsNLP* genes spread unevenly throughout the lucerne genome. This is a big increase compared to other plants because to whole-genome duplication and tandem duplication events. Phylogenetic analysis put these *MsNLPs* into three different groups. Members of Group I showed the strongest reactions to the availability of nitrate. The study found that some *MsNLP* genes, such as *MsNLP5* and *MsNLP12*, were mostly expressed in roots and nodules. This suggests that they have unique roles in taking up nitrogen and fixing nitrogen in symbiosis. Several *MsNLPs* showed a robust response to abiotic stress, with *MsNLP3* and *MsNLP8* being strongly activated by drought and salt stress, respectively. The study also found many stress-responsive cis-elements in *MsNLP* promoters, which are molecular proof that they play a role in stress adaptation. These results greatly improve our understanding of how nitrogen signaling works in perennial legumes. They also show that *MsNLPs* are good targets for molecular breeding methods that could make lucerne better at using nitrogen and dealing with stress [54].

# Chapter 3

## Methodology

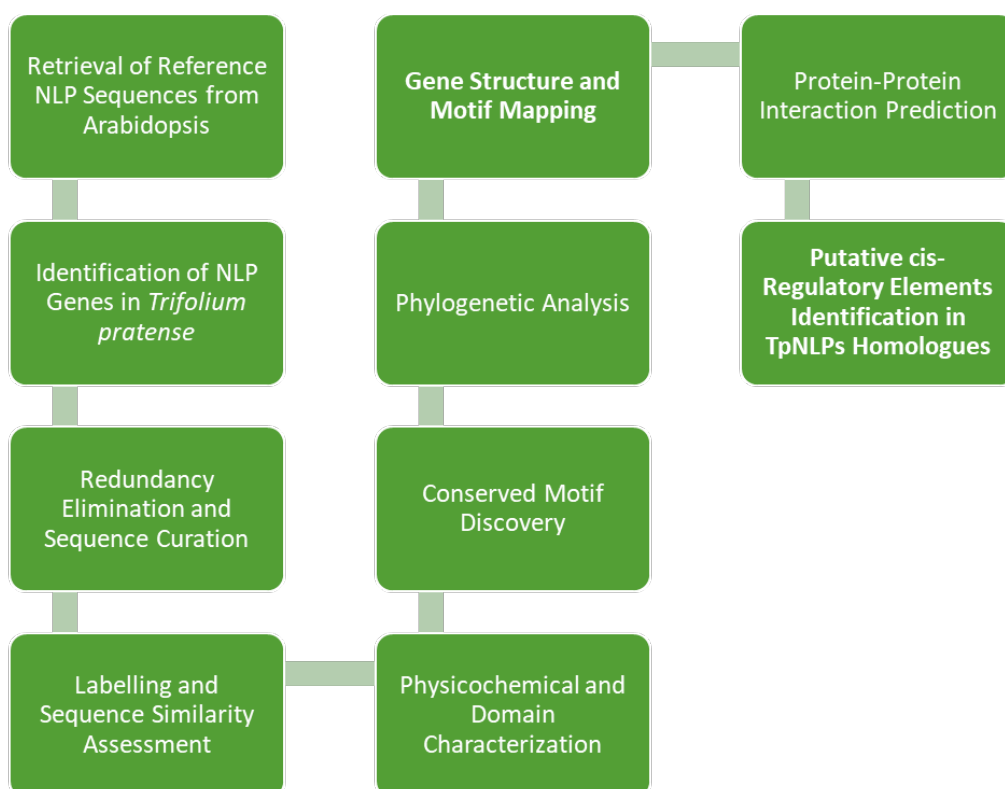


FIGURE 3.1: Research Methodology

This study aims to identify and characterize *NLP* (NIN-like protein) gene family members in *Trifolium pratense* (red clover) using a comprehensive *in silico* approach. The methodology will include genome screening, sequence validation, structural and functional analysis, and interaction prediction to ensure a thorough investigation.

### 3.1 Retrieval of Reference NLP Sequences from *Arabidopsis*

*Arabidopsis thaliana* NLP protein sequences were taken from the TAIR database (<http://arabidopsis.org>) in order to provide a trustworthy reference collection. Sequences that are longer than 500 amino acids and that have both the PB1 (protein interaction) and RWP-RK (DNA-binding) domains were chosen. *Trifolium pratense* homologous genes found using these well-characterized *Arabidopsis* NLP sequences as templates.

### 3.2 Identification of NLP Genes in *Trifolium pratense*

We looked for possible NLP genes in the *Trifolium pratense* genome (found in NCBI or Phytozome) using BLASTp (Protein Basic Local Alignment Search Tool). Using the NCBI Conserved Domain Database (CDD) ([www.ncbi.nlm.nih.gov/Structure/.html](http://www.ncbi.nlm.nih.gov/Structure/.html)), we verified their identities by looking for the existence of conserved RWP-RK and PB1 domains, which are essential characteristics of NLP proteins. For additional analysis, only sequences with both domains were kept.

### 3.3 Redundancy Elimination and Sequence Curation

Using BLASTp (Protein Basic Local Alignment Search Tool), we searched the *Trifolium pratense* genome for potential NLP genes (identified in NCBI or Phytozome). In order to confirm their identities, we searched the NCBI Conserved Domain Database (CDD) ([www.ncbi.nlm.nih.gov/Structure/.html](http://www.ncbi.nlm.nih.gov/Structure/.html)) for the presence of conserved RWP-RK and PB1 domains, which are crucial features of NLP

proteins. Only sequences containing both domains were retained for further examination.

### 3.4 Labelling & Sequence Similarity Assessment

Following the elimination of sequences that were repetitive, spliced, or incomplete, Lab-IDs were initially assigned to each consecutive sequence. In the second step of the process, a comparison table was constructed in order to ascertain the sequence similarities that exist between the *TpNLP* protein sequences. There would be no consideration given to sequences that had a similarity level that was lower than the threshold, which is 90%.

### 3.5 Physicochemical and Domain Characterization

Use ExPASy ProtParam (<https://web.expasy.org/protparam/>) to look at molecular weight, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY), and other traits. We were used CD-Search to look for conserved domains, focusing on confirming the presence of PB1 and RWP-RK domains. Using CELLO (<http://cello.life.nctu.edu.tw/>), were able to guess where in the cell something is located.

### 3.6 Conserved Motif Discovery

Using the MEME Suite (<https://meme-suite.org/meme/tools/meme>), motifs can be found with widths ranging from 6 to 50 amino acids and a maximum of 15 motifs per sequence. NLP sequences from *Arabidopsis* and *Trifolium pratense* will be used in the investigation. The detected motifs were functionally annotated

using InterPro, and motif conservation was displayed using WebLogo's sequence logos.

### 3.7 Phylogenetic Analysis

NLP protein sequences from *Trifolium pratense* and other plants (such as *Zea mays*, *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor*, *Solanum lycopersicum*, *Physcomitrella patens*, *Brassica napus*, *Populus trichocarpa*, and *Triticum aestivum*) were aligned using Clustal Omega in order to investigate evolutionary relationships. Phylogenetic trees were constructed in MEGA-X ([megasoftware.net](http://megasoftware.net)) using the neighbor-joining technique with 1000 bootstrap repeats, and they were displayed using the iTOL platform iTOL; <https://itol.embl.de/>.

### 3.8 Gene Structure and Motif Mapping

GSDS 2.0 (<http://gsds.gao-lab.org/>) was used to align the genomic and CDS sequences of *TpNLP* genes in order to visualize gene structures. Conserved motifs identified in the MEME analysis are going to be mapped to exon-intron boundaries to explore structure-function relationships and possible alternative splicing events.

### 3.9 Protein-Protein Interaction Prediction

To anticipate the cellular proteins that interact with *TpNLP* and *AtNLPs*, a web-based tool known as STRING (<https://string-db.org/>) was used. The tool's figures and explanation were exported for future reference and comparison analysis. A functional homology was found by comparing the interacting proteins of *TpNLPs* and *AtNLPs*. Pink and blue indicate known interactions, whereas dark green, blue, and red indicate predicted interactions. The color lines display both known and predicted interactions. While the colored lines on the edges represent protein

sequences, the variously colored balls represent nodes. Proteins involved in N absorption, assimilation, and transport interact with NLP in PPI; as a result, they predict the involvement of NUE overall.

### 3.10 Putative cis-Regulatory Elements Identification in *TpNLPs* Homologues

The upstream promoter region of *TpNLPs* (2000 bps) obtained from the online database Plant Ensemble (<http://www.plants.ensembl.org/>) was used to identify gene regulatory elements in promoter regions of *TpNLPs*. Cisregulatory elements in promoter regions were examined online (<http://bioinformatics.psb.ugent.be/webtools/plantcare/>).

# Chapter 4

## Results

### 4.1 Identification of *NLP* genes in *Trifolium pratense*

In this study, the accession number of the *NLP* gene in the TAIR database ([TAIR: http://arabidopsis.org/](http://arabidopsis.org/)) was used to retrieve the full-length protein-coding sequences of the *NLP* gene from the *Arabidopsis thaliana* *NLP* gene family. The *NLP* gene had nine different variations, and the longest sequence was selected. Protein Blast against the *Trifolium pratense* genome available in NCBI ([www.ncbi.nlm.nih.gov/genome/gdv/](http://www.ncbi.nlm.nih.gov/genome/gdv/)). An Excel file containing the findings was downloaded in CSV format.

An Excel file including the results' names, accession numbers, and lab IDs was saved. To create a percentage similarity table of all downloaded sequences, they were self-aligned. Six sequences of *Trifolium pratense* were downloaded in FASTA format, and a new Excel sheet containing the selected sequences along with their lab IDs, gene names, and accession numbers was created after the similarity table, which excluded all spliced variants, repeated/redundant sequences, and short or incomplete fragments from the retrieved sequences concurrently validated through conserved domain identification. The *Arabidopsis thaliana* *NLP* protein sequences

and the pfam accessions of RWP-RK (PF02042) and PB1 domain (PF00564) were used as searches to identify *NLPs* in the *Trifolium pratense* genome.

## 4.2 Conserved Domain Identification

In the end, six *TpNLPs* were found that had both RWP-RK and PB1 domains. They were given numbers from 1 to 6. The accession numbers of the identical or duplicate sequences identified in certain databases are provided. The CDD ([www.ncbi.nlm.nih.gov/Structure/cdd.html](http://www.ncbi.nlm.nih.gov/Structure/cdd.html)) database was used to find conserved domains in the obtained sequences. These domains were then used to choose potential *TpNLPs*. We pasted all six of these sequences and the query sequences into CD search and sent them in.

The result is summarized in (Table 4.1) which includes query, hit type, position, E- value, bit-score, Accession, and short name of both plants.

TABLE 4.1: Conserved domains of NLP gene families in *Arabidopsis thaliana* and *Trifolium pratense*

Organ- ism	Query	Hit type	Position		E-value	Bitscore	Accession	Short name
			From	To				
Arabidopsis thaliana	<i>AtNLP1</i>	Specific	812	893	6.21E-41	144.773	cd06407	PB1_NLP
		Specific	608	656	1.10E-23	94.0888	pfam02042	RWP-RK
	<i>AtNLP2</i>	Specific	864	944	1.19E-41	146.699	cd06407	PB1_NLP
		Specific	648	696	1.82E-23	93.7036	pfam02042	RWP-RK
	<i>AtNLP3</i>	Specific	674	758	1.48E-40	143.232	cd06407	PB1_NLP
		Specific	498	546	1.45E-23	93.7036	pfam02042	RWP-RK
	<i>AtNLP4</i>	Specific	745	826	6.79E-43	150.166	cd06407	PB1_NLP
		Specific	558	606	1.28E-23	94.0888	pfam02042	RWP-RK
	<i>AtNLP5</i>	Specific	711	787	3.53E-36	130.906	cd06407	PB1_NLP
		Specific	549	597	3.77E-24	95.6296	pfam02042	RWP-RK
	<i>AtNLP6</i>	Specific	742	822	2.85E-34	125.513	cd06407	PB1_NLP
		Specific	556	604	4.91E-24	95.2444	pfam02042	RWP-RK
	<i>AtNLP7</i>	Specific	864	944	4.11E-34	125.128	cd06407	PB1_NLP
		Specific	591	639	1.20E-24	97.1704	pfam02042	RWP-RK
	<i>AtNLP8</i>	Specific	835	915	6.63E-39	138.995	cd06407	PB1_NLP
		Specific	590	638	1.37E-24	96.7852	pfam02042	RWP-RK
	<i>AtNLP9</i>	Specific	793	874	3.20E-34	125.513	cd06407	PB1_NLP
		Specific	535	583	2.37E-24	96.0148	pfam02042	RWP-RK

Table 4.1 continued from previous page

Organ-ism	Query	Hit type	Position		E-value	Bitscore	Accession	Short name
			From	To				
Trifolium pratense	<i>TpNLP1</i>	Specific	806	887	1.745e-40	143.232	cd06407	PB1_NLP
					7.55848e-23	92.1838	pfam02042	
						39.372		
		Specific	586	634	0.00204687		pfam13185	RWP-RK
		Specific	186	286				GAF_2
	<i>TpNLP2</i>	Specific	849	930	1.43641e-38	138.225	cd06407	PB1_NLP
					1.22556e-22	91.4134	pfam02042	
		Specific	623	671				RWP-RK
	<i>TpNLP3</i>	Specific	749	829	4.35E-36	130.906	cd06407	PB1_NLP
						94.8802		
		Specific	578	626	6.15E-24		pfam02042	RWP-RK
	<i>TpNLP4</i>	Specific	913	994	7.63807e-34	124.743	cd06407	PB1_NLP
1.75992e-17					76.7758			
	Specific	573	649			pfam02042	RWP-RK	
<i>TpNLP5</i>	Super family	826	906	2.7165e-29	111.646	cd06407	PB1_NLP	
				2.87199e-23				
	Specific	578	626		93.3394	pfam02042	RWP-RK	
<i>TpNLP6</i>	super family	868	946	8.46736e-28	107.409	cl02720	PB1 super family	
				5.49564e-23				
	Specific	603	651		92.569	pfam02042	RWP-RK	

### 4.3 Physiochemical Properties

The ExPasy ProtParam tool (<https://web.expasy.org/protparam/>) looked at the physical and chemical properties of the NLP gene families in *A. thaliana* and *Trifolium pratense*. It looked at things like protein molecular weight (MW), hydropathicity (GRAVY), and theoretical isoelectric point (pI) values for both plants. These results are shown in Table 4.2. The lengths of the genes, proteins, and molecules, as well as the pI and GRAVY values, were very similar for both plants. The readings of both plants had negative GRAVY values, which meant that the proteins were hydrophilic. The PI and GRAVY readings for both plants are very similar. CELLO (<http://cello.life.nctu.edu.tw/>) was used to look at the sub-cellular localizations, which suggested that *A. thaliana* and *T. pratense* NLPs were found in the nucleus.

The average length of the proteins in *AtNLPs* and *TpNLPs* was 808 and 1008 amino acids, respectively. The lengths of the genes in *AtNLPs* and *TpNLPs*, on the other hand, were very different.

All of the *AtNLPs* and *TpNLPs* except *AtNLP3* had pl values below 7, which means they are acidic proteins. *AtNLP3*, on the other hand, had a pl value of 8.14, which means it is a basic protein.

TABLE 4.2: Physical and chemical properties of NLP gene families of *Arabidopsis thaliana* and *Trifolium pratense*

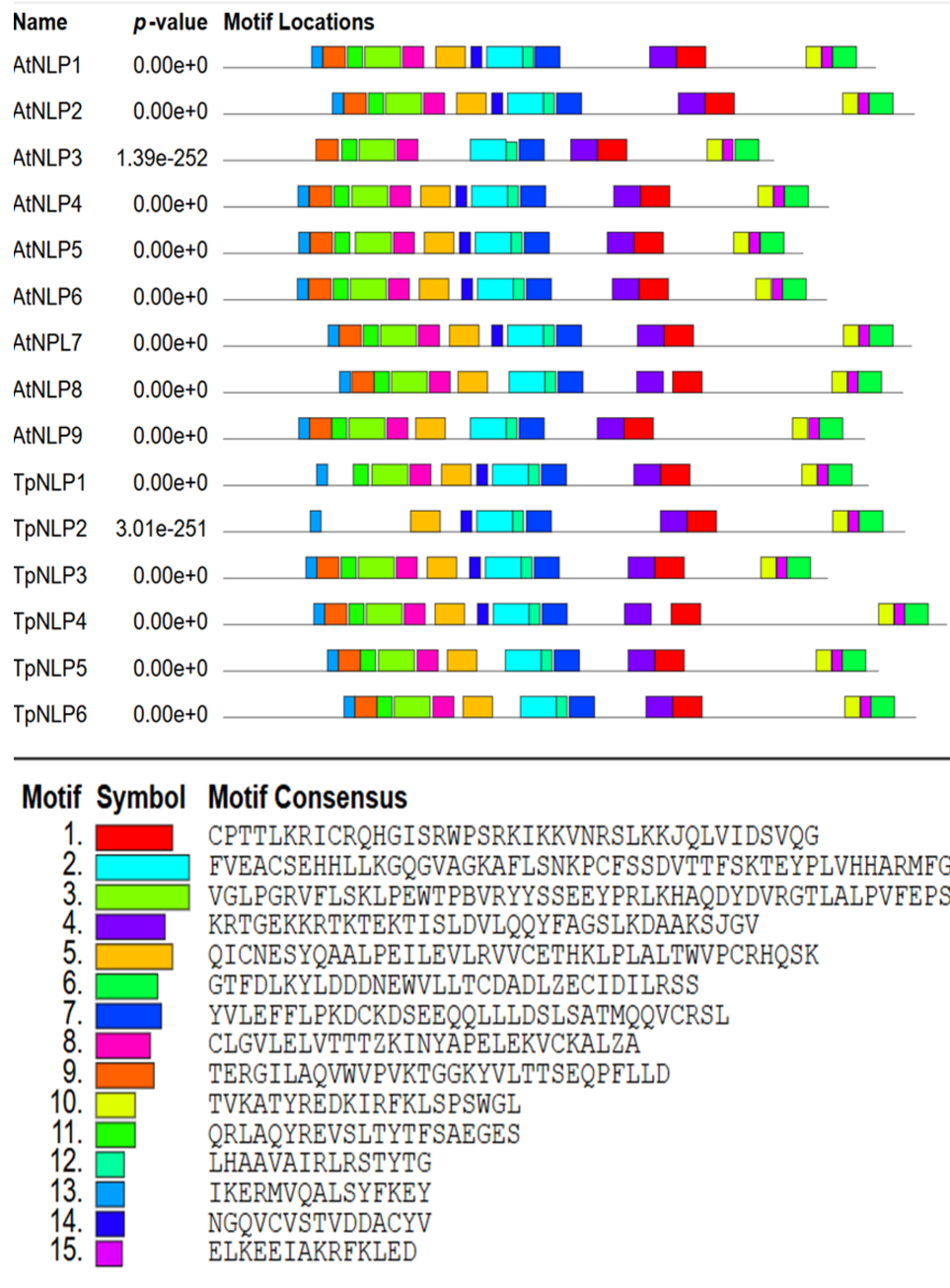
Plant	Gene name	Chr.	Position	Gene Length (bp)	Protein Length (aa)	Molecular Weight	Iso - electric Point	GRAVY
Arabidopsis thaliana	<i>AtNLP1</i>	2	7466687 - 7471586	4900	909	100885.3	4.83	-0.443
	<i>AtNLP2</i>	4	16777264 - 16782054	4791	963	107277.6	5.76	-0.476
	<i>AtNLP3</i>	4	17954710 - 17958063	3354	767	85065.7	8.14	-0.271
	<i>AtNLP4</i>	1	7154425 - 7158284	3860	844	94231.1	5.45	-0.472
	<i>AtNLP5</i>	1	28639453 - 28643086	3634	808	90683.4	6.13	-0.467
	<i>AtNLP6</i>	1	23959627 - 23963083	3457	841	93862.6	6.3	-0.356
	<i>AtNLP7</i>	4	12479528 - 12484049	4522	959	105741.1	5.69	-0.42
	<i>AtNLP8</i>	2	18061716 - 18066692	4522	934	103284.1	5.45	-0.436
	<i>AtNLP9</i>	3	22009010 - 22012791	3782	894	98712.1	5.29	-0.383
Trifolium pratense	<i>TpNLP1</i>	7	5054760 - 5059659	4900	899	100760.13	5.62	-0.539
	<i>TpNLP2</i>	4	196210 - 201256	5047	950	103819.66	6.1	-0.439
	<i>TpNLP3</i>	2	21560468 - 21566247	5780	842	94791.5	5.39	-0.343
	<i>TpNLP4</i>	1	8251714 - 8256441	4728	1008	111985.05	5.36	-0.345
	<i>TpNLP5</i>	3	14964897 - 14969623	4727	913	101717.95	5.16	-0.437
	<i>TpNLP6</i>	7	29669937 - 29677255	7319	965	107002.03	5.63	-0.504

## 4.4 Conserved Motif Discovery

Using MEME Suite (<https://meme-suite.org/meme/tools/meme>), up to 15 consensus motifs were identified in *TpNLP* proteins (Fig. 4.2 and Fig. 4.1) in comparison to *AtNLPs*. Significantly conserved motifs in the proteins of *A. thaliana* and *T. pratense* were found in every sequence. Every *AtNLP* and every *TpNLP* had motifs, with the exception of *AtNLP3*, -5, and -8, which have 14 motifs, and *AtNLP9*, which has 13. *TpNLP5* has 10 motifs, while *TpNLP2* and -4 have 14.



FIGURE 4.1: Logos of identified motifs

FIGURE 4.2: Consensus motifs of *AtNLPs* and *TpNLPs*

## 4.5 Gene Structure

The structural properties of *AtNLPs* and *TpNLPs*, such as the identification of exons, introns, and un-translated regions (UTRS), were analyzed using their gene and coding sequences. An online server known as the Gene Structure Display Server (<http://gsds.gao-lab.org/>) was used for this.

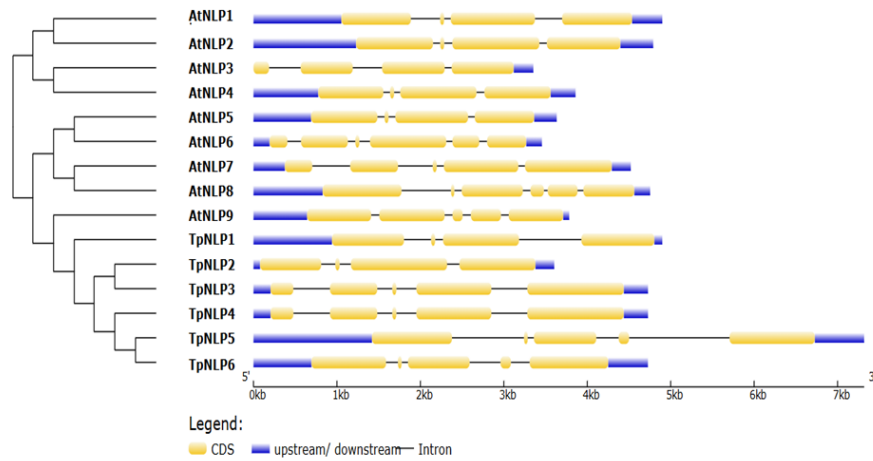


FIGURE 4.3: Display the result of Gene structure determination

## 4.6 Protein-Protein Interaction

Proteins do not function in isolation; they operate as part of a collective. The network was forecasted online via STRING (<https://www.expasy.org/resources/string>). It has been suggested that all TpNLP proteins interact with various N-interacted genes. The color lines indicate both known and predicted interactions, with pink and blue representing known interactions, while dark green, blue, and red denote predicted interactions. The various colored balls represent nodes, while the colored lines signify the edges, indicating protein sequences. In PPI, the proteins that interact with NLP are those that play a role in nitrogen uptake, assimilation, and transport, thereby indicating their contribution to overall nitrogen use efficiency.

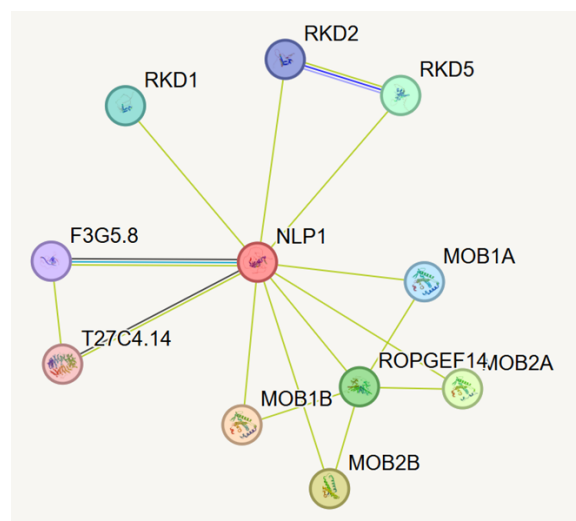
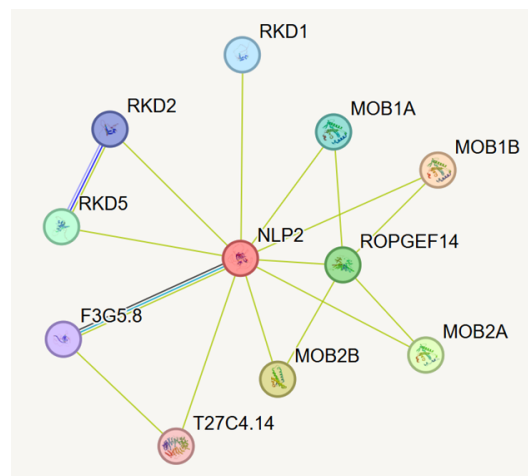


FIGURE 4.4: Protein Protein interaction of *AtNLP1*

TABLE 4.3: Protein Protein interaction of *AtNLP1*

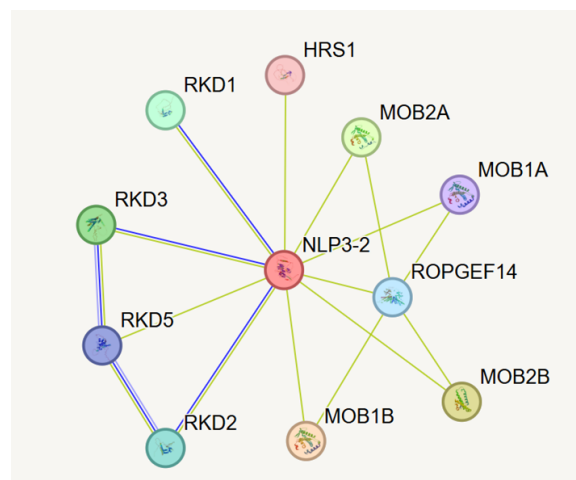
Predicted Functional Partners	
MOB1B	<i>MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP, hence activating Rop (plant Rho) GTPases.</i>
RKD5	<i>Protein RKD5; Putative transcription factor.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
MOB1A	<i>MOB kinase activator-like 1A is essential for controlling cell division and proliferation.</i>
RKD2	<i>Protein RKD2; Putative transcription factor.</i>
F3G5.8	<i>Ypt/Rab-GAP domain of gyp1p superfamily protein.</i>
T27C4.14	<i>Exportin-4 protein.</i>

FIGURE 4.5: Protein Protein interaction of *AtNLP2*TABLE 4.4: Protein Protein interaction of *AtNLP2*

Predicted Functional Partners	
MOB1B	<i>MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP, hence activating Rop (plant Rho) GTPases.</i>

Table 4.4 continued from previous page

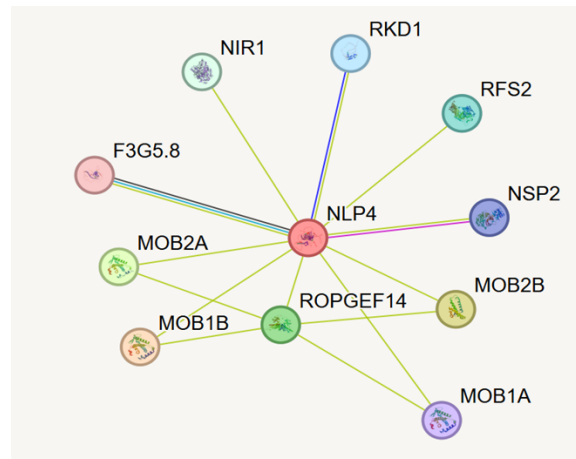
Predicted Functional Partners	
RKD5	<i>Protein RKD5; Putative transcription factor.</i>
MOB1A	<i>MOB kinase activator-like 1A is essential for controlling cell division and proliferation.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
RKD2	<i>Protein RKD2; Putative transcription factor.</i>
F3G5.8	<i>Ypt/Rab-GAP domain of gyp1p superfamily protein.</i>
T27C4.14	<i>Exportin-4 protein.</i>

FIGURE 4.6: Protein Protein interaction of *AtNLP3*TABLE 4.5: Protein Protein interaction of *AtNLP3*

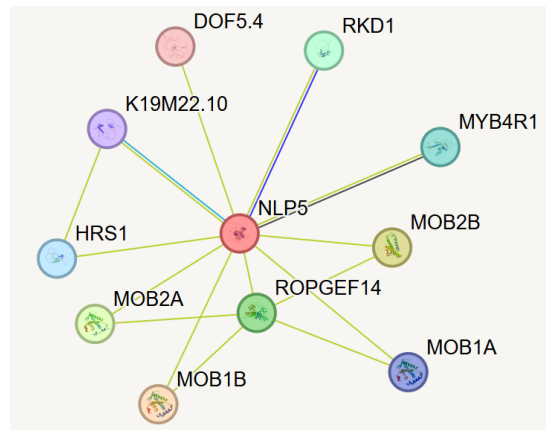
Predicted Functional Partners	
MOB1B	<i>MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
RKD3	<i>Protein RKD3; Putative transcription factor.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
RKD2	<i>Protein RKD2; Putative transcription factor.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP, hence activating Rop (plant Rho) GTPases.</i>
RKD5	<i>Protein RKD5; Putative transcription factor.</i>
MOB1A	<i>MOB kinase activator-like 1A is essential for controlling cell division and proliferation.</i>

Table 4.5 continued from previous page

Predicted Functional Partners	
HRS1	<i>Transcription factor HRS1; Transcription factor involved in nitrate and phosphate signaling in roots.</i>

FIGURE 4.7: Protein Protein interaction of *AtNLP4*TABLE 4.6: Protein-Protein interaction of *AtNLP4*

Predicted Functional Partners	
MOB1B	<i>MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP, hence activating Rop (plant Rho) GTPases.</i>
NIR1	<i>The six-electron reduction of nitrite to ammonium is catalyzed by ferredoxin-nitrite reductase, a chloroplastic.</i>
RFS2	<i>The ping-pong reaction mechanism is used by transglycosidase and likely galactinol-sucrose galactosyltransferase 2.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
NSP2	<i>Nitrile-specifier protein 2; Promotes simple nitriles, but not epithionitrile or thiocyanate formation.</i>
MOB1A	<i>MOB kinase activator-like 1A is essential for controlling cell division and proliferation.</i>
F3G5.8	<i>Ypt/Rab-GAP domain of gyp1p superfamily protein.</i>

FIGURE 4.8: Protein Protein interaction of *AtNLP5*TABLE 4.7: Protein Protein interaction of *AtNLP5*

Predicted Functional Partners	
MOB1B	<i>MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
MYB4R1	<i>Putative transcription factor.</i>
HRS1	<i>Transcription factor HRS1; Transcription factor involved in nitrate and phosphate signaling in roots.</i>
MOB1A	<i>MOB kinase activator-like 1A is essential for controlling cell division and proliferation.</i>
DOF5.4	<i>A transcription factor known as Dof zinc finger protein DOF5.4 selectively binds to a consensus core sequence of 5'-AA[AG]G-3'.</i>

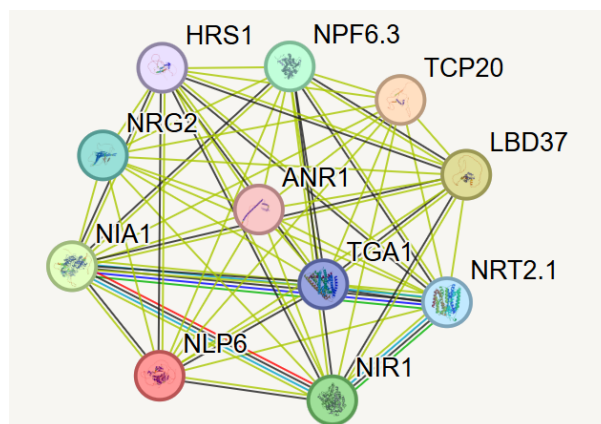
FIGURE 4.9: Protein Protein interaction of *AtNLP6*

TABLE 4.8: Protein Protein interaction of *AtNLP6*

Predicted Functional Partners	
TCP20	<i>Transcription factor TCP20: This transcription factor attaches itself to the 3'-GCCCG/A-5' elements in the promoters of cyclin CYCB1-1 and ribosomal protein genes, as well as to the site II motif (3'- TGGGCC/T-5') in the promoter of PCNA-2.</i>
LBD37	<i>LOB domain-containing protein 37.</i>
NIA1	<i>One important enzyme involved in the initial stage of nitrate assimilation in bacteria, fungi, and plants is nitrate reductase [NADH] 1.</i>
NIR1	<i>Ferredoxin–nitrite reductase, a chloroplastic, catalyzes the six-electron reduction of nitrite to ammonium.</i>
NPF6.3	<i>Protein NRT1/PTR Family 6.3: Nitrate transporter with dual affinity.</i>
NRG2	<i>Nitrate regulatory gene2 protein; Required for nitrate signaling.</i>
NRT2.1	<i>Although it is involved in nitrate transport, high-affinity nitrate transporter 2.1 does not appear to be capable of mediating transport on its own.</i>
TGA1	<i>The transcription factor TGA1 is a transcriptional activator that binds to the 5'-TGACG-3' DNA sequence.</i>
HRS1	<i>HRS1 is a transcription factor that has a role in the signaling of phosphate and nitrate in roots.</i>
ANR1	<i>Probable transcription factor; MADS-box transcription factor ANR1. necessary for the nitrate-responsive plasticity of roots.</i>

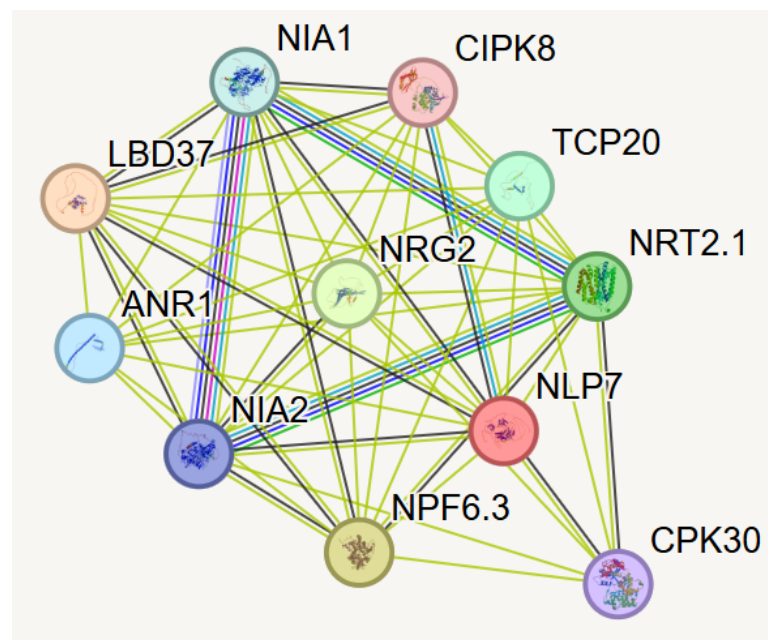
FIGURE 4.10: Protein Protein interaction of *AtNLP7*

TABLE 4.9: Protein Protein interaction of *AtNLP7*

Predicted Functional Partners	
LBD37	<i>LOB domain-containing protein 37.</i>
NPF6.3	<i>Protein NRT1/PTR Family 6.3: Nitrate transporter with dual affinity.</i>
NRG2	<i>Nitrate regulatory gene2 protein; Required for nitrate signaling.</i>
NRT2.1	<i>Although it is involved in nitrate transport, high-affinity nitrate transporter 2.1 does not appear to be capable of mediating transport on its own.</i>
TCP20	<i>Transcription factor TCP20: This transcription factor attaches itself to the 3'-GCCCG/A-5' elements in the promoters of cyclin CYCB1-1 and ribosomal protein genes, as well as to the site II motif (3'- TGGGCC/T-5') in the promoter of PCNA-2.</i>
NIA1	<i>One important enzyme involved in the initial stage of nitrate assimilation in bacteria, fungi, and plants is nitrate reductase [NADH] 1.</i>
ANR1	<i>Probable transcription factor; MADS-box transcription factor ANR1. necessary for the nitrate-responsive plasticity of roots.</i>
NIA2	<i>One important enzyme in the initial stage of nitrate assimilation is nitrate reductase [NADH] 2.</i>
CPK30	<i>Protein kinase 30 that is dependent on calcium may be involved in signal transduction pathways where calcium is used as a second messenger.</i>
CIPK8	<i>CBL proteins interact with CIPK serine-threonine protein kinases and CBL-interacting serine/threonine-protein kinase 8.</i>

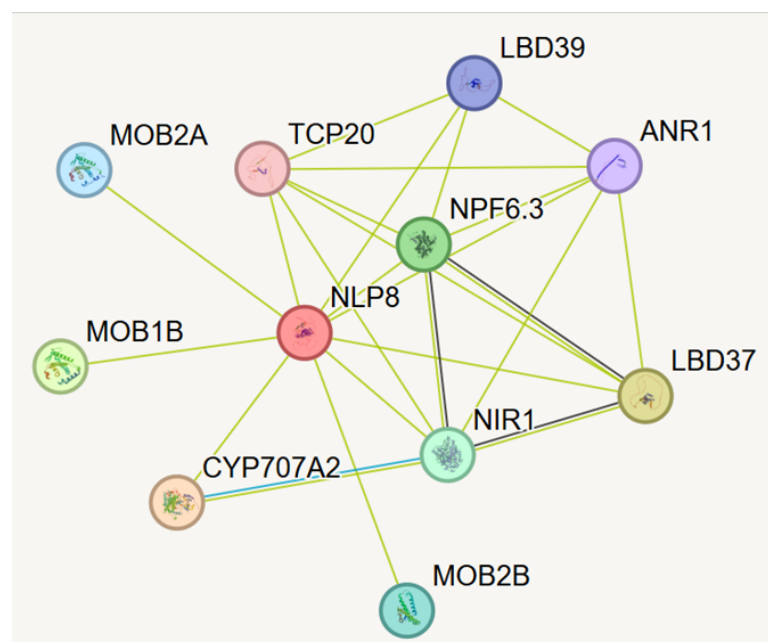
FIGURE 4.11: Protein Protein interaction of *AtNLP8*

TABLE 4.10: Protein Protein interaction of *AtNLP8*

Predicted Functional Partners	
CYP707A2	<i>Abscisic acid 8'-hydroxylase 2: The enzyme that breaks down abscisic acid oxidatively,</i>
LBD37	<i>LOB domain-containing protein 37.</i>
MOB1B	<i>Activator-like MOB kinase 1B is a member of the MOB1/phocein family.</i>
NPF6.3	<i>Protein NRT1/ PTR FAMILY 6.3; Dual affinity nitrate transporter. Involved in proton- dependent nitrate uptake and in the regulation of the nitrate transporter NRT2.1</i>
NIR1	<i>Ferredoxin-nitrite reductase, chloroplastic; Catalyzes the six-electron reduction of nitrite to ammonium.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
LBD39	<i>LOB domain-containing protein 39.</i>
ANR1	<i>Probable transcription factor; MADS-box transcription factor ANR1. necessary for the nitrate-responsive plasticity of roots.</i>
TCP20	<i>Transcription factor TCP20: This transcription factor attaches itself to the 3'-GCCCG/A-5' elements in the promoters of cyclin CYCB1-1 and ribosomal protein genes, as well as to the site II motif (3'- TGGGCC/T-5') in the promoter of PCNA-2.</i>

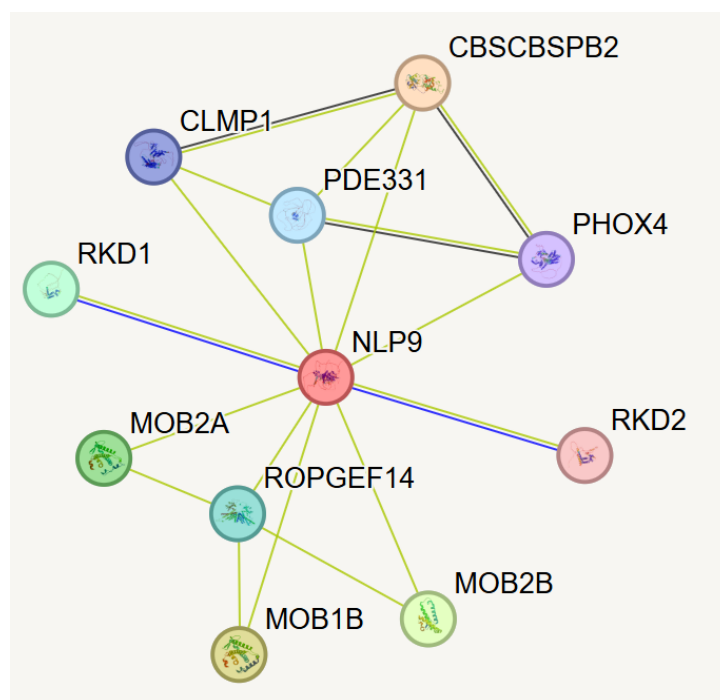
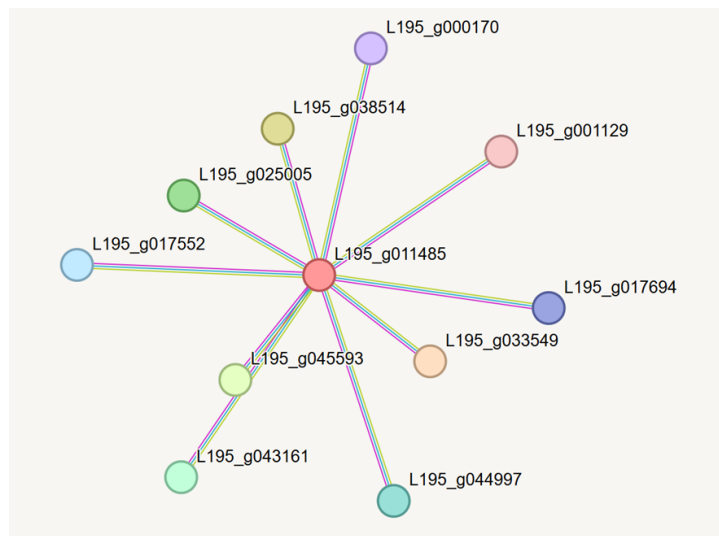
FIGURE 4.12: Protein Protein interaction of *AtNLP9*

TABLE 4.11: Protein Protein interaction of *AtNLP9*

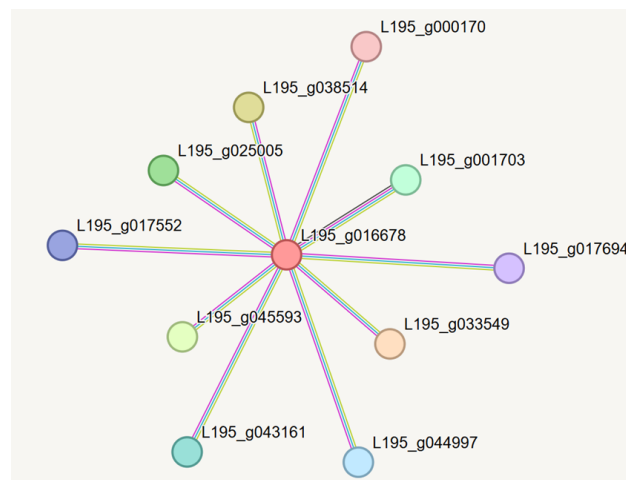
Predicted Functional Partners	
CBSCBSPB2	<i>CBS domain-containing protein CBSCBSPB2.</i>
MOB1B	<i>Activator-like MOB kinase 1B is a member of the MOB1/phocein family.</i>
MOB2B	<i>Putative MOB kinase activator-like 2B.</i>
MOB2A	<i>MOB kinase activator-like 2A.</i>
RKD1	<i>Protein RKD1; Putative transcription factor.</i>
ROPGEF14	<i>Guanine-nucleotide exchange factor (GEF), also known as Rop guanine nucleotide exchange factor 14, facilitates the conversion of GDP into GTP, hence activating Rop (plant Rho) GTPases.</i>
PDE331	<i>Octicosapeptide/Phox/Bem1p family protein.</i>
CLMP1	<i>Protein CLMP1; Required for plastid separation and partitioning during cell division.</i>
RKD2	<i>Protein RKD2; Putative transcription factor.</i>
PHOX4	<i>Potential Hsp90/Hsp70 co-chaperone PHOX4 is a carboxylate clamp-type tetratricopeptide repeat protein. Causes root hairs to develop polarly.</i>

FIGURE 4.13: Protein Protein interaction of *TpNLP1*TABLE 4.12: Protein Protein interaction of *TpNLP1*

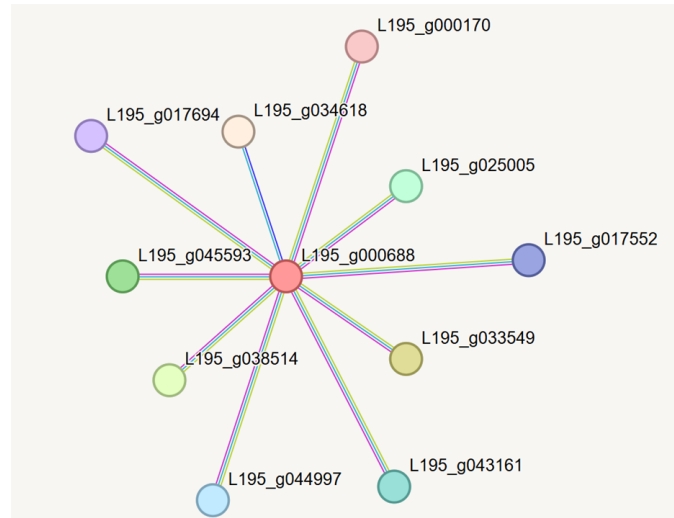
Predicted Functional Partners	
L195_g033549	<i>Mitochondria fission 1 protein</i>
L195_g038514	<i>Uncharacterized protein.</i>
L195_g045593	<i>Uncharacterized protein</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>

Table 4.12 continued from previous page

Predicted Functional Partners	
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein.</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>
L195_g000170	<i>Beta-adaptin C-like protein.</i>
L195_g001129	<i>A subunit of the clathrin-associated adaptor protein complex, beta-adaptin-like protein aids in the sorting of proteins in endosomes and/or the late-Golgi/trans-Golgi network (TGN).</i>

FIGURE 4.14: Protein Protein interaction of *TpNLP2*TABLE 4.13: Protein–Protein interaction of *TpNLP2*

Predicted Functional Partners	
L195_g033549	<i>Mitochondria fission 1 protein.</i>
L195_g045593	<i>Uncharacterized protein.</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>
L195_g001703	<i>Cellulose synthase A catalytic subunit 1.</i>
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein.</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>
L195_g000170	<i>Beta-adaptin C-like protein.</i>

FIGURE 4.15: Protein Protein interaction of *TpNLP3*TABLE 4.14: Protein Protein interaction of *TpNLP3*

Predicted Functional Partners	
L195_g034618	<i>E3 ubiquitin-protein ligase prt1-like protein</i>
L195_g033549	<i>Mitochondria fission 1 protein.</i>
L195_g038514	<i>Uncharacterized protein.</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein.</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>
L195_g000170	<i>Beta-adaptin C-like protein.</i>

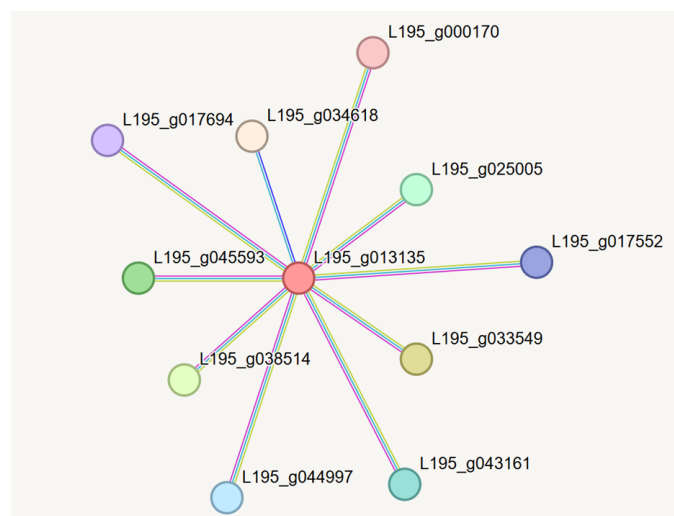
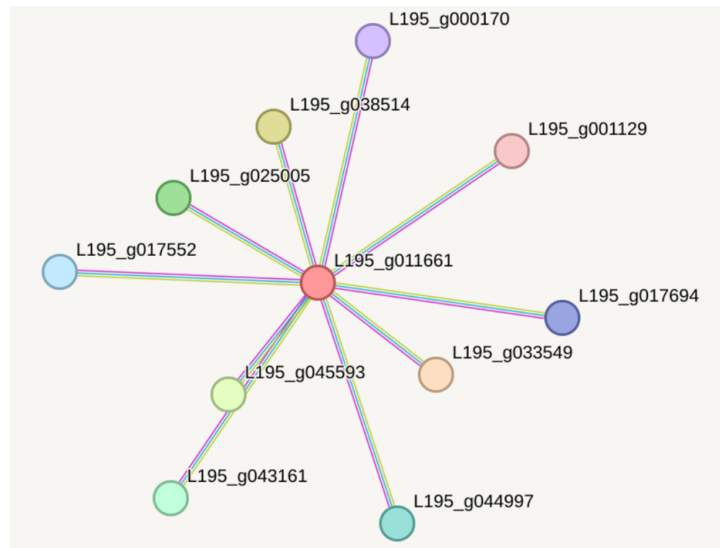
FIGURE 4.16: Protein Protein interaction of *TpNLP4*

TABLE 4.15: Protein Protein interaction of *TpNLP4*

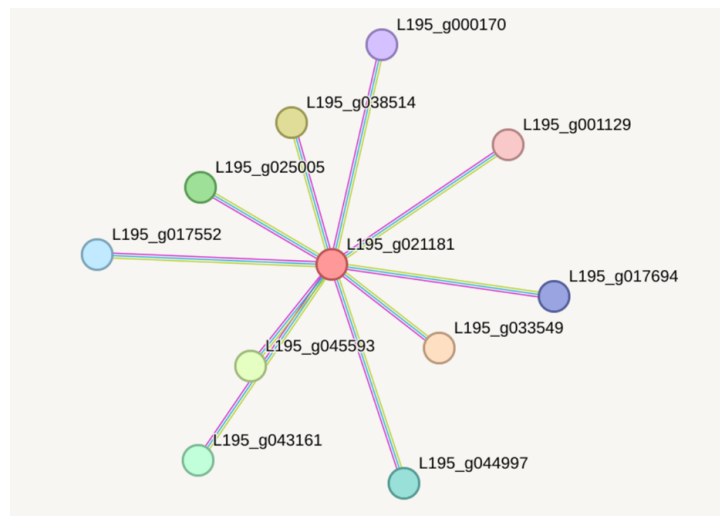
Predicted Functional Partners	
L195_g034618	<i>E3 ubiquitin-protein ligase prt1-like protein.</i>
L195_g033549	<i>Mitochondria fission 1 protein.</i>
L195_g038514	<i>Uncharacterized protein.</i>
L195_g045593	<i>Uncharacterized protein.</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>
L195_g000170	<i>Beta-adaptin C-like protein</i>

FIGURE 4.17: Protein Protein interaction of *TpNLP5*TABLE 4.16: Protein Protein interaction of *TpNLP5*

Predicted Functional Partners	
L195_g033549	<i>Mitochondria fission 1 protein.</i>
L195_g038514	<i>Uncharacterized protein</i>
L195_g045593	<i>Uncharacterized protein.</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein.</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>

Table 4.16 continued from previous page

Predicted Functional Partners	
L195_g000170	<i>Beta-adaptin C-like protein.</i>
L195_g001129	<i>A subunit of the clathrin-associated adaptor protein complex, beta-adaptin-like protein aids in the sorting of proteins in endosomes and/or the late-Golgi/trans-Golgi network (TGN).</i>

FIGURE 4.18: Protein Protein interaction of *TpNLP6*TABLE 4.17: Protein Protein interaction of *TpNLP6*

Predicted Functional Partners	
L195_g033549	<i>Mitochondria fission 1 protein.</i>
L195_g038514	<i>Uncharacterized protein.</i>
L195_g045593	<i>Uncharacterized protein.</i>
L195_g025005	<i>Mitochondria fission 1 protein.</i>
L195_g043161	<i>Beta-adaptin A-like protein.</i>
L195_g044997	<i>Ddb1-and cul4-associated factor-like protein.</i>
L195_g017552	<i>Beta-adaptin A-like protein.</i>
L195_g017694	<i>Beta-adaptin C-like protein.</i>
L195_g000170	<i>Beta-adaptin C-like protein.</i>
L195_g001129	<i>A subunit of the clathrin-associated adaptor protein complex, beta-adaptin-like protein aids in the sorting of proteins in endosomes and/or the late-Golgi/trans-Golgi network (TGN).</i>

## 4.7 The *TpNLPs* Gene Family's Evolutionary Relationships and Sequence Alignment

To verify the proper selection and uniqueness of each found *TpNLP* gene used for additional analysis, the percentage similarities between *TpNLPs* and *AtNLPs* were compared (Table 4.19). Less than 78% of the protein sequences of the *AtNLPs* and *TpNLPs* were comparable, ensuring both the evolutionary variety of the *TpNLP* gene family and the uniqueness of each gene. In MEGA-X v10.1.8, a rooted neighbor-joining phylogenetic tree was constructed using the alignment output of the *TpNLP* gene family and the *NLP* gene families of *Zea mays*, *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor*, *Solanum lycopersicum*, *Physcomitrella patens*, *Brassica napus*, *Populus trichocarpa*, and *Triticum aestivum*.

In MEGA-X v10.1.8, a rooted neighbor-joining phylogenetic tree was constructed using the alignment output of the *TpNLP* gene family and the *NLP* gene families of *Zea mays*, *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor*, *Solanum lycopersicum*, *Physcomitrella patens*, *Brassica napus*, *Populus trichocarpa*, and *Triticum aestivum*.

The tree was then redesigned using itol ([itol.embl.de/upload.cgi](http://itol.embl.de/upload.cgi)). The phylogenetic tree was used to examine the evolutionary relationships between *TpNLPs* and various species that have *NLP* gene families.

The tree was then redesigned using itol ([itol.embl.de/upload.cgi](http://itol.embl.de/upload.cgi)). The phylogenetic tree was used to examine the evolutionary relationships between *TpNLPs* and various species that have *NLP* gene families. All evolutionary trees show divergence in the *Trifolium pratense* (Fig 4.6). *Trifolium pratense*'s *NLP* gene family demonstrated evolutionary relationships with the other nine plants that were chosen, and it is anticipated that *Populus trichocarpa* will be more comparable to them structurally and functionally. The phylogenetic tree shows that all *NLPs* share a common root, indicating that they evolved from one another. Sister evolutionary divergence of *NLP* genes is also seen.

	AtNLP1	AtNLP2	AtNLP3	AtNLP4	AtNLP5	AtNLP6	AtNLP7	AtNLP8	AtNLP9	TpNLP1	TpNLP2	TpNLP3	TpNLP4	TpNLP5	TpNLP6
AtNLP1	100	68.66	44.13	46.39	41.26	35.88	36.54	37.5	37.08	41.18	39.21	32.52	37.32	36.7	38.25
AtNLP2	69.08	100	45.88	40.8	39.31	35.22	40.28	35.98	36.54	42.18	39.74	35.63	36.01	36.7	36.63
AtNLP3	44.13	45.43	100	39.17	40.58	31.53	33.99	32.66	33.65	40.72	40.37	33.38	32.42	36.42	34.1
AtNLP4	41.35	41.08	43.94	100	70.57	35.58	42.39	40.11	39.26	48.93	38.99	36.91	39.93	41.82	42.68
AtNLP5	45.04	43.24	41.25	69.33	100	36.3	44.19	38.46	39.67	45.89	43.37	37.16	40.6	41.13	42.44
AtNLP6	39.63	39.58	34.1	34.32	35.59	100	60.14	38.83	40.26	42.57	32.2	54.33	60.54	40.97	41.12
AtNLP7	37.27	40.28	34.99	42.75	44.19	60.21	100	38.53	40.37	42.91	33.07	59.49	53.16	39.89	39.22
AtNLP8	37.5	35.98	32.82	39.38	38.46	34.4	38.53	100	64.69	37.01	35.07	35.66	38.15	46.69	47.06
AtNLP9	37.08	36.54	32.94	38.53	39.67	40.26	40.37	65.14	100	32.29	35.62	39.93	38.03	49.42	53.55
TpNLP1	42.43	41.75	38.34	48.93	46.44	37.67	35.49	37.01	32.61	100	40.16	38.09	40.47	39.17	38.06
TpNLP2	39.35	40.65	30.23	38.99	38.22	32.2	32.66	33.44	33.12	36.14	100	32.73	34.04	32.71	28.27
TpNLP3	32.39	38.94	32.03	34.99	37.31	54.21	59.11	35.5	39.75	40.99	32.47	100	65.1	36.94	54.21
TpNLP4	37.32	36.01	33.57	39.87	40.6	60.75	53.28	35.97	38.03	40.64	33.33	65.81	100	38.26	39.39
TpNLP5	36.7	36.7	36.03	41.45	41.13	40.97	40.04	46.9	49.42	39.17	32.89	37.1	38.26	100	59.41
TpNLP6	38.25	36.63	34.82	42.13	42.63	41.12	38.81	47.27	53.42	39.53	29.58	39.78	39.39	59	100

FIGURE 4.19: Sequence similarity between *Trifolium pratense* and *Arabidopsis thaliana* NLP proteins (similarity greater than 60% is indicated with red font color).

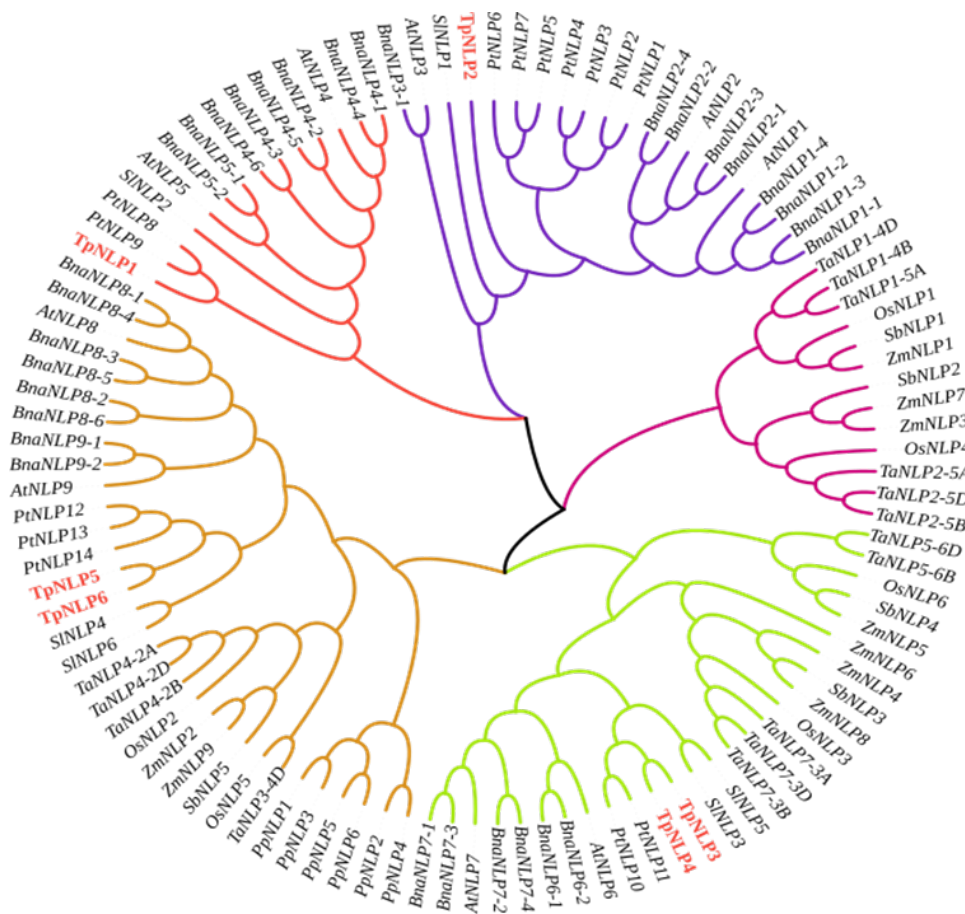


FIGURE 4.20: Phylogenetic tree

## 4.8 Cis-regulatory elements

One important method for speculating about the function and regulation of genes is the identification of cis-regulatory elements in upstream promoter regions (2000 bp). To classify the found cis-regulatory elements in the promoter regions of both *AtNLPs* and *TpNLPs*, three groups were developed: phytohormone (PR), stress (SR), and plant growth and development (PGD), as indicated in Table 4.18. In contrast to *TpNLPs*, *AtNLPs* have a greater number of regulatory elements. The *AtNLPs* cis-elements that responded to phytohormones were the most numerous (87), whereas the *AtNLPs* cis-elements that responded to SR and PGD were 45 and 46, respectively (table 4.18). With the exception of *AtNLP7*, which had more PGD responsive cis-elements than SR and PR, all *AtNLPs* had more PR cis-elements. *TpNLPs* have been found to include 30, 30, and 39 PGD, SR, and PR cis-elements in total.

TABLE 4.18: Number of cis-regulatory elements identified in promoter regions of AtNLPs and TpNLPs Gene Families.

Gene	Plant Growth & Development								Stress Responsive					Phytohormone Responsive									
	Box 4	MRE	CAT-Box	O2-site	circadian	GCN4_Motif	MSA-Like	WUN-Motif	ARE	MBS	TC-rich repeats	LTR	GC-motif	CGTCA-Motif	TGACG-Motif	GARE-Motif	P-Box	TATC-Box	ABRE	ERE	TGA-element	TCA-element	
AtNLP1		1		2	1				3	2	1			1	1				3			1	1
AtNLP2	1			1					5	1	1			6	6		1	1	2			1	
AtNLP3	3	1		1					2		1						1	1	5			1	1
AtNLP4	2	1	1						1		1							1	1		3		1
AtNLP5	5	1	1						3	1				1	1	1			2				1
AtNLP6	1	1					1		2		1	1		2	2		2		1			1	1
AtNLP7	6	2	1		1			1	4		3	1					1	2	2		1	2	
AtNLP8	2		2						1	3		1		3	3				1				1
AtNLP9	4	1				1	1		2					2	2		1		5		2	1	4
TpNLP1	3								1	3		1	1						1				
TpNLP2	3	1			1				1	1	1	1		2	2	2	2		1		1		
TpNLP3	1	1	2					3	2								1				4		2
TpNLP4	1	1				1			4	1				1	1				2				
TpNLP5	6	1			1			3	4								2				2		
TpNLP6	6		1						2		1			1	1	1			5		1		

# Chapter 5

## Discussion

In order for plants to develop properly, they require nitrogen since it enables them to produce energy, proteins, and chlorophyll. The efficiency with which plants use nitrogen, also known as nitrogen usage efficiency (NUE), has a significant impact on the efficiency with which farms function and the integrity of the environment. *Trifolium pratense*, often known as red clover, is a legume that is significant for both farming and the environment. This study investigated the Nodule Inception-Like Protein (NLP) gene family in *Trifolium pratense*. As a result of the findings, it appears that NLP genes have the potential to enhance NUE and offer us fresh perspectives on the ways in which plants regulate nitrogen.

NLP is a family of genes that is well-known for its function in nitrogen signaling and assimilation, particularly in situations where nitrate is already present. From *TpNLP1* to *TpNLP6*, researchers discovered six different NLP genes in red clover. Every single one of these genes possessed the RWP-RK and PB1 domains, which are essential for the binding of DNA and the interaction of proteins. These *TpNLPs* have been shown to be related to NLP genes from other plants, such as *Arabidopsis thaliana*, *Zea mays*, and *Oryza sativa*, according to a phylogenetic research. By demonstrating that the manner in which plants react to nitrogen are comparable across a wide range of species, this conservation underscores the significance of nitrogen-loving proteins (NLPs) in the process of nitrogen metabolism.

After doing structural studies, it was discovered that *TpNLPs* possess the same molecular weight, isoelectric point, and hydrophobicity as their counterparts found in Arabidopsis. The nucleus, which is the location where transcription factors are present, was the place where it was anticipated that all *TpNLPs* would be located. The fact that these proteins contain a number of motifs that are conserved makes it even more plausible that they play a regulatory role. Depending on the differences between these motifs, it is possible that the *TpNLPs* serve distinct purposes. In order to appreciate the manner in which these proteins regulate nitrogen-related functions through their interactions with other molecules, it is necessary to have an understanding of their structural features.

The extent to which NLP genes influence the rates at which plants take in nitrogen is the subject of a great deal of research. Due to the fact that it regulates genes that are involved in the process of absorbing and using nitrate, AtNLP7 plays a significant role in the nitrate signaling process in Arabidopsis. In a similar fashion, *TpNLPs* are likely to perform comparable tasks in red clover due to the fact that their structure and function are comparable to those of AtNLPs. Predictions of protein-protein interactions revealed that *TpNLPs* interact with proteins that assist the body in absorbing, transporting, and utilizing nitrogen. These proteins include nitrate transporters and nitrate reductases, among others. The fact that these interactions take place demonstrates that *TpNLPs* are significant components of the nitrogen regulatory network, which assists the plant in coping with variations in the amount of nitrogen present in the soil.

Taking a look at the promoter regions of *TpNLPs* revealed that there are a number of cis-regulatory components that react to phytohormones, stress, and the growth and development of plants. Considering that *TpNLPs* contain components that are associated with abscisic acid and auxin responses, it is possible that these proteins are responsible for connecting hormone pathways with nitrogen signaling. Elements that are responsive to stress imply that these genes may be able to assist plants in coping with issues that are present in their environment, whereas elements that are related to growth imply that these genes are involved in the processes that are associated with growth. Because there are so many different

regulatory components, it is likely that *TpNLPs* are engaged in more than only nitrogen metabolism. This is because there are so many different regulatory elements. It is likely that they make a connection between nitrogen signaling and the way in which plants react to their surroundings in general.

In order to better understand the context of *TpNLPs*, it is helpful to compare them to NLP genes from other species. For instance, it has been discovered that *OsNLP1* in rice can boost nitrogen use efficiency (NUE) and grain yield even when there is a lack of nitrogen. A number of well-known NIN LIKE PROTEIN (NLPs) have functional domains that are comparable to those of the *TpNLPs*. In light of this, it appears that the systems that govern nitrogen are stable. Differences in the structure and expression patterns of genes, on the other hand, demonstrate that different species have adapted to meet their unique requirements. When compared to non-leguminous plants such as *Arabidopsis*, the NLP gene family found in red clover is significantly larger. It's possible that this is due to the evolutionary changes that took place when legumes created symbiotic partnerships with bacteria that fix nitrogen. It can be deduced from this that red clover has developed incredibly intricate mechanisms to regulate not only the process of nitrogen fixation that takes place in symbiosis but also the process of nitrogen absorption from the soil.

From an agricultural perspective, improving NUE is essential for long-term farming since it helps farming to be more sustainable. It is now conceivable to utilize genetic approaches to improve the way that red clover and other crops that are comparable to it use nitrogen because to the characterization of *TpNLPs*. Over-expressing certain *TpNLP* genes in order to boost nitrogen absorption is one approach that might be taken. Altering the regulatory roles of these genes through gene editing is another possibility. Because of its ecological benefits, such as the ability to fix nitrogen and improve soil quality, red clover is an excellent choice for sustainable agriculture. When farmers have a greater understanding of how *TpNLPs* function, they may be able to make better use of this plant in crop rotations or intercropping systems, which will allow them to reduce their reliance on synthetic fertilizers while still achieving satisfactory yields.

# Chapter 6

## Conclusion and Future Work

### 6.1 Conclusion

Based on what we found in this study and what other studies have found, we think that *Trifolium pratense* *NLP* genes (*TpNLPs*) respond to nitrogen levels and probably have a big influence in how efficiently plants use nitrogen (NUE). Similar to transcription factors in other plant species, *TpNLPs* are an interesting group that could be targeted to make crops use nitrogen better. Our *in silico* research gives us useful information about their structure and their functions, but it mostly sets up a theoretical framework that needs to be tested more in the lab.

Our research fills in an essential gap in the knowledge of how red clover, a crucial legume for farming and the environment, controls nitrogen. This study is different from others that looked at model plants like *Arabidopsis* or cereals. It focuses on the specific ways that nitrogen signaling works in a nitrogen-fixing perennial crop. The conserved domains and projected protein interactions suggest that *TpNLPs* work in a way similar to known *NLPs*. However, more research is needed to find out what precise roles they play in red clover's symbiotic nitrogen fixation system.

This study raises a number of important questions that future research should answer. First and foremost, we need to do mutant studies to get a better understanding of the molecular processes by which *TpNLPs* control nitrogen responses.

It would be very useful for this goal to create gene-edited red clover lines. Also, comparing *TpNLPs* to *NLPs* from other legumes could show both general and specific roles in fixing nitrogen.

Red clover is a great plant to examine how nitrogen moves and is used because it can get nitrogen from the soil and fix it with symbiotic bacteria. This special feature makes it especially fascinating for figuring out how plants use these multiple nitrogen sources together. Future studies should look at whether *TpNLPs* help balance these pathways. This could have big effects on how well nitrogen works in farming.

This research could be useful in a lot of different ways. If we can figure out how *TpNLPs* control nitrogen utilization, we might be able to create new types of red clover that use nitrogen more efficiently, which would mean we wouldn't need as many synthetic fertilizers. Also, what we learn from studying these transcription factors in red clover could help us make similar changes in other crops, which would lead to more environmentally friendly farming around the world.

Our computer analysis gives us a good place to start, but the actual value of our research will come from testing it out in the real world and doing field surveys. The next stages should be to do expression analysis under different nitrogen levels, validate protein interactions, and finally, use targeted genetic techniques to make better red clover types. By making the best use of nitrogen in one of agriculture's most valuable forage crops, these efforts could have real benefits for farmers and the environment.

## 6.2 Future Direction

In the future, this discovery opens up a lot of significant research paths that will help us learn more about how *Trifolium pratense* controls nitrogen. First and foremost, we need to use gene editing and transgenic methods to test the functionality of *TpNLPs* in the lab. Making knockout or overexpression lines would show exactly what their responsibilities are in taking in and using nitrogen.

Then, field trials might see if the changed plants do better in low-nitrogen settings, which could lead to more efficient crop types.

Another important area of study is how *TpNLPs* affect the nitrogen-fixing system that red clover uses to help other plants grow. This legume can get nitrogen from the soil and from bacteria that live in its roots. Learning how *TpNLPs* control this balance could help make natural nitrogen fixation work well. This could lead to clover types that use less fertilizer but yet produce a lot.

The bigger picture is just as thrilling. Finding out how *TpNLPs* work with similar genes in other crops could help us find ways to control nitrogen that work in all crops. Also, looking into how *TpNLPs* and stress responses are related could help us understand how nitrogen use efficiency is linked to disease or drought resistance.

These future experiments, which will combine lab work with field work, could help create crops that use nitrogen more efficiently, which would increase yields and lower the impact on the environment. This is a very useful area of research to pursue because of the possible benefits for sustainable agriculture.

# Bibliography

- [1] P. A. Matson, P. M. Vitousek, J. J. Ewel, M. J. Mazzarino, and G. P. Robertson, “Nitrogen transformations following tropical forest felling and burning on a volcanic soil,” *Ecology*, vol. 68, no. 3, pp. 491–502, 1987.
- [2] M. Tegeder and C. Masclaux-Daubresse, “Source and sink mechanisms of nitrogen transport and use,” *New Phytol.*, vol. 217, no. 1, pp. 35–53, 2018.
- [3] A. Alfatih, “Rice NIN-LIKE PROTEIN 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency,” *Journal of Experimental Botany*, vol. 71, no. 19, pp. 6032–6042, 2020.
- [4] J. Liu, X. Yu, Q. Qin, R. D. Dinkins, and H. Zhu, “The impacts of domestication and breeding on nitrogen fixation symbiosis in legumes,” *Front. Genet.*, vol. 11, p. 00973, 2020.
- [5] E. T. Wang, C. F. Tian, W. F. Chen, J. P. W. Young, W. X. Chen, and E. T. Wang, “Symbiosis between rhizobia and legumes. Ecology and evolution of rhizobia: Principles and applications,” pp. 3–19, 2019.
- [6] K. A. Congreves, O. Otchere, D. Ferland, S. Farzadfar, S. Williams, and M. M. Arcand, “Nitrogen use efficiency definitions of today and tomorrow,” *Front. Plant Sci.*, vol. 12, p. 637108, 2021.
- [7] P. Govindasamy et al., “Nitrogen use efficiency-a key to enhance crop productivity under a changing climate,” *Front. Plant Sci.*, vol. 14, p. 1121073, 2023.
- [8] J. Erisman, A. Leach, A. Bleeker, B. Atwell, L. Cattaneo, and J. Galloway, “An integrated approach to a nitrogen use efficiency (NUE) indicator for the

- food production–consumption chain,” *Sustainability*, vol. 10, no. 4, p. 925, 2018.
- [9] R. Blanc-Mathieu, R. Dumas, L. Turchi, J. Lucas, and F. Parcy, “Plant-TFClass: a structural classification for plant transcription factors,” *Trends Plant Sci.*, vol. 29, no. 1, pp. 40–51, 2024.
- [10] J. M. Alvarez et al., “Systems approach identifies TGA1 and TGA4 transcription factors as important regulatory components of the nitrate response of *Arabidopsis thaliana* roots,” *Plant J.*, vol. 80, no. 1, pp. 1–13, 2014.
- [11] A. A. Chowdhary, S. Mishra, S. Mehrotra, S. K. Upadhyay, D. Bagal, and V. Srivastava, “Plant transcription factors: an overview of their role in plant life,” in *Plant Transcription Factors*, Elsevier, 2023, pp. 3–20.
- [12] T. Nazish et al., “Transporters and transcription factors gene families involved in improving nitrogen use efficiency (NUE) and assimilation in rice (*Oryza sativa* L.),” *Transgenic Res.*, vol. 31, no. 1, pp. 23–42, 2022.
- [13] Y. Ueda and S. Yanagisawa, “Transcription factor-based genetic engineering to increase nitrogen use efficiency. Engineering nitrogen utilization in crop plants,” pp. 37–55, 2018.
- [14] K. Lluga-Rizani, D. Šoljan, N. Berisha, K. Kurteshi, and K. Letaj, “Morphological variability of *Trifolium repens* L.(Fabaceae),” *Hacquetia*, vol. 20, no. 2, pp. 281–290, 2021.
- [15] A. D. Heslop, J. L. Ford, Z. Jahufer, and R. W. Hofmann, “From Root to Shoot: Morphological Evaluation of an International Collection of Red Clover (*Trifolium pratense* L.) Populations,” *Journal of Agronomy and Crop Science*, vol. 211, no. 3, 2025.
- [16] S. O. Solberg, F. Yndgaard, and A. Palme, “Morphological and phenological consequences of ex situ conservation of natural populations of red clover (*Trifolium pratense* L.). Plant genetic resources,” vol. 15, pp. 97–108, 2017.
- [17] “Taxon Page,” Efloras.org. [Online]. Available: <http://www.efloras.org/florataxon.aspx>? [Accessed: 11-Jul-2025].

- [18] E. D. Kubasova et al., “Pharmaceutical potential of red clover (*Trifolium pratense* L.). Ekologiya cheloveka,” *Human Ecology*, vol. 31, no. 3, pp. 179–190, 2024.
- [19] N. Prokhorenko, L. Kadyrova, and G. Demina, Reserves Of *Trifolium Pratense* And *Trifolium Repens* As Medicinal Raw Materials In Different Landscape Zones Of The Middle Volga Region.
- [20] S. Chaturvedi, N. Sultana, M. Rashid, Z. Naseem, S. K. Singh, and M. Wahajuddin, “Phytochemistry, pharmacology, and pharmacokinetics of phytoestrogens from red clover extract: An exhaustive overview,” in *Medicinal and Aromatic Plants*, Cham: Springer International Publishing, 2021, pp. 325–351.
- [21] G. P. Ramos et al., “In vivo and in vitro anti-inflammatory activity of red clover *Trifolium pratense* dry extract,” *Rev. Bras. Farmacogn.*, vol. 22, no. 1, pp. 176–180, 2012.
- [22] K. A. Wickham et al., “Short-term supplementation with fermented red clover extract reduces vascular inflammation in early post-menopausal women,” *Front. Cardiovasc. Med.*, vol. 9, p. 826959, 2022.
- [23] C. Harris and F. L. W. Ratnieks, “Clover in agriculture: combined benefits for bees, environment, and farmer,” *J. Insect Conserv.*, vol. 26, no. 3, pp. 339–357, 2022.
- [24] T. Gounden, R. Moodley, and S. B. Jonnalagadda, “Elemental analysis and nutritional value of edible *Trifolium* (clover) species,” *J. Environ. Sci. Health B*, vol. 53, no. 8, pp. 1–6, 2018.
- [25] S. Jing, P. Kryger, and B. Boelt, “Review of seed yield components and pollination conditions in red clover (*Trifolium pratense* L.) seed production,” *Euphytica*, vol. 217, no. 4, 2021.
- [26] A. Fathi, “Role of nitrogen (N) in plant growth, photosynthesis pigments, and N use efficiency: A,” *Agrisost*, vol. 28, pp. 1–8, 2022.

- [27] J. R. Evans and V. C. Clarke, “The nitrogen cost of photosynthesis,” *J. Exp. Bot.*, vol. 70, no. 1, pp. 7–15, 2019.
- [28] H. R. Heydari, E. Chamani, and B. Esmailpour, “Effect of total nitrogen content and  $\text{NH}_4^+/\text{NO}_3^-$  ratio on biomass accumulation and secondary metabolite production in cell culture of *Salvia nemorosa*,” *Iranian Journal of Genetics and Plant Breeding*, vol. 9, no. 1, pp. 17–27, 2020.
- [29] J. Y. Ye, W. H. Tian, and C. W. Jin, “Nitrogen in plants: from nutrition to the modulation of abiotic stress adaptation,” *Stress Biol.*, vol. 2, no. 1, p. 4, 2022.
- [30] W. Wei et al., “Appropriate nitrogen application enhances saponin synthesis and growth mediated by optimizing root nutrient uptake ability,” *J. Ginseng Res.*, vol. 44, no. 4, pp. 627–636, 2020.
- [31] M. I. Makarov, “The role of mycorrhiza in transformation of nitrogen compounds in soil and nitrogen nutrition of plants: A review,” *Eurasian Soil Sci.*, vol. 52, no. 2, pp. 193–205, 2019.
- [32] S. Abbas, Amna, M. T. Javed, Q. Ali, M. Azeem, and S. Ali, “Nutrient deficiency stress and relation with plant growth and development,” in *Engineering Tolerance in Crop Plants Against Abiotic Stress*, Boca Raton: CRC Press, 2021, pp. 239–262.
- [33] A. Gaudinier et al., “Transcriptional regulation of nitrogen-associated metabolism and growth,” *Nature*, vol. 563, no. 7730, pp. 259–264, 2018.
- [34] L. Luo, Y. Zhang, and G. Xu, “How does nitrogen shape plant architecture?,” *J. Exp. Bot.*, vol. 71, no. 15, pp. 4415–4427, 2020.
- [35] Q. Wang, S. Li, J. Li, and D. Huang, “The utilization and roles of nitrogen in plants,” *Forests*, vol. 15, no. 7, p. 1191, 2024.
- [36] L. Gent and B. G. Forde, “How do plants sense their nitrogen status?,” *J. Exp. Bot.*, vol. 68, no. 10, pp. 2531–2539, 2017.

- [37] M. Anas et al., “Fate of nitrogen in agriculture and environment: agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency,” *Biol. Res.*, vol. 53, no. 1, p. 47, 2020.
- [38] C. Muratore, L. Espen, and B. Prinsi, “Nitrogen uptake in plants: The plasma membrane root transport systems from a physiological and proteomic perspective,” *Plants*, vol. 10, no. 4, p. 681, 2021.
- [39] P. Omara, L. Aula, F. Oyebiyi, and W. R. Raun, “World cereal nitrogen use efficiency trends: Review and current knowledge,” *Agrosyst. Geosci. Environ.*, vol. 2, no. 1, pp. 1–8, 2019.
- [40] M. R. Yadav et al., “Strategies for improving nitrogen use efficiency: A review,” *Agric. Rev.*, no. OF, 2017.
- [41] D. W.-K. Ng, J. K. Abeysinghe, and M. Kamali, “Regulating the regulators: The control of transcription factors in plant defense signaling,” *Int. J. Mol. Sci.*, vol. 19, no. 12, p. 3737, 2018.
- [42] T. Javed, R. Shabbir, A. Ali, I. Afzal, U. Zaheer, and S.-J. Gao, “Transcription factors in plant stress responses: Challenges and potential for sugarcane improvement,” *Plants*, vol. 9, no. 4, p. 491, 2020.
- [43] Y. Ueda et al., “Gene regulatory network and its constituent transcription factors that control nitrogen-deficiency responses in rice,” *New Phytol.*, vol. 227, no. 5, pp. 1434–1452, 2020.
- [44] L. Liu, “Transcription factors and their genes in higher plants: Functional domains, evolution and regulation,” *European Journal of Biochemistry*, vol. 262, no. 2, pp. 247–257, 1999.
- [45] M. L. Sámano, K. Nanjareddy, and M.-K. Arthikala, “NIN-like proteins (NLPs) as crucial nitrate sensors: an overview of their roles in nitrogen signaling, symbiosis, abiotic stress, and beyond,” *Physiol. Mol. Biol. Plants*, vol. 30, no. 7, pp. 1209–1223, 2024.

- [46] X. Mu and J. Luo, “Evolutionary analyses of NIN-like proteins in plants and their roles in nitrate signaling,” *Cell. Mol. Life Sci.*, vol. 76, no. 19, pp. 3753–3764, 2019.
- [47] D. Yan and E. Nambara, “Conserved and unique functions of NIN-like proteins in nitrate sensing and signaling,” *Plant Sci.*, vol. 336, p. 111842, 2023.
- [48] B. Jagadhesan et al., “Genome wide analysis of NLP transcription factors reveals their role in nitrogen stress tolerance of rice,” *Sci. Rep.*, vol. 10, no. 1, p. 9368, 2020.
- [49] M. Liu et al., “Genome-wide identification and characterization of NODULE-INCEPTION-like protein (NLP) family genes in *Brassica napus*,” *Int. J. Mol. Sci.*, vol. 19, no. 8, 2018.
- [50] M. Ge, “Genome-wide analysis of maize NLP transcription factor family revealed the roles in nitrogen response,” *Plant Growth Regulation*, vol. 84, no. 1, pp. 95–105, 2021.
- [51] M. Liu, X. Zhi, Y. Wang, and Y. Wang, “Genome-wide survey and expression analysis of NIN-like Protein (NLP) genes reveals its potential roles in the response to nitrate signaling in tomato,” *BMC Plant Biol.*, vol. 21, no. 1, p. 347, 2021.
- [52] S. Mishra, G. Singroha, R. Tiwari, and P. Sharma, “Comprehensive genome-wide expression analysis of NLP transcription factors elucidates their crucial role in enhancing nitrogen response in wheat (*Triticum aestivum* L.),” *Curr. Plant Biol.*, no. 100463, p. 100463, 2025.
- [53] D. Li, Y. Jin, Q.-H. Lu, N. Ren, Y.-Q. Wang, and Q.-S. Li, “Genome-wide identification and expression analysis of NIN-like protein (NLP) genes: Exploring their potential roles in nitrate response in tea plant (*Camellia sinensis*),” *Plant Physiol. Biochem.*, vol. 207, no. 108340, p. 108340, 2024.
- [54] J. Yu, Y. Yuan, L. Dong, and G. Cui, “Genome-wide investigation of NLP gene family members in alfalfa (*Medicago sativa* L.): evolution and expression

---

profiles during development and stress,” *BMC Genomics*, vol. 24, no. 1, p. 320, 2023.