CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Functional and Evolutionary Analysis of *Nodule Inception-Like Proteins* in Soybean

by

Kainat Raja

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

I would like to dedicate my work to my beloved parents whose support and unconditional love have inspired me to reach this milestone and foundation of my every success

CERTIFICATE OF APPROVAL

Functional and Evolutionary Analysis of *Nodule Inception-Like Proteins* in Soybean

by

Kainat Raja

(MBS233001)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. M. Zeeshan Qureshi	MIU, AJK
(b)	Internal Examiner	Dr. Sohail Ahmad Jan	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, **Kainat Raja** hereby state that my MS thesis titled "**Functional and Evolutionary Analysis of *Nodule Inception-Like Proteins* in Soybean**" 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.

(Kainat Raja)

Registration No: MBS233001

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Functional and Evolutionary Analysis of *Nodule Inception-Like Proteins* in Soybean**" 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.

(Kainat Raja)

Registration No: MBS233001

Acknowledgement

Alhamdullilah, I am extremely thankful to Allah Almighty for granting me the strength, patience and perseverance to complete this thesis. Without His countless blessings and guidance, this accomplishment would not have been possible. I offer my gratitude to **Holy Prophet (PBUH)** who preached us to seek knowledge for the betterment of mankind. I owe gratitude to my University, **Capital University of Science and Technology (CUST)** Islamabad for giving me this opportunity to do MS Biosciences. A Special thanks to the Dean & Head of the Department of Bioinformatics and Biosciences **Dr. Sahar Fazal** & **Dr. Marriam Bakhtiar** for their encouragement.

I would like to express my sincere gratitude to my supervisor, **Dr. Sami Ullah Jan**, for his guidance and supervision, without which this work would not have been possible. His direction was invaluable at every stage of this research, and his involvement throughout the process is sincerely appreciated. It has been a great honor to work with him, and the experience has been truly worthwhile. I am also deeply grateful to **Dr. Sohail Ahmad Jan** for his guidance, support, and inspiration during this work.

Last but not least, I want to express my wholehearted thanks to my parents who have provided me generous support throughout my life and specially in doing my research. Because of their unconditional love, support and prayers, I have completed this thesis. I also want to acknowledge my beloved brother, Raja Umer Abdullah, for his companionship and encouragement during challenging times.

(Kainat Raja)

Abstract

Nitrogen is commonly applied to agricultural soil in the form of nitrogenous fertilizers to increase crop yields. However, plants often fail to fully utilize the applied nitrogen, leading to economic losses. Additionally, the unutilized nitrogen can leach or percolate into water reservoirs, causing environmental concerns. Since cleaning nitrogen-contaminated agricultural soil is impractical, a more feasible solution is to enhance the plant's Nitrogen Use Efficiency (NUE). NUE measures how effectively plants convert available nitrogen into agricultural output, such as net yield. Plant NUE can be improved by engineering of genes associated with nitrogen uptake, assimilation, and utilization. One such group of genes encodes Nodule Inception-like Proteins (*NLPs*), which are transcription factors (TFs) crucial for regulating nitrogen responses in plants. *NLPs* have been extensively studied in various plant species, yet their presence and functions in soybean (*Glycine max)* remain unclear. In this study, we identified and characterized *GmNLPs* using *In Silico* approaches. A total of 8 NLP genes (*GmNLP1-GmNLP8*) were identified, exhibiting similarities in physicochemical properties with *Arabidopsis thaliana NLPs* (AtNLPs). Both *AtNLPs* and the selected *GmNLPs* shared conserved domains, including RWP-RK and PB1, confirming their membership in the same gene family. The comparative analysis revealed that *GmNLPs*, on average, possess longer gene sequences (4743 bp), greater protein lengths (905 aa), and higher molecular weight (100 kDa) compared to *AtNLPs*, which exhibit shorter gene sequences (4142 bp), smaller protein lengths (880 aa), and lower molecular weight (97.75 kDa). All *GmNLPs* exhibited pI values below 7, identifying them as acidic proteins, while *NLPs* from both plants displayed negative GRAVY values, indicating their hydrophilic nature. Subcellular localization predictions indicated that all *GmNLP* proteins are nuclear-localized. Gene structure analysis showed variations in exon-intron organization, while motif composition analysis highlighted functionally conserved regions across *GmNLPs*. Phylogenetic analysis demonstrated a close evolutionary relationship between *GmNLPs* and *NLP* homologs from *Populus trichocarpa, Triticum aestivum*, and *Solanum lycopersicum*. Protein-protein interaction analysis revealed that *GmNLPs* strongly interact with nitrogen-responsive

genes, as well as with auxin response factors (ARFs) and GRAS domain proteins, implicating their roles in nitrogen transport and nodulation pathways. Additionally, cis-acting regulatory element prediction indicated that *GmNLP* genes may play critical roles in phytohormone signaling and abiotic stress responses. This study provides a comprehensive understanding of the structural, functional, and evolutionary patterns of *NLPs* in soybean and proposes further *In Vitro* and *In Vivo* investigations to enable genetic engineering of *GmNLPs* for improved NUE in soybean.

Key words: Nitrogen, Nodule Inception (NIN), NIN-Like Protein (NLP), Nitrogen Use Efficiency (NUE), *Glycine max*, Soybean.

Contents

List of Figures

List of Tables

Abbreviations

CLC	Chloride Channel
GOGAT	Glutamine Oxoglutarate Aminotransferase
GRAVY	Grand Average of Hydropathicity
GS	Glutamine Synthetase
GSDS	Gene Structure Display Server
NCBI	National Centre for Biotechnology Information
NLP	Nodule Inception-Like Protein
NRE	Nitrate Responsive Cis-element
NUE	Nitrogen Use Efficiency
NUpE	Nitrogen Uptake Efficiency
NUtE	Nitrogen Utilization Efficiency
NiR	Nitrate Reductase
NiR	Nitrate Reductase
pI	Isoelectric point
PNR	Primary Nitrate Response
UTRs	Un-translated Regions

Chapter 1

Introduction

1.1 Background

Nitrogen is necessary for the formation of amino acids, chlorophyll, hormones and nucleic acids. It has a huge impact on development, growth and production, and so is recognized as among the most important and significant elements limiting plant growth and production. Plants have evolved sophisticated and sensitive regulatory systems to gradually adjust to changes in the amount of nitrogen in the surrounding environment. Nitrogen is often found in soil as nitrogenous compounds such as ammonium and nitrate nitrogen. Nitrate nitrogen, the principal type of nitrogen in soil, has crucial biological and the nutritious role in plant development and growth [1].

Plants are capable of absorbing inorganic nitrogen in two distinct forms: nitrate ($N0_3^-$), and ammonium (NH_4^+). Large quantities of nitrogen-rich fertilizer are applied to crops to help achieve sufficient yields and meet the growing demand for food. Fifty percent of the food consumed by people worldwide is either directly or indirectly influenced by nitrogenous fertilizers. FAO (2021) reports that the global consumption of nitrogen fertilizer has increased to 109 million tons worldwide, and by 2050, it is expected to have increased by 125–236 million tons.

Currently, the increase in agricultural investment is primarily attributable to the usage of nitrogen fertilizer, which has a direct impact on production. Use of nitrogen fertilizers has been increasing since the beginning of the 1960s, but it has steadied slightly during the last decade. Plants may only ingest 30-40% of the nitrogen fertilizer provided, while more than 40% is lost to the environment, ground water, lakes, and rivers due to leakage. Such leakage causes substantial environmental degradation by producing volatile N oxides and releasing nitrate ions from agricultural areas. Excessive use of nitrogen fertilizers pollutes the environment, particularly the aquatic ecosystem, and reduces farmer revenue, whereas underuse of nitrogen is related with reduced agricultural output. As a result, in order to obtain the largest potential and sustainable output, the response to application of nitrogen and its efficacy must be extensively assessed. According to the United Nations Environment Programme, the three most serious global risks are nitrogen pollution, water scarcity, and global warming [2]. As a result, nitrogen management is necessary to enhance crop yield while also reducing environmental concerns associated with nitrogen losses. Improving crop nitrogen use efficiency (NUE) is critical to resolving these issues and promoting sustainable agriculture.

1.2 NUE

Two biological components of NUE are N uptake efficiency (NUpE), represent the rate at which N is absorbed or taken up and N utilization efficiency (NUtE), which is the effectiveness of absorption and remobilization of plant N in order to produce grain. It is defined as the entire quantity of grain that is obtained for each gram of nitrogen that is available to the crop through the soil and fertilizer that is applied to it [3].

It also describes the link between the quantity of nitrogen a crop absorbs and holds until harvest and the amount of nitrogen accessible to the crop. This relationship is most evident when comparing the amount of fertilizer given to soils to the amount of N crops retain. Crop NUE is influenced by the surrounding

environment, plant physiological activity and interaction. Biochemical N transformations in soil are complicated and should be viewed as being in constant flux. NUE is primarily influenced by biochemical transformations fluxes in the soil system. However, physiological nitrogen losses from plants or soil systems also affect NUE. N loss mostly occurs by ammonia (NH_3) gas volatilization, dissolved NO_3^- leaching, overland discharge of all soluble types [4]. Improving NUE of crop is usually acknowledged as an economical, viable, and desirable technique for tackling nitrogen-related agricultural and environmental problems. Improving NUE by just 1% can result in considerable annual savings of approximately 1.1 billion dollars.

Plants have well-coordinated molecular systems for obtaining, consuming, transporting, and utilizing nitrogen, that are controlled by a number of transcription factors and gene families. Plants use nitrate transporters, such as NRT1 and NRT2, to transport inorganic nitrate (NO_3^-) from the soil into cells via channels such as CLC and SLAH. Nitrite reductase (NiR) and nitrate reductases (NIA1, NIA2) convert inorganic nitrate to ammonium (NH_4^+). GOGAT (glutamate synthase) and GS (glutamine synthetase) convert ammonium into amino acids including glutamate and glutamine. In the process of producing plant macromolecules such as nucleic acid, chlorophyll, and vital amino acids, these absorbed amino acids, which function as nitrogen donors, are important. Both ingested nitrate and assimilated amino acids function as signaling molecules, controlling related TFs and cellular activities. These discussions emphasize the importance of N and transcription factors that are sensitive to N, about the overall plant structure, function, and NUE [5].

1.3 NIN (Nodule Inception) Genes

NIN (*NODULE INCEPTION*) genes have been identified to be defective in bacterial recognition, formation of infection threads and beginning of nodulation in the *L. japonicus* as well as the presence of *NIN* genes was later confirmed to be essential for the formation of rhizobia in legume plants [6]. Nuclear-targeted proteins

that attach to DNA via *bZIP* domains are encoded by NIN genes, and the most common characteristic of *NIN* proteins is the *RWP-RK* sequence. *NLP* (NIN-like proteins) are a group of genes that have been found in legumes that have a high degree of similarity to *NIN* [6].

1.4 Nodule Inception-Like Proteins (NLP)

NLP is a family of TFs unique to plants, and its proteins share similarities with NIN in both the N-terminal regions and the *RWP-RK* domain. In NLP, two of the most conserved domains are PB1 and *RWP-RK*. The PB1 domain helps proteins interact with one other, whereas N-terminal regions of *NLPs*, conversely, function as a transcriptional regulatory domain that boosts nitrate signaling. However, the *RWP-RK* domain does not have anything to do with nitrate signaling, rather, it is a domain that binds certain DNA molecules [7]. The nitrogen termini of certain NLP proteins also feature a GAF domain that is extremely conserved, a universal signaling motif, and a unique cyclic GMP receptor that guarantees photosensitizing light reversal and optical signal transmission function properly. This domain may be linked to dimerization or signal transduction [1].

NLPs play a critical role in nitrate signaling. When nitrate becomes available, *NLPs* activate the transcription of genes that help plants take in nitrates, including nitrate transporters (NRTs), NR and NIR, or nitrate and nitrite reductases. This facilitates the absorption of nitrate from the soil by plants, which then convert it into useful compounds like amino acids and ammonium [8].

NLPs also contribute to the coordination of nitrogen signaling with growth and development of plants. They regulate genes participating in nitrogen absorption as well as hormonal signaling pathways including auxin and cytokinin, which influence root architecture, shoot development, and overall biomass accumulation. This shows that *NLPs* serve as integrative centers for several signaling pathways, allowing plants to adapt to fluctuating nitrogen levels and maximize growth accordingly [9].

In *Arabidopsis thaliana*, *AtNLP7* is an extensively studied component that is involved in nitrate signaling. When nitrogen is scarce, *AtNLP7* is found in the cytoplasm; however, when nitrate is present, it moves to the nucleus and adheres to elements in target gene promoters that respond to nitrate, stimulating nitrate absorption pathways. This nitrate-dependent nucleus localization mechanism enables *NLPs* to operate as molecular shifts improving gene expression according to the nitrogen status [10]. *NLPs* are essential for the formation and growth of nodules that fix nitrogen in leguminous plants in conjunction with rhizobia. Furthermore, *NLPs* are important in how plants react to abiotic stressors like cold and drought [11].

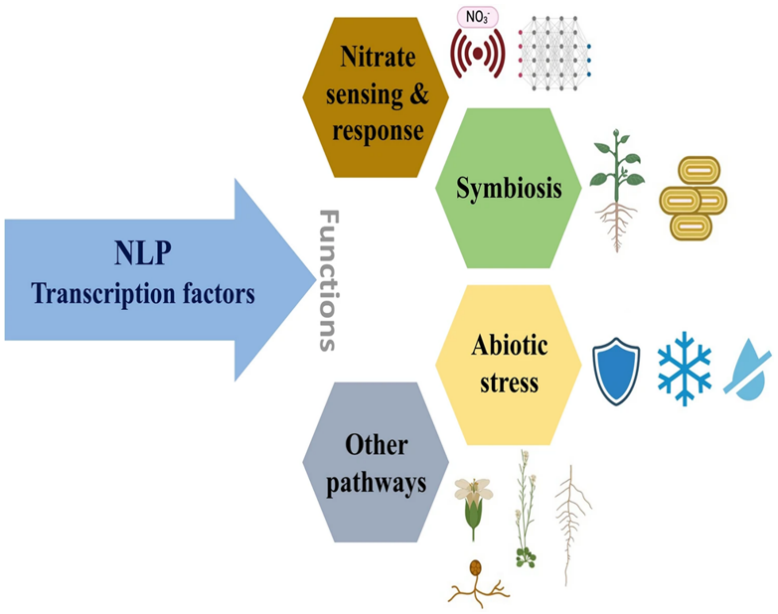

FIGURE 1.1: An outline demonstrating how NLP transcription factors work in a variety of processes related to nitrate signaling, symbiosis, abiotic stress responses, and other related roles [11].

This research focuses on the legume plant usually referred to as soybean. Because of its nutritional content and numerous applications, it has gained several names, including miracle bean, golden bean, queen of pulses, wonder crop, farmer's friend and agriculture's Cinderella [12]. Soybean is an annual dicotyledonous plant whose size can range from small, bushy variety to tall, vine-like cultivars, depending on the intended purpose and environmental conditions.

The plant normally grows to be between 1 and 1.5 meters tall. Soybeans develop pods that slowly mature and fill with protein-rich seeds. Soybean inflorescences can contain two to twenty pods, with a single plant capable of producing up to 400 pods. Soybeans have a certain pattern for root development. Growth of the primary taproot comes first, and then a vast network of subsidiary roots proliferates. Since the development of specialized root nodules is a crucial mechanism for enhancing nitrogen fixation, these roots and the nitrogen-fixing bacteria *Bradyrhizobium japonicum* have a mutually advantageous symbiotic relationship [13].

Glycine max (L.) usually known as soybean is a significant legume crop due to the fact that it is the primary source of plant-based protein that is consumed by both animals and humans around the world and hence the focus of intensive scientific investigation [14]. Soybean is the only legume that contains a substantial amount of omega-3 fatty acids and alpha-linolenic acids.

Furthermore, soybeans contain a higher concentration of biotin, a recognized important dietary element, than fruits, vegetables, and most meat items. Soybeans are used all over the world as a component of a healthy diet due to their high isoflavonoids and folic acid content. Because they are rich in essential amino acids and offer many health benefits, soybeans and their derivatives are considered important sources of plant protein.

The quality lipids in soybeans and polyunsaturated fatty acids are also crucial from a nutraceutical standpoint [15]. Soybean's ability to form symbiotic interactions with nitrogen-fixing bacteria is essential for its growth in nitrogen-deficient soils. This link is supported by numerous gene families, one of them is the Nin-like Protein (NLP) gene family. NLP genes are important for nitrogen absorption and consumption, which are vital for the development and reproduction of plants. Soybean holds both economic importance and scientific relevance, serving as a model organism for studying photoperiodism and symbiotic nodulation with rhizobial bacteria, particularly in relation to nitrogen fixation. Additionally, soybeans help to sustain healthy soil ecosystems, increase soil productivity, and support sustainable farming methods [16]. This research sheds light on the evolution of *NLPs* in

plants and establishes the groundwork for clarifying how *NLPs* control nitrogen responses in *Glycine max*.

1.5 Problem Statement

With reference to pressing issues of climate change, as well as global food security, application of nitrogenous fertilizers is necessary but the chemical fertilizers are potential pollutants. The biotechnology offers a robust system of cleaning the nitrogenous pollutants by improving NUE-responsive genes and TFs. One of such TFs is NLP which has been widely recognized in various plants, yet the *GmNLP* gene family has not been identified through genome-wide analysis.

1.6 Aim

The primary aim of the study is to identify and assess the functional and evolutionary features of *GmNLPs*.

1.7 Objectives

- To genome-wide identify the *NLP* gene family in *Glycine max*.
- To assess the structural attributes of identified *GmNLPs* and compare with *AtNLPs*.
- To assess the functional attributes of identified *GmNLPs* and compare with *AtNLPs*.

1.8 Scope of Study

This study intends to comprehensively characterize the soybean's NLP gene family. Soybeans are a highly significant agricultural legume because of their high

protein content and capacity to fix atmospheric nitrogen through rhizobia symbiosis. Through the use of bioinformatics tools, this work advances our knowledge of NLP genes' suitability for NUE improvement, which is essential for enhancing soybean production in a sustainable manner. Therefore, our discovery may facilitate the both the *In Vitro* and *In Vivo* assessment of *GmNLPs* in subsequent research, which could be advantageous for agricultural output, NUE, and other genetic engineering methods.

1.9 Impact on Society

Understanding the NLP gene family in *Glycine max* could help address both agricultural and environmental challenges. Soybean is a major source of plant-based protein and oil, and improving their NUE can reduce reliance on commercial fertilizers. This not only lowers production costs for farmers but also helps minimize environmental pollution caused by excessive nitrogen runoff. By identifying key *NLP* genes, this study supports the development of soybean varieties that are more effective in nitrogen uptake and fixation. Improved nitrogen utilization contributes to more sustainable farming practices, enhancing food security and conserving resources. In the end, this research might help farmers, consumers, and the environment by making agricultural production more efficient and favorable to the environment. In addition, this research paves the way for creating climate-resilient crop varieties that can perform well even in soils with low nutrient availability especially nitrogen. Such advancements could greatly benefit areas facing soil degradation or limited fertilizer access. These insights may also support future breeding efforts focused on developing environmentally sustainable and cost-effective agricultural solutions globally.

Chapter 2

Literature Review

2.1 General Overview of Soybean

The most significant seed legume in the world, soybeans (*Glycine max* L.) are members of the Fabaceae family. They supply two-thirds of the protein concentrate used in cow feed and around 25% of the edible oil produced worldwide. Soybean meal is a crucial element of poultry and fish feed formulations. The soy protein is known as a whole protein due to its amino acid makeup. Its nutritional importance in diabetes and heart disease is well known. Soybean also contains bioactive chemicals such isoflavones, anthocyanins, and phytosterols, which have antioxidant characteristics and provide numerous health advantages. It is interesting to note that Chinese newborns who consume soy milk instead of cow's milk are almost completely spared from developing rickets [17].

Soybeans are used in traditional Asian cuisines to make miso, tofu, soymilk, soy sauce, and soy paste. In Western societies, soybeans are most frequently processed into soy meal and seed oil. However, during the past ten years, interest in growing this crop has significantly increased in several nations outside of Southeast and East Asia due to its excellent nutritional content and pleasant flavor characteristics. Vegetable soybeans are a legume crop that works well with a range of cropping systems and can increase agricultural sustainability because they can be grown quickly, almost all of their parts are useful, and they can fix atmospheric

nitrogen into the soil when combined with Rhizobia. For instance, vegetable soybeans are commonly grown as a commercial crop in Thailand by farmers after primary crops such as maize and rice. The plant fragments are then harvested and used to build compost and feed cows that produce milk [18].

FIGURE 2.1: Soybean (*Glycine max* (L.)) Merr [19]

2.2 Taxonomy of Soybean

Soybean is a commercially important dicot legume that is a member of Fabaceae family and the genus Glycine Willd. Wild annual and cultivated soybeans are classified as described below:

TABLE 2.1: Classification of Soybean [20]

Order	Fabales
Family	Fabaceae
Subfamily	Papilionoideae
Genus	Glycine Willd
Subgenus	Soja (Moench)
Botanical name	*Glycine max (L.) Merr.*

2.3 Soybean Production and Marketing Worldwide

Vegetable soybean production has increased globally due to increased awareness of the crop's health and nutritional benefits, as well as demand from both domestic and foreign markets. As a result, soybeans are a major crop used around the world. Soybeans are one of the world's ten most widely produced crop species. It is vital for food and nutrition. 353.5 million metric tons of soybeans were produced worldwide in 2020. The expected agriculture area was 127 million hectares. The leading countries that produce soybeans are Bolivia, Russia, Ukraine, India, China, Argentina, Brazil, and the US. In 2020-21, soybeans generated 28.88% plant-derived oil and 70.86% plant-derived protein meals [21].

2.4 Worldwide Export and Import

With a need for more than 135,000 tons of vegetable soybeans in 2020, Japan is the biggest importer of soybeans worldwide. Taiwan, China, Thailand, and Indonesia are the top four exporters of vegetable soybeans to Japan; in 2019 and 2020, they imported around 77,600 and 71,100 tons, respectively, for USD 161.0 and USD 199.2 million. In the form of frozen pods and shelled beans, Thailand exports more than 70% of its vegetable soybeans to Japan. In recent years, however, domestic consumption of vegetable soybeans has increased due to rising consumer awareness of their health benefits. Taiwan and China are the main suppliers of the green vegetable soybeans that make up more than 70% of the US imports. Brazil is now the world's biggest producer of soybean grain. The market for vegetable soybeans is gradually rising. The World Vegetable Center has created awareness on vegetable soybeans in a variety of locations throughout the world. New nations have started commercially producing vegetable soybeans for both home and export markets. Most newly developing countries pursue export potential rather than promoting and utilizing domestically [18].

2.4.1 Production of Soybean in Pakistan

In Pakistan, soybeans were first used as an oilseed in the early 1960s, but planting was not allowed till the 1970s, when production and adaptation trials carried out all throughout the country showed encouraging results. The large areas of Punjab, Sindh, and Khyber Pakhtunkhwa were considered ideal for commercial soybean production based on the results of these studies [22].

Currently, the crop is farmed in a very limited area in the country's north for domestic usage as animal feed and human consumption. Production ranges from 12 to 20 mounds/acre from the irrigated terrain of Punjab to the steep provinces of Khyber Pakhtunkhwa, FATA, and Hazara division. Because soybeans are an unusual crop, the country's overall output cannot match demand.

Despite these barriers, soybean growth has significant potential in areas with favorable agroclimatic conditions. Soybeans are a leguminous crop that not only produces grain but also improves soil fertility and subsequent crop output.

In addition to providing a platform for the research-based promotion of high-yielding, versatile soybean cultivars in the nation, the Ministry of National Food Security and Agricultural Research collaborates with the Pakistan Agricultural Research Council (PARC) to improve the technical proficiency of employees engaged in basic seed manufacturing and new variety development.

As part of this initiative, the company is collaborating with provincial partners to implement a soybean crop commercialization project at the National Agricultural Research Centre (NARC) in Islamabad, Pakistan. This initiative intends to increase agricultural productivity and output by utilizing superior cultivars and modern mechanized production equipment.

The development of new soybean highly productive and adaptable varieties would provide new options for small-scale farmers whose land remains idle following the wheat crop in the Kharif season [23].

2.5 Marketing and Import of Soybean in Pakistan

In 2014–2015, Pakistan imported over one million tons of soymeal with a value of approximately one hundred fifty million United States dollars for usage in poultry and livestock. Demand for soybean seed rose slightly to 1.1 million tons in 2015–16, worth $1.02 billion, to satisfy an increasing need of the solvent and poultry industries. Rather than purchasing soybean meal from Brazil and US, the nation's growing feed industry is instead purchasing whole grain [23].

2.6 Nutrient Content / Nutritional Profile

2.6.1 Protein

By dry weight, soybean seeds typically contain 40–41% protein. Based on their functions, seed proteins are divided into four classes: membrane, storage, structural (including ribosomal), and metabolic enzymes.

Storage proteins make up around 65–80% of the total seed protein. Glycinin and β-conglycinin, the two primary seed storage proteins, are members of the vicilin (7S globulins) and legumin (11S globulins) protein families, respectively [24].

2.6.2 Lipids

Plants use soybean seed lipids for energy, membranes, signaling, pathogen defense, and other purposes. Another kind of bioactive lipid component present in soybean seeds that has been the subject of much investigation because of its ability to decrease blood LDL cholesterol is phytosterols. When soybean oil is purified, phytosterols are recovered as byproducts and used to make commercial goods for the food and health sectors.

2.6.3 Carbohydrates

Carbohydrates are the third most prevalent component in soybean seeds, accounting for approximately 35% of their dry weight. They are abundant in the soybean seed coat (86% of its dry weight), but they may also be identified in embryonic parenchyma cells.

Although the hulls remove some of the seed's carbohydrates, soybean meal may still contain up to 40% of total carbs [24].

2.6.4 Fat

The lipid content, or fat, of soybeans is composed of 15-20% saturated fatty acids, 20-41% monounsaturated fatty acids, and 46-63% polyunsaturated fatty acids. The proportions of omega-3 -linolenic acid and omega-6 linoleic acid in PUFA are 88.2% and11.8% respectively.

More than 40% of the omega-3 fatty acids consumed in the US come from soybean oil, the most commonly used vegetable oil [25].

2.6.5 Soy Oil

Soybeans contain approximately 20% oil, making them a valuable source of plant-based fats. Soybean oil is taken out, processed, and combined for a number of reasons, most of which have to do with food. There are no trans fats in soy oil, and it has a minimal amount of saturated fat [26].

2.6.6 Vitamins

Vitamins A, B6, B12, C, and K are found in soybeans. These chemicals must be obtained through diet since an organism is unable to produce them in sufficient amounts [26].

2.6.7 Isoflavone

Soybeans contain significant levels of isoflavones (1-5 μg/g dried soybean). Soy isoflavones are tiny, bioactive compounds that are both nonsteroidal and phenolic. Phytoestrogens are named after their molecular similarities to 17-β-estradiol, which binds to the estrogen receptor. Genistein, daidzein, and glycitein are the three main isoflavones produced by soybeans [27]. They have the potential to improve health by improving postmenopausal symptoms, reducing the risk of osteoporosis, avoiding heart diseases, and having antimutagenic qualities.

2.7 Health Implications

Consuming soybean or soy products has numerous health benefits, including preventing breast cancer, prostate cancer, symptoms associated with menopause, cardiovascular disease, and osteoarthritis [28].

Soyfoods and soybean products are high in isoflavones, which are categorized as a major group of phytoestrogens. Epidemiological studies show that isoflavones present in soybeans and soybean derivatives, like daidzein and genistein, protect against breast and prostate cancer [28].

2.7.1 Anti-Diabetic Effects

Soybeans and soy derivatives have been shown to have significant diabetes-prevention advantages. In type 1 diabetes, utilizing soy proteins rather than animal proteins reduces GFR and proteinuria. Soy products containing significant isoflavonoids have been shown to have anti-diabetic properties.

Soybean extract has also been shown to effectively prevent the incorporation of glucose into brush border membrane vesicles. Stigmasterol, a phytosterol obtained from edible soybean oil, has demonstrated possible advantages in treating type 2 *diabetes mellitus* [29].

2.7.2 Anti-Oxidant Effects

Soybeans and their derivatives are particularly effective at lowering oxidative stress and eliminating free radicals. Dou-chi, an old Chinese soybean food fermented with *Aspergillus sp.*, has been proven to have antioxidant potential. A crucial soy product, soy milk, has been shown in human studies to reduce oxidative stress in Type 2 *Diabetes Mellitus* patients. Through the modulation of antioxidant levels, fermented soy milk helps individuals with type 2 diabetes minimize oxidative stress [30].

2.7.3 Reduce Risk of Cardiovascular Disease

According to certain research, long-term soy protein consumption can help decrease blood cholesterol and avoid cardiovascular disease. Furthermore, scientific study has revealed that soy protein helps lower blood total cholesterol and LDL levels in adults.

2.7.4 Cancer Prevention

Consuming soybeans is one of the main reasons why hormone-related malignancies like endometrial, breast, and prostate cancer are less common. Soyfoods and soybean derivatives contain significant quantities of isoflavones, a type of phytoestrogen. Epidemiological studies have shown that isoflavones present in soybean and soy products, which include daidzein and genistein, protect against breast and prostate cancers [26].

2.7.5 Protection Against Osteoporosis

Several studies have demonstrated the positive effects of soybean isoflavones on bone health, enhancing bone strength and preventing estrogen-induced bone loss [26].

2.8 Role of Nitrogen in Plants

Nitrogen is an important component of proteins, chlorophyll, vitamins, hormones, and nucleic acids, and it is required for every step of plant growth and development. Nitrogen is also required for cell division. Nitrogen is a critical component for plant development and growth, in addition to how plants react to various abiotic stressors. Although nitrogen is thought to be the most significant nutrient for plant development in terms of quantity, plants have evolved effective strategies to regulate nitrogen levels in response to a variety of complex stressors. As a result, knowing the interplay between N and abiotic stress in plants is critical for optimizing N fertilizer use, while maintaining a balance between application and the negative impacts of abiotic stresses. This understanding is critical for enhancing advanced agricultural systems and promoting sustainable farming practices [31].

2.8.1 Plant Growth

N is a nutrient that crops need to grow and develop, and if insufficient amounts are absorbed, plant growth is limited. However, it has been shown that too much nitrogen in the soil negatively affects crop growth and output. Growth is a complicated process regulated by nutrient uptake and moisture availability. Because nitrogen is one of the most consumed elements, it is critical for assimilation and transport to developing organs. The presence of this element constantly promotes development and performance by giving the necessary water. Furthermore, excess nitrogen in the soil encourages the lateral buds of older leaves and the terminal meristem of the stem to produce new leaves. It eventually boosts the production of aerial components. N shortage causes phenological development to be delayed in both, the stages of vegetation and reproduction. Utilizing the appropriate quantity of nitrogen fertilizer may considerably enhance biomass, and high biomass is only feasible under N fertilization circumstances. N seems to sustain the survival of the leaf surface, as the durability of the leaf surface rises, so does the length and rate of photosynthesis, enabling the plant to create more dry matter [32].

2.8.2 Yield

Since nitrogen is essential to plants, it is critical to properly manage N consumption in order to boost plant productivity. Reduced leaf size due to a lack of nitrogen results in less light absorption and less efficient utilization of light for plant photosynthesis, which lowers biological production and vice versa. The amount of N consumed should match the demands of the plant. N loss in the soil is increased by excessive N usage via leaching, inadequate efficiency of N and insufficient use of excess N by plants.

Consequently, adequate management entails giving the plant the ideal amount of N so that it can utilize it. By influencing plant morphological characteristics, N increases grain yield; the optimal plant yield is obtained when N is applied at various plant phenological phases [32].

Because farmers think that adding N fertilizer will always boost crop yields, they usually apply it more frequently than is advised, which has a negative impact on the production system's sustainability and raises costs. From the perspective of biomass production and yield, the quantity of N that produces the best growth rate and yield is the ideal level. However, often, the appropriate N level is determined by the product's quality and health characteristics, such as the quantity of nitrate in the product. Root development and bulking, soil moisture absorption, and dry matter and grain yield all increase with increased N intake. Additionally, consuming more nitrogen speeds up green development, increases aerial part volume, and boosts plant evapotranspiration.

According to Ying et al. (1998), a plant should be able to take in more nitrogen as its biomass and yield increase. For example, a plant that produces more than 13 tons of biomass per hectare needs to take in more than 250 kg of nitrogen per hectare. Research indicates that the more N fertilizer is used, the more of this element builds up in the grain and shoots of wheat varieties. As a result, the protein content in the grains continues to increase gradually.

2.8.3 Chlorophyll

N is the primary component of plant cells' proteins and chlorophyll. Dry matter generation and photosynthetic rate are directly correlated with chlorophyll concentration. Since 70% of the nitrogen in leaves is stored in chloroplasts, which create the pigments that make up chlorophyll, the amount of chlorophyll and nitrogen in plants are tightly connected. Amino acids, proteins, nucleic acids, and chlorophyll structure all mostly include nitrogen. Because foliar chlorophyll content and leaf N concentrations are strongly correlated, it would be beneficial to track plant chlorophyll and N concentrations during production to improve growth, yield, and marketability [32].

2.8.4 Photosynthesis

The primary component of plants' photosynthetic mechanism is nitrogen, a common fertilizer. Since photosynthesis is the most significant activity for plant development and biomass production, it is the primary driver of yield formation. To put it another way, the efficiency of photosynthesis, absorption, and distribution are the factors that influence the level of agricultural productivity. N ions are therefore essential for these reactions. N fertilization affects growth patterns and leaf longevity, which in turn affect photosynthetic efficiency, in addition to increasing leaf area. Enough nitrogen added to plants throughout their late growth stages delays the breakdown of protein and chlorophyll solutions, prolongs photosynthesis, fortifies the defenses of the leaves, and prevents leaf senescence [32].

2.8.5 Nitrogen Acquisition in Leguminous Plants

Legumes get their nitrogen in two ways, unlike most other plants. Inorganic nitrogen is first taken up by them from the soil. Second, legumes coexist symbiotically with rhizobia, which are microorganisms that fix nitrogen. Rhizobia thrive in the nodules produced by the roots of leguminous plants as a result of this symbiotic relationship. Rhizobia in these nodules carry out the essential biological function

of nitrogen fixation, which transforms atmospheric nitrogen (N_2) into a usable form for plants. This nitrogen-fixing ability is very advantageous to leguminous plants, ultimately improving their growth and nitrogen nutrition in general [33].

2.9 Role of Transporters in Plants / Mechanism of Nitrogen Uptake in Plants

Plants have several advanced inorganic N absorption systems, such as nitrate and ammonium transporters, as well as extremely complicated regulatory mechanisms, to adjust to the diverse types and amounts of nitrogen in the environment.

2.9.1 NO_3^- Uptake in Plants / Nitrate Transporters

Plant nitrate transporters are divided into nitrate transporter 1 (NRT1/PTR), nitrate transporter 2 family (NRT2), slow anion channel homologues (SLAC1/SLAH) and chloride channel (CLC), based on their structure, location, and function [34]. Plant nitrate transporters (NRTs) are critical for NO_3^-_N absorption by roots, along with transport over long distances and redistribution in stems and roots. The NO_3^-_N transport systems found in plants are classified as either low-affinity (LATS) or high-affinity (HATS). The three families of NRT genes found in plants are NRT1/PTR, NRT2, and NRT3.

Additionally, the nitrate transporter 1/peptide transporter family, also known as the NRT1/PTR Family, is generally referred to as the NPF family, since NRT1 and PTR have similar gene structures and functions. In *A. thaliana*, a total of 53 NRT1 members and seven NRT2 members have been discovered. NAR2 is NRT2's companion protein, which creates an association with the protein and ensures its stability [34]. In addition to transporting nitrate nitrogen, NRT1 is responsible for transporting a wide range of substrate molecules, which demonstrates the protein's participation in a number of activities related to the progress and development of plants.

All of the NRT2 family transporters and a few of NRT1 family transporters are capable of transporting NO_3^-_N. The first NRT component identified in *A. thaliana* was *AtNRT1.1*, a double-affinity transporter that is capable of carrying both high and low NO_3^-_N concentrations [35].

Chloride along with other anion channel proteins, which are frequently referred as CLC transporters, are present in organelles and cytoplasmic membranes. They typically have ten to twelve transmembrane domains having conserved amino acid sequences mediating anion binding. In Arabidopsis, the CLCa-g family consists of seven members. CLCa stores nitrate on the vacuolar membrane while exchanging chloride in opposite direction. Phospholipids and nucleotides regulate its action.

In *Xenopus oocytes*, CLCb, a vacuolar protein, promotes reverse cotransport of Cl^- and NO_3. Downregulating HHO2 leads to higher expression of CLCb in low-N conditions, improving nitrogen use efficiency (NUE) and promoting vacuolar NO_3^- efflux. Domain of 10 helices and the guard cell plasma membranes include SLAC1/SLAH transporters, that are implicated in ABA-controlled closure of the stomata. One out of five Arabidopsis SLAC1/SLAH proteins (*SLAC1*, *SLAH1-4)*, *SLAH2*, is detected in root cells and may aid in "transporting nitrate" to shoots. *SLAH3*, which controls Cl and NO - efflux during closure of stomatal pores, has a higher capacity for transport than SLAC1. The nitrate-dependent decrease of ammonium toxicity is regulated by a functional component composed of *NRT1.1* and *SLAH3* [36].

2.9.2 NH_4^+ Uptake in Plants / Ammonium Transporters

Ammonium transporters (*AMTs*) take up NH_4^+_N from the soil. Ninnemann found the first ammonium transporter in 1994, which he named *AtAMT1.1*. The plant AMT gene family is divided into two subdivisions: *AMT1* and *AMT2*. Each subfamily has a different total number of members depending on the type of plant. There are five members in the *AMT1* family and two in the *AMT2* family in *A. thaliana*.

The *AMT1* family has three members, while the *AMT2* family has seven. Despite having different protein sequences, *AMT1* and *AMT2* members have 9–11 transmembrane structural regions in their high-level structures that are similar and their tertiary structures are quite similar. Different *AMT* family members are expressed in different plant organs and tissues. The root epidermis and cortical cell membranes include *AMT1.1* and *AMT1.3*, which make up around 30% of the high-affinity absorption system in *Arabidopsis*. Crop yield and NH_4^+_N absorption can be significantly increased by overexpressing *OsAMT1.1*. The primary functions of *AMT1.2*, which is located in the root cortex, are either absorbing ammonium ions from the roots' non-protoplasts or transferring them to the vascular system [37].

2.9.3 Assimilation of Nitrogen in Plants

The following phases are usually part of the assimilation process for NO_3^- (i) taking in NO_3^- and turning it into nitrite; and (ii) then turning it into NH_4^+. (iii) Glutamine synthase (GS) makes glutamine (Gln) from NH_4^+. (IV) Glutamate synthetase (GOGAT) makes glutamate (Glu) and changes it into other amino acids that plants need.

Ammonium is an important component in the assimilation of nitrogen into organic molecules, whether the nitrogen comes from soil or nitrogen fixation. Nitrate taken by roots is either transformed in the roots to nitrite before being converted to ammonium, or it is transferred to shoots and decreased in the leaves to ammonium. The main mechanism by which plants absorb ammonium is probably the GS/GOGAT cycle. GS has a high affinity towards ammonium and can act in low ammonium concentrations. This is significant since excessive concentrations of ammonium are hazardous. The enzyme GOGAT has been identified in the bacteria *Aerobacter aerogenes* and has also been found in nitrogen-fixers.

Nitrate reductase (NR), the rate-limiting enzyme, is activated by NO_3^-_N concentration and triggers the reduction of NO_3^-_N to NO_2^-_N in plants [38]. Abiotic stress has a sensitive impact on NR. Nitrite reductase (NiR) is a catalytic enzyme that converts NO_2^- to NH_4^+ and is mostly present in plant chloroplasts.

The NO_2 supply influences NiR activity. GS stimulates the reduction of NH_4^+ with ATP, which then combines with Glu to form Gln. The two basic sources of NH_4^+ employed in the preceding processes are NO_3^- reductions and immediate root absorption. In plant cells, there are two different kinds of GS: GS1 is located in the cytoplasm, and GS2 is found in the plastid. Glutamate dehydrogenase (GDH) catalyzes reversible amination/deamination reactions in plants, controlling Glu production and breakdown. Masclaux-Daubresse et al. report that the GS/GOGAT pathway is the primary Glu production pathway, while GDH is predominantly responsible for Glu deamination. Several genes, including NR, NiR, and GS, regulate N metabolism-related enzyme activity [39].

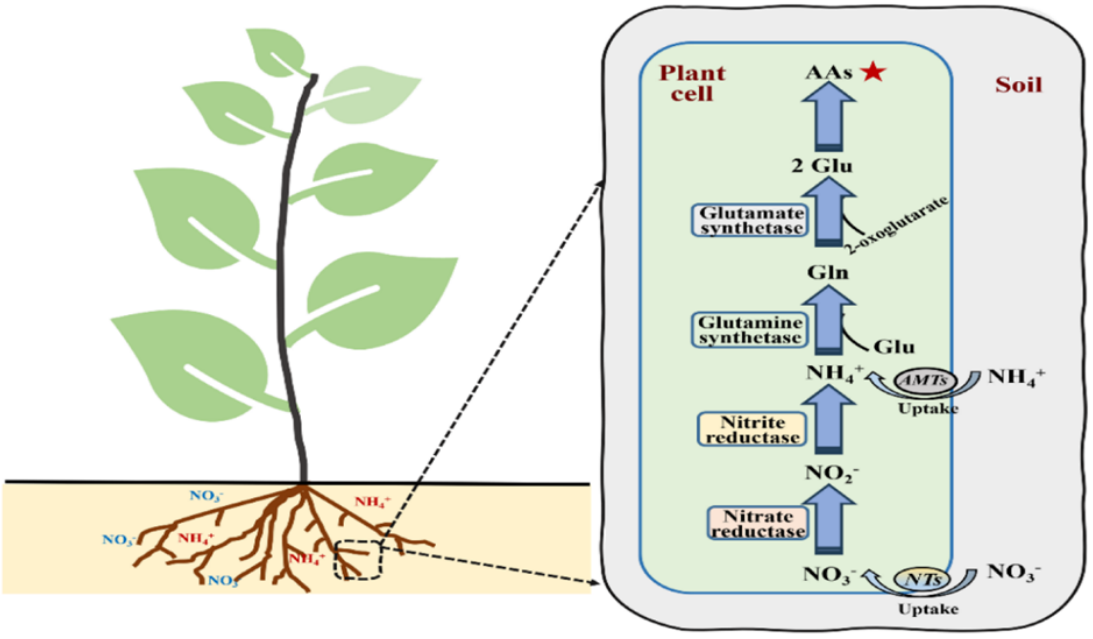

FIGURE 2.2: Mechanism of Nitrogen Assimilation in Plants [37]

2.10 Importance of Nitrogen and NUE

A variety of nitrogen sources, such as nitrate, ammonium, nitrogen metabolites, and amino acids, are available to plants, but only inorganic nitrate can efficiently stimulate plant biomass development and root remodeling. Since nitrate is the autotrophs' main source of nitrogen, it is crucial to the structure and development of roots. Nitrate treatment has the potential to stimulate the development of lateral roots and encourage their proliferation. The absorption of plant N fertilizer

is less than 50%, depending on crop species and soil conditions. Excess nitrogen contributes to climate change by polluting the environment and releasing greenhouse gases like N_2O. In order to achieve a balance between high crop production and a reduced nitrogen application rate, it is necessary to improve the efficiency with which plants use nitrogen.

2.10.1 Nitrogen Use Efficiency

The examination of NUE provides information about how plants react to various nitrogen availability situations. Since nitrogen use efficiency may be divided into two parts, namely nitrogen uptake efficiency (NUpE) and nitrogen utilization efficiency (NUtE), it is necessary to compute both parts in order to evaluate NUE. NUpE is calculated by dividing the total amount of nitrogen in above-ground plant biomass at harvest by the amount of nitrogen available in the soil. NUtE is calculated by dividing the amount of nitrogen present in grain tissues by the amount of nitrogen in the above-ground plant biomass. So NUE is determined at the time of harvest, i.e., end of the crop cycle. The efficiency of nitrogen usage is linked to nitrate acquisition, which may be further improved by modifying the proteins and enzymes involved in nitrate absorption by various biotechnological techniques. Instead of focusing on single-point rate-limiting control, it is imperative to target numerous processes, enzymes, and variables in order to improve NUE [40].

2.10.2 Genes Related to NUE

It is known that several gene sets in agricultural plants control the processes linked to NUE, including nitrogen uptake, accumulation, and remobilization. The genes that control NUE in many cereal crops, including rice and wheat, may be roughly categorized into six groups: transporters, signal molecules, nitrate absorption, amino acid biosynthesis, transcription factors, and other genes. Transporters and nitrate absorption genes play important roles in nitrogen uptake, whereas amino acid biosynthesis genes play a role in nitrogen metabolism. Signaling molecules, transcription factors, and other genes all play a passive role in nitrogen absorption and utilization [40].

2.11 Consequences of Poor NUE

Excessive nitrogen fertilizer application is the sole cause of modern agriculture, which damages ecosystems and pollutes the environment. It is believed that seventy percent of the nitrogen fertilizer that is sprayed is lost in the environment, influencing atmospheric chemistry locally and globally. Groundwater nitrate pollution has caused a slew of environmental and socioeconomic issues. Drinking nitrate-contaminated water is a severe problem, particularly among young people [41]. Adults who continue to ingest NO_3^- contaminated water get stomach cancer, whereas children acquire methemoglobinemia. Furthermore, by lowering the oxygen concentration of the water, NO_3^- or NH_4^+ poisoning of water bodies encourages a proliferation of algae and other types of aquatic vegetation. Nitrogen's oxide forms are extremely reactive and harm the environment in many ways. Nitric acid, a major component of acid rain, is produced in part by excessive nitric oxide and nitrous oxide emissions. It damages infrastructure and has a significant effect on soil microbial communities. Additionally, the interaction between volatile organic pollutants and nitrous oxide produces the atmospheric pollutant ozone. Consequently, nitrogen depletion leads to serious environmental and health problems. Crop NUE must be addressed globally in order to prevent these effects [4].

2.12 Strategies to Improve NUE

2.12.1 Conservation of the Land Through Tilling

The degree of soil disturbance that is brought about by the many different tillage techniques influences plant nitrogen availability and soil nitrogen dynamics. According to Francis (1993), conservation tillage practices reduce nitrogen availability compared to conventional tillage systems. By limiting soil disturbance, conservation tillage can reduce the rate of N mineralization, lowering crop N availability and loss. However, under traditional tillage methods, soil organic matter loss is

hastened by increased soil erosion and oxidation caused by exposure and disturbance. Conventional tillage methods remove soil organic matter, reducing soil quality and nitrogen availability. Thus, the tillage system's operation will be critical to increasing NUE [4].

2.12.2 Managing N Inputs for NUE

Changes in fertilizer application timing, rate, source, and location have been shown to improve NUE. Although current and future technology may alter or enhance these methods, they are regarded as essential components of nitrogen control and cannot be replaced by other components.

2.12.3 Crop Production System

Enhancement of nitrogen usage efficiency is also influenced by cropping system capabilities. Variety in crops can improve soil health, structure, mycorrhizal fungal interactions, upward nutrient stratification, and crop residue diversity. The proposed farming strategy has the potential to increase plant uptake and nitrogen delivery. The best strategy to store leftover nitrogen is to grow cereals and legumes. The interaction of legumes and cereals, as well as shallow and deep-rooted crops, makes crop rotation and intercropping advantageous. Therefore, rotating crops with varying root depths can increase the efficiency of nitrogen use while simultaneously improving soil stability and structure. In general, tap-rooted crops absorb nitrogen more efficiently than crops with shallow roots because they may more readily pierce layers of compacted soil [4].

2.12.4 Cropping System using Cover Crops

The rate at which roots take up nitrogen from the soil and the quantity of nitrogen that accumulates in various areas of the plant such the stem, leaf, and cultivated sections, affect how well plants use nitrogen. Because of this, cover crops have an

effect on NUE in a cropping system. Adding cover crops and multipurpose fodder crops, which produce a lot of biomasses, to any system can raise its total NUE. Cover crops may help keep roots alive in the soil for as long as possible, stop soil erosion, and recycle nutrients. Legume, a cover crop with low residues of the C:N ratio, may accelerate the breakdown of organic nitrogen, which might account for the high NUE of the main crops [42].

2.12.5 Optimizing Fertilizer Management

Slow-release nitrogen fertilizers tend to regulate N release and uptake based on crop needs, while reducing N runoff and denitrification losses. Coated N sources, such as sulfur-coated urea and synthetic urea-based fertilizers with progressive release, have also improved NUE. Furthermore, polymer-coated urea has been shown to enhance NUE (3%-34%) while decreasing N volatilization loss (23%-62%) and ammonia emissions (51.3%-91.3%).

2.12.6 Genetic Engineering

It is significantly easier to functionally validate NUE now that we have high-throughput genomics technology and rapid transformation methods for model crops. MiRNAs have been shown to play a key function in controlling N-responsive gene expression in N-limited environments. As a result, identifying gene regulatory networks, like short RNAs that participate in stress response regulation, would aid our understanding of how to create stress-responsive crops with high NUE [43].

2.13 Transcription Factors

The role of RWP-RK proteins in nitrate sensing and signaling has recently come to light. Many plant species have been shown to contain RWP-RK proteins, often referred to as NIN-like proteins, according to extensive genome-wide investigations.

Green algae, slime molds, and all vascular plant species include the RWP-RK protein family of transcription factors, which are identified by the existence of RK (Arg-Lys) in addition to RWP (Arg-Trp-Pro) motifs. These are classified as a plant-specific TF family. The DNA binding-related RWP-RK domain is categorized as a novel transcription factor. The two categories of RWP-RK proteins are RKDs and *NLPs*. Since *NLP* proteins attach to promoters and alter their expression, they are known to regulate genes linked to nitrogen use efficiency (NUE). RKD genes, on the other hand, play a crucial role in gametogenesis and embryogenesis [11].

Many studies have been conducted on the functional control of genes linked to nitrogen transport and absorption. Within the RWP-PK family, for example, the *NLP* transcription factor subfamily is responsible for responding to nitrogen deprivation and plays an essential role in the regulation of plant nitrogen signaling. Nodule Inception (*NIN*) was the focus of the initial research on Nin-like Protein (*NLP*), which was conducted in a legume model plant. The gene known as NIN was initially identified as a key regulator gene that controlled the growth of root nodules in plants. After further investigation, it has been shown that other non-leguminous plants, including *A. thaliana*, *Oryza sativa, Triticum aestivum, Zea mays,* and *Brassica napus*, possess homologs that are similar to it [44].

Nitrate signaling depends critically on members of the *NLP* family. Furthermore, a lot of research has been undertaken to study the roles of nodule inception-like proteins (*NLP*). The nucleus contains both *NLP7* and its homolog, *NLP6*, which have the ability to trigger the nitrate response cis-acting element. Marchive et al. (2013) investigated the crucial role that *AtNLP7* plays in *A. thaliana*'s early nitrate response. *AtNLP7* also plays a role in nitrogen deficiency. The TCP20-NLP6/7 complex controls nitrate absorption and signaling, as well as the synthesis of NRT1.1, NIA1, and NIA2, when there is a nitrogen deficit. Additionally, *NLP7* stimulates nitrate-dependent lateral root development and the expression of BT1 and BT2. *NLP7*-overexpressing transgenic plants in *A. thaliana* had a considerably higher density of lateral roots than the wild-type control, which interfered with nitrate-stimulated stem growth and root remodeling by altering

Ca2+. Through direct activation of abscisic acid lyase production, signaling pathways demonstrated that NLP8 is essential for nitrate-mediated seed germination. Liu and colleagues discovered that the NO_3^-CPK-NLP signaling pathway plays a role in plant vegetative growth. The aforementioned investigations have confirmed that *NLP* transcription factors play critical roles in nitrate response, plant growth, and development.

An investigation of *NLP* gene families from different species found that *NLP* transcription factors share two distinctive domains: the highly conserved RWP-RK domain and the carboxyl-terminal PB1 domain. The PB1 domain is believed to be involved in protein-protein interactions, whereas the RWP-RK domain is known to bind to DNA. The protein connections between *MtNLPs* and *MtNIN* in alfalfa and between *AtNLP6/7* and *TCP20* in *Arabidopsis* are also controlled by the PB1 domain. *NLP* proteins are present in nearly all examined tissues, including rice leaves, nodes, and inflorescences, as well as *Arabidopsis thaliana* seedlings, roots, stems, and flowers [7]. In *Oryza sativa*, *OsNLP1* and *OsNLP3* are primarily expressed in source tissues, while in *A. thaliana*, *AtNLP8* and *AtNLP9* are predominantly expressed in senescent leaves and seeds.

2.14 Role of *NLP* in Nitrate Response

The primary nitrate response is mediated by a number of genes, including nitrate sensors like NRT1.1, nitrate reductase (NR), nitrite reductase (NiR), glutamine synthase (GS), and glutamate synthase (GOGAT). PNR has been identified to be dependent on NIN-like protein transcription factors (TFs), which are among these factors. *NLPs*' role as nitrate-responsive transcription factors is strongly suggested by molecular evidence from *Arabidopsis* transgenic lines expressing the *NLP6* promoter, which shows selective β-Glucuronidase (GUS) activation in response to nitrate.

Additionally, substantial study has been conducted on the nitrate responsiveness of *NLPs* in diverse plants. The nine *NLPs* identified in *A. thaliana* (*AtNLP1* to

AtNLP9) are believed to bind to nitrate-responsive cis-elements (NREs), thereby initiating NRE-dependent gene expression in response to nitrate.

In a similar manner, it has been revealed that five *NLPs*, one of which is NIN, bind directly to the NRE motif in *Lotus japonicus*, therefore increasing the transcription of the target gene.

While the N-terminal region of NIN appears unresponsive to nitrate signaling, the corresponding regions in *LjNLP1* and *AtNLP* do show responsiveness. Interestingly, NIN has also been found to repress the expression of nitrate-inducible genes in *Lotus japonicus*.

NLPs are regarded as principal regulators of the nitrate response. Their interactions with nitrate-responsive cis-elements (NREs), which are found either adjacent to or within nitrate-inducible genes, result in the activation of nitrate transporters and enzymes that are responsible for nitrate and ammonium assimilation.

Generally speaking, *NLP* mutations decrease nitrate-induced gene expression, which in turn slows down the growth of the vegetative stage. Recent research has shown that in addition to its role in transcriptional control, *NLP7* may also directly bind to nitrate and act as a nitrate sensor inside cells [45].

Other Arabidopsis *NLPs* contain nitrate binding residues, which can also be found on proteins from other plants and microorganisms. However, functional differentiation among *NLP* family members has received little attention. Researchers found that *NLP2* and *NLP7* are very important for vegetative growth [46].

New research conducted by Mickael Durand and colleagues [47] analyzed the regulatory network of genes associated to N, to better understand *NLP2*-dependent regulation. The researchers observed that *NLP2* and *NLP7* perform comparable roles in regulating the major nitrate response (PNR), but *NLP2* performs a distinct role in linking the PNR with both carbon and energy metabolism in the presence of nonlimiting nitrogen.

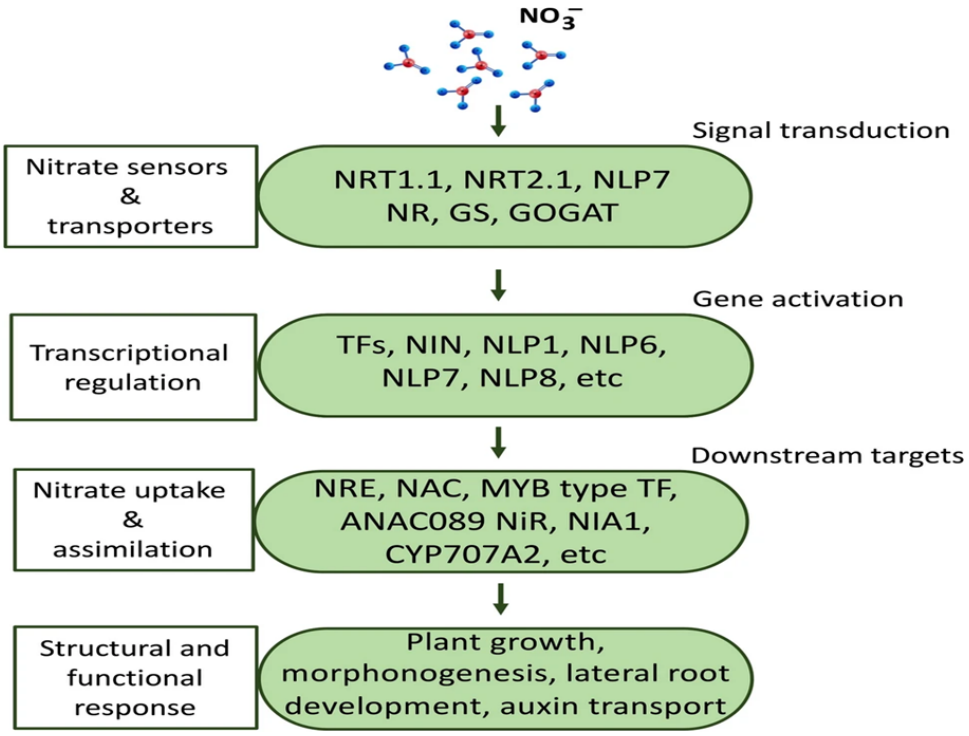

FIGURE 2.3: A summary of how plants react to nitrate, including how nitrate sensors are activated, how they are then transported to the cytoplasm, and how genes are regulated transcriptionally [11]

2.15 Mechanism of Post-Translational Induction of NLP Activity

Post-translational activation of *NLP* activity is necessary for nitrate signaling and gene expression. This is due to the fact that *NLPs* are the key transcription factors that govern the nitrate response prior to nitrate signaling. Recent research is shedding light on how *NLP* transcription factor activity is regulated post-translationally. The initial investigation found that *A. thaliana NLP7* accumulates in the nucleus in a nitrate-dependent manner. Following nitrate treatment, *NLP7* linked to green fluorescent protein accumulates rapidly in the nucleus, which is reversed when nitrate is withdrawn. The addition of a nuclear export inhibitor enhanced *NLP7* nuclear accumulation, indicating that *NLP* retention in the presence of nitrate is a critical component of the nitrate signaling cascade.

The second investigation found that *NLP* phosphorylation was Ca^{2+}-dependent, as was the induction of *NLP* activity [48].

A few Arabidopsis calcium-dependent protein kinases, CPK10, 30, and 32, rapidly phosphorylated Arabidopsis *NLP6* and *NLP7* in response to nitrate treatment. Ca^{2+} sensor inhibitors and Ca^{2+} channel blockers stopped this phosphorylation. All *NLPs*, but not NIN, share the nitrate-responsive domain, which contained the phosphorylated amino acid residue, a serine. *NLP7*'s nitrate-inducible nuclear accumulation was eliminated and *NLP7*-dependent transactivation was greatly decreased when this serine residue was mutated to alanine. Mutant *NLP7* expression has ceased to complement the *NLP7* mutant phenotype, which was consistent with the loss of activity. The *cpk10*, *cpk30*, *cpk32* triple mutant displayed a nitrate-insensitive phenotype, indicating that *NLP* phosphorylation is essential for nitrate response. Despite the likelihood that CPK10, CPK30, and CPK32 are N-myristoylated and associate at the plasma membrane in the absence of nitrate signaling, as are other CPKs [48], some CPKs were translocated into the nucleus and interacted with NLP7 in response to nitrate application. These findings support the existing paradigm for nitrate signaling, which includes Ca^{2+} signaling, *NLP* nuclear accumulation, *NLP* phosphorylation, and the consequent NRE-dependent activation of many nitrate-responsive genes [49].

2.16 Characterization of *NLP* Gene Family in *Arabidopsis thaliana*

The *Arabidopsis* genome has nine *NLP* genes, though only three have been functionally characterized. Due to its higher susceptibility to nitrogen shortage and alterations in nitrate-induced gene expression, *NLP7* has attracted the most attention. Additionally, *NLP7* also helps to maintain root cellulose and pectin levels.

Research indicates that *NLP* transcription factors have a role in the expression of genes in *A. thaliana* that are activated by nitrate. A rapid increase in the concentration of calcium ions within the cell is brought about by the administration

of nitrate, which in turn activates calcium-dependent protein kinases (CDPKs) of group III. Phosphorylation of *Arabidopsis NLP7* occurs at amino acid position 205 (Ser205) when a number of CDPKs, including CPK10, CPK30, and CPK32, enter the nucleus. In order to activate gene expression, *NLP7* is phosphorylated, which means it is active, and it accumulates in the nucleus. On the other hand, *NLP8*, which is only produced in imbibed seeds, has a unique role in *Arabidopsis*. Its purpose is to promote the germination of dormant seeds that have a particular function on the plant. The *NLP8* protein is responsible for inducing the expression of a variety of nitrate-inducible genes in seeds that have been consumed. One of these genes is the CYP707A2 gene, which is responsible for encoding an abscisic acid (ABA) breakdown enzyme [46].

2.17 *Brassica napus NLP* Gene Family Characterization

The author found 31 *BnaNLP* family genes throughout their analysis. When compared to the NLP gene family found in other plant species, the *BnaNLP* gene family is significantly larger. This is due to the fact that *B. napus* has an additional WGT and merger event. During the course of this research, a number of different gene structures were found in the members of the *BnaNLP* family. The fact that members of group III included five to six exons, but members of group I and II contained four to five exons, is indicative of the structural variety that exists within the *BnaNLP* subfamily. Both the *BnaNLP8-6* and the *BnaNLP9* genes include seven exons. Both PB1 and RWP-RK are examples of domains that are conserved in NLP proteins.

The *BnaNLP* proteins that were discovered in this investigation all have two domains, with the exception of *BnaNLP4-4*, which only possesses the RWP-RK domain. According to sequencing analysis, *BnaNLP4-4*'s C-terminus is missing around 50 amino acids, which also includes the PB1 domain. This might be due to sequencing mistake in the genome of *B. napus* or sequence fragmentation during

evolution. The amphipathic leucine zipper and helix-turn-helix motif seen in the RWP-RK domain may play a role in DNA binding. All of the *BnaNLP* proteins in this investigation had the three domains (with the exception of *BnaNLP4–4*), indicating their critical roles in *B. napus* quick reaction and adaptation to nitrogen deficiency [1] .

2.18 Evaluation of *Oryza sativa* (Rice) NLP Gene Family

NLP protein *OsNLP1* is required for nitrogen metabolism. *OsNLP1* is located in the nucleus, and nitrogen shortage rapidly raises its transcript level. Under varied N conditions, *OsNLP1* overexpression promotes plant growth and yield, but *OsNLP1* deletion affects grain yield and NUE under nitrogen-limited environments.

Through the coordination of several genes that are involved in the process of nitrogen absorption and assimilation, *OsNLP1* determines how nitrate and ammonium are used. *OsNLP* may directly bind to these genes promoters to increase their expression, according to chromatin immunoprecipitation. *OsNLP1* is therefore a feasible target for raising NUE and rice production [50].

2.19 Characterization of NLP Gene Family in Watermelon (*Citrullus lanatus*)

An analysis of the whole watermelon genome revealed three NLP genes. This indicates the variety in the number of NLP between species. Watermelon had a much lower family number of NLP than *.thaliana* (9), rice (6), maize (9), and *B. napus* (31). Using phylogenetic analysis, 3 *ClNLPs* were separated into three groups, which is in line with previous results. Five exons were found in all three *ClNLP* genes, which was in line with the four or five exons seen in the majority

of NLP genes. With the exception of the C-terminal PB1 domain structure, all of the *ClNLPs* that were investigated in this work possessed RWP-RK and PB1 domains that were conserved, which may have contributed to the differences in their functions compared to those of the NLP family members of other species. This study found that *ClNLP1* expression levels reached their highest point in the roots of numerous nitrogen-consuming varieties of watermelons and under a variety of different conditions regarding the availability of nitrogen. Based on this discovery, it appears that *ClNLP1* is the protein that is accountable for controlling the nitrogen metabolic pathway in watermelons. It was discovered that *ClNLP1* in particular was connected to the majority of genes and positively affected the expression of *ClNRT2.1*. This finding lends credence to the idea that it is a crucial component in the process of regulating the watermelon N response [44] .

2.20 *Solanum lycopersicum* (Tomato) NLP Gene Family Characterization

In this work, six NLP members were found in the tomato genome by genome-wide study. The phylogenetic tree showed that these individuals were divided into three clades. *SlNLP5* is unique for the double RWP-RK and PB1 domains. Orthologous gene pairs that are connected with *SlNLP1*, *SlNLP2*, or *SlNLP5* were shown to have existed prior to the divergence of the ancestral lineage of monocotyledonous and dicotyledonous plants.

Tissue-dependent expression patterns demonstrated that all six *SlNLP* genes were expressed in all of the tissues that were examined, which included the roots, stems, leaves, flowers, and fruits. *SlNLP3*, which is a close homolog of *AtNLP6/7*, was found to be strongly expressed in roots during both the seedling and flowering stages. Additionally, *SlNLP4* and *SlNLP6* were found to exhibit preferential expression in stems and leaves, and *SlNLP6* was found to be expressed at high levels in fruits.

Chapter 3

Methodology

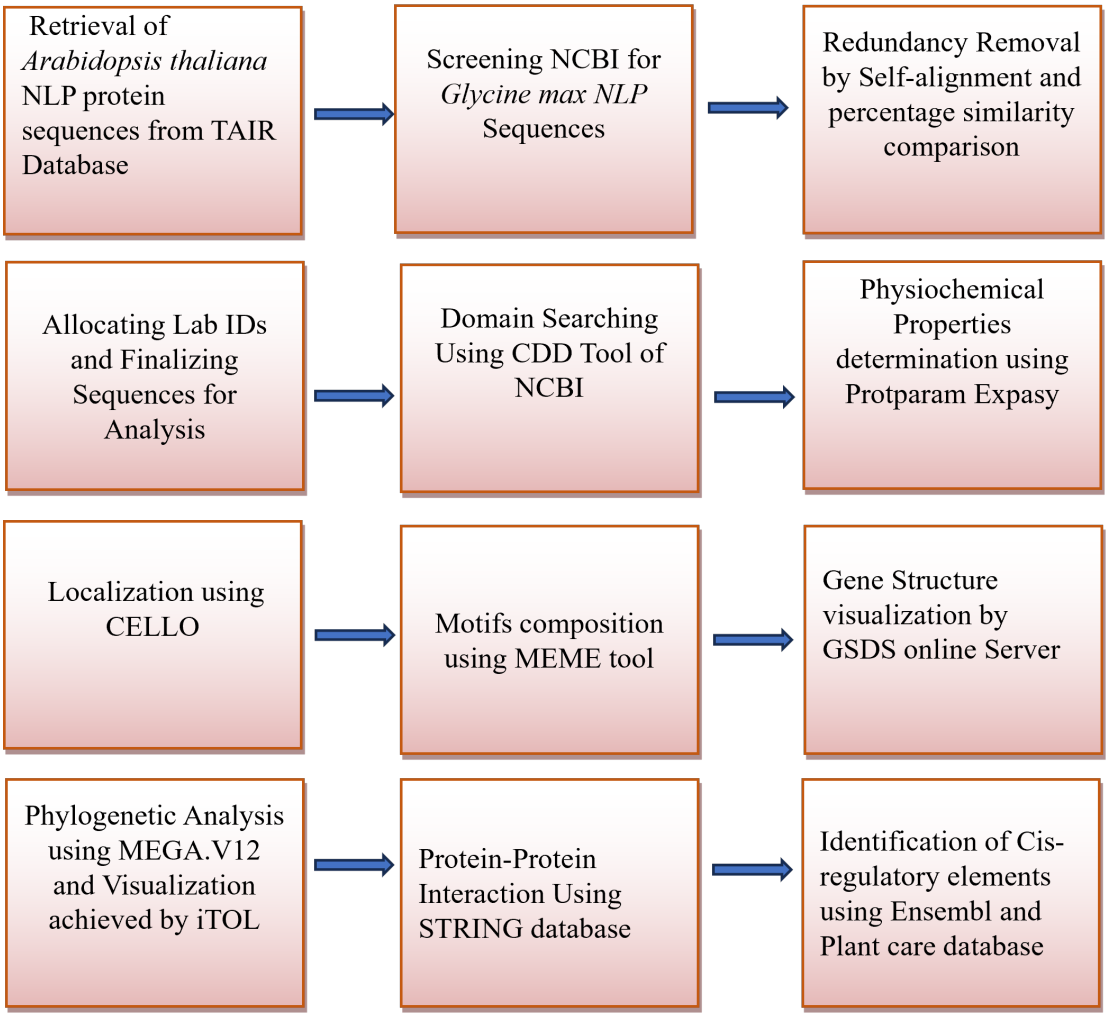

FIGURE 3.1: Overview of Methodology

3.1 Database Screening for Genes and Transcription Factors

The TAIR genomic library for *Arabidopsis thaliana* (https://www.arabidopsis.org/) was used to get the whole gene, amino acid, and coding sequence for each member of the *Arabidopsis thaliana NLP* gene family.

The recognition of potential *AtNLPs* was investigated using three protein databases: The NCBI (www.ncbi.nlm.nih.gov), TFDB (https://planttfdb.gao-lab.org/), and TAIR (https://www.arabidopsis.org/).

3.2 GmNLP Protein Sequences Retrieval

The *AtNLP* protein sequences were utilized as query sequences in two databases, NCBI and TFDB, to identify the *NLP* protein sequences for *Glycine max*. NCBI was scanned using accession numbers for *GmNLP* protein sequences.

The sequences were obtained in FASTA format and were saved for use in all sampling and analysis as scale. Two databases were searched to identify and characterize any suspected *GmNLPs*. All *GmNLP* sequences were retrieved and aligned.

3.3 *GmNLP* Sequence Redundancy Removal

All the protein sequences of *GmNLP* were aligned, in order to remove the redundant, splice and incomplete sequences. The repetition among the accession numbers of each sequence taken from different databases was removed. All the sequences whether from NCBI or TFDB were rearranged and matched with the identical one from other database to ensure no sequence is repeated and samples are specific and accurate.

3.4 Allocating Lab IDs and Establishing Sequence Similarity of *GmNLP* Protein Sequences

Following the elimination of redundant, spliced, and incomplete sequences, each sequence was assigned a Lab ID. Second, a comparison table was created to identify sequence similarities between *GmNLP* protein sequences. Sequences with a similarity of less than 80% were excluded.

3.5 Domain Searching for *Glycine max* NLPs/ Conserved Domain Side Nitrification in *GmNLPs*

For domain identification, the *AtNLP* scale sequences and the chosen *Glycine max* sample sequences (*GmNLPs*) were uploaded. The CD Database at the NCBI, or CDD (https://www.ncbi.nlm.nih.gov/cdd), was utilized to do this. Sequences were sent to CD-search in bulk, and the conserved domains were used to choose the possible *GmNLPs* from the retrieved sequences. The results described the common domain between the sequences, showing that the sequences are from the same family. The selected genes had both PB1 and RWP-RK domains.

3.6 Physiochemical Properties and Localization of *GmNLP*

The physiochemical characteristics of the sequences that are chosen based on conserved domains were examined further using metrics like theoretical isoelectric point, (pI), molecular weight, and GRAVY values. This was achieved by an online tool called Protparam Expasy (https://web.expasy.org/protparam/). Their

localization within the cell was anticipated utilizing another online tool known as CELLO (http://cello.life.nctu.edu.tw/).

3.7 Motif Composition in *GmNLP* Gene Family

An online tool known as MEME (meme-suite.org/tools/meme) v5.5.7, was utilized to determine whether a consensus motif is present. It is the most accurate online motif elicitation tool.

All parameter settings were left unchanged with the exception of the motif-finding threshold, which was kept at 15 to ensure specificity and accuracy. The identification of consensus motif will provide evidence of local regions that are similar between scale and study organism.

3.8 Phylogenetic Analysis of *GmNLP*

The MEGA-X v12 tool (https://www.megasoftware.net/index.phpz) was used to align a variety of species, including *Oryza sativa*, *Arabidopsis thaliana*, *Physcomitrella patens* and *Brassica napus* etc, in order to generate the phylogenetic tree for *GmNLP* genes. The tree generated from MEGA.X V12 tool was used as a query for visual representation of the phylogenetic tree in another online tool called Interactive Tree of Life v6 (https://itol.embl.de/). This tool illustrates the evolutionary link between *GmNLP* and other plants.

3.9 Gene Structure Determination in *GmNLP* Gene Family

The full-length genome and coding sequences of *GmNLPs* were used to investigate gene structural components using GSDS software (https://gsds.gao-lab.org/). This tool aid in identifying introns, exons, and untranslated regions (UTRs) within the sequences.

3.10 Protein Protein Interactions in *GmNLP* Family

The *GmNLP* protein sequences were analyzed. Proteins do not act independently, instead they work in the form of a group. The network was predicted online using STRING (https://string-db.org/), and their figures and descriptions were exported for comparison and future reference. It was postulated that all *GmNLP* proteins will interact with multiple N-interacted genes. The proteins that interact with *NLP* are responsible for nitrogen absorption, utilization, and transport.

3.11 Identification of Cis-regulatory elements in *GmNLPs*

Gene regulatory elements in *GmNLP* promoter regions were identified by analyzing upstream promoter regions (2000 bps) obtained from an online database, Plant Ensembl (https://plants.ensembl.org/index.html). Cis-regulatory elements were examined in promoter regions using a database (bioinformatics.psb.ugent.be/plantcare.html).

Chapter 4

Results

4.1 Identification and Evaluation of *NLP* Genes in Soybean

In this study, the complete protein coding sequences of the *NLP* gene from the *A. thaliana NLP* gene family were obtained by entering the accession number of *NLP* on the TAIR database (TAIR: https://www.arabidopsis.org/). There were 9 variants of the *NLP* gene and the longest sequence was chosen, then the longest sequence was protein Blast in the NCBI.

NCBI database (https://www.ncbi.nlm.nih.gov/) was examined to find *NLPs* in *Glycine max* genome (Taxonomy ID: 3847) using RWP-RK and PB1 domains as queries, in addition to *A. thaliana NLPs* protein sequences. Ninety sequences were first acquired via NCBI. All the results were downloaded and saved with names, accession numbers and Lab Ids in an excel file. They were self-aligned to make a percentage similarity table of all downloaded sequences. Short or incomplete fragments, repetitive or duplicated sequences, and spliced variants were eliminated from the obtained sequences that were concurrently verified by conserved domain identification. Eight *GmNLPs* with both PB1 and RWP-RK domains were ultimately found and assigned labels 1 to 8. Protein sequences of these 8 *GmNLPs* were downloaded in FASTA format.

4.2 Conserved Domain Identification

The common domain between the scale organism *Arabidopsis thaliana* and the study organism *Glycine max* for this particular gene family was found to be RWP-RK (PF02042) and PB1(cd06407) domains.

The CDD tool (https://www.ncbi.nlm.nih.gov/cdd), was used for this for this purpose, where the *A. thaliana's NLP* protein sequences in FASTA format were combined with the *NLP* protein sequences of *G. max* and uploaded in bulk search. The results show two common domains as shown in Table 4.2 confirming that the common domain of *NLP* as present in *A. thaliana* protein sequences, is present in *G. max* sequences. The result is summarized in (Table 4.1) which includes query, hit type, E-value, Accession, Bit score and position of both plants.

TABLE 4.1: NLP gene family conserved domains in *Glycine max* and *A. thaliana*

Organism	Query	Hit type	Position		E-value	Bitscore	Accession	Short name
			From	To				
Arabidopsis thaliana	*AtNLP1*	Specific	812	893	6.21E-41	144.773	cd06407	PB1_NLP
		Specific	608	656	1.10E-23	94.0888	pfam02042	RWP-RK
	AtNLP2	Specific	864	944	1.19E-41	146.699	cd06407	PB1_NLP
		Specific	648	696	1.82E-23	93.7036	pfam02042	RWP-RK
	AtNLP3	Specific	674	758	1.48E-40	143.232	cd06407	PB1_NLP
		Specific	498	546	1.45E-23	93.7036	pfam02042	RWP-RK
	AtNLP4	Specific	745	826	6.79E-43	150.166	cd06407	PB1_NLP
		Specific	558	606	1.28E-23	94.0888	pfam02042	RWP-RK
	AtNLP5	Specific	711	787	3.53E-36	130.906	cd06407	PB1_NLP
		Specific	549	597	3.77E-24	95.6296	pfam02042	RWP-RK
	AtNLP6	Specific	742	822	2.85E-34	125.513	cd06407	PB1_NLP
		Specific	556	604	4.91E-24	95.2444	pfam02042	RWP-RK
	AtNLP7	Specific	864	944	4.11E-34	125.128	cd06407	PB1_NLP
		Specific	591	639	1.20E-24	97.1704	pfam02042	RWP-RK
	AtNLP8	Specific	835	915	6.63E-39	138.995	cd06407	PB1_NLP
		Specific	590	638	1.37E-24	96.7852	pfam02042	RWP-RK
	AtNLP9	Specific	793	874	3.20E-34	125.513	cd06407	PB1_NLP
		Specific	535	583	2.37E-24	96.0148	pfam02042	RWP-RK
	GmNLP1	Specific	817	898	7.19E-40	141.692	cd06407	PB1_NLP
		Specific	597	645	3.95E-23	92.9542	pfam02042	RWP-RK
	GmNLP2	Specific	719	800	3.15E-39	139.766	cd06407	PB1_NLP
		Specific	518	597	5.11E-17	75.235	pfam02042	RWP-RK
		Specific	290	403	3.96E-05	43.9944	pfam13185	GAF_2
	GmNLP3	Specific	961	1041	2.73E-37	134.373	cd06407	PB1_NLP
		Specific	650	698	6.59E-24	95.2654	pfam02042	RWP-RK

continued on next page

Table 4.1 continued from previous page

Organism	Query	Hit type	Position From	Position To	E-value	Bitscore	Accession	Short name
Glycine max	*GmNLP4*	Specific	983	1063	2.19E-34	126.284	cd06407	PB1_NLP
		Specific	672	720	1.57E-22	91.4134	pfam02042	RWP-RK
	GmNLP5	Specific	603	687	7.90E-38	135.143	cd06407	PB1_NLP
		Specific	499	547	9.17E-23	91.4134	pfam02042	RWP-RK
	GmNLP6	specific	940	1020	5.94E-31	116.268	cd06407	PB1_NLP
		Specific	670	718	3.40E-23	93.3394	pfam02042	RWP-RK
	GmNLP7	Specific	519	602	3.92E-39	138.225	cd06407	PB1_NLP
		Specific	408	456	2.57E-22	89.8726	pfam02042	RWP-RK
	GmNLP8	Specific	905	985	2.94E-30	114.342	cd06407	PB1_NLP
		Specific	641	689	5.72E-23	92.569	pfam02042	RWP-RK

4.3 Physiochemical Properties of the *NLP* Genes in the *Glycine max* Genome

Eight *NLP* genes, which we named *GmNLP1* to *GmNLP8*, were found in the soybean genome based on the existence of conserved RWP-RK and PB1 domains. Table 4.2 displayed the gene properties, chromosomal locations, protein length, molecular weight (MW), isoelectric point (pI), and subcellular locations.

Compared to *AtNLPs*, *GmNLPs* were shown to have greater gene lengths, protein lengths, and molecular weights (MW). The average gene lengths of *AtNLPs* and *GmNLPs* were 4141 and 4743 bp, respectively. *AtNLPs* and *GmNLPs* were found to have significantly different protein lengths, with averages of 880 and 905 amino acids, respectively. The average MW of *GmNLPs* was 100612 MW, which was higher than the average MW of *AtNLPs*, which was 97749.

The average pI and GRAVY values of both plants were close to each other. All *AtNLPs* (with the exception of *AtNLP3*) and *GmNLPs* had pI values less than 7, which showed that they were acidic proteins, however *AtNLP3* had a pI value of 8.14, which suggested that it was a basic protein.

The examination of the sub-cellular localization of both *A. thaliana* and *G. max* *NLPs* suggested that they were located in the nucleus.

Furthermore, all of the *NLPs* from both plants showed negative GRAVY values, which indicated that *NLPs* are hydrophilic proteins. *GmNLP1* and *GmNLP2* were distributed on Chr06 and Chr4, while *GmNLP3/4/5/6/7/8* were distributed on Chr20, Chr16, Chr02, Chr15, Chr14, and Chr11, respectively.

TABLE 4.2: Physiochemical characteristics of the *A. thaliana* and *G. max* NLP gene families

Plant	Gene name	Chr.	Position	Gene length (bp)	Protein length (aa)	Molecular weight	Iso - electric point	GRAVY	Loc
Arabidopsis thaliana	AtNLP1	2	7466687-7471586	4900	909	100885.3	4.83	-0.443	Nucleus
	AtNLP2	4	16777264 - 16782054	4791	963	107277.6	5.76	-0.476	Nucleus
	AtNLP3	4	17954710 - 17958063	3354	767	85065.7	8.14	-0.271	Nucleus
	AtNLP4	1	7154425 - 7158284	3860	844	94231.1	5.45	-0.472	Nucleus
	AtNLP5	1	28639453 - 28643086	3634	808	90683.4	6.13	-0.467	Nucleus
	AtNLP6	1	23959627 - 23963083	3457	841	93862.6	6.3	-0.356	Nucleus
	AtNLP7	4	12479528 - 12484049	4522	959	105741.1	5.69	-0.42	Nucleus
	AtNLP8	2	18061716 - 18066692	4977	934	103284.1	5.45	-0.436	Nucleus
	AtNLP9	3	22009010 - 22012791	3782	894	98712.1	5.29	-0.383	Nucleus
Glycine max	*GmNLP1*	6	1338317 - 1343018	4702	909	101290	5.6	-0.464	Nucleus
	GmNLP2	4	72328 -75769	3442	816	90237.01	5.51	-0.246	Nucleus
	GmNLP3	20	39860747 - 39866224	5478	1055	117013.15	5.73	-0.313	Nucleus
	GmNLP4	16	34323618-34328331	4714	1082	120147.94	6.23	-0.272	Nucleus
	GmNLP5	2	48440620 - 48443772	3153	710	79084.29	5.87	-0.312	Nucleus
	GmNLP6	15	2228855 - 2237078	8224	1039	115034.17	5.87	-0.354	Nucleus
	GmNLP7	14	202374 - 204618	2245	628	69883	6.12	-0.307	Nucleus
	GmNLP8	11	9530122 - 9536109	5988	1004	111208.24	5.8	-0.431	Nucleus

4.4 Gene Structure Determination

The gene and coding sequence of *AtNLPs* and *GmNLPs* were used to analyze their structural features which included the identification of exons, introns and untranslated regions (UTRs). This was done using an online tool called GSDS (https://gsds.gao-lab.org/).

It was observed that the lowest number of exons present was in *GmNLP5*, which was 2, followed by the presence of 3 exons in *GmNLP7*. The number of exons for the rest of the *GmNLPs* range from 4 to 5, with *GmNLP1* and *GmNLP2* containing 4 exons each, while 5 exons were observed in *GmNLP3, GmNLP4, GmNLP6 and GmNLP8*.

The range of exons in *AtNLPs* is four to six, with *AtNLP1, AtNLP2, AtNLP3, AtNLP4* and *AtNLP5* were observed to have 4 exons each, while 5 exons were observed in *AtNLP7* and *AtNLP9*.

The highest number of exons were present in *AtNLP6* and *AtNLP8* which was 6. It was also observed that *AtNLP3* and *GmNLP7* do not have a 5'UTR.

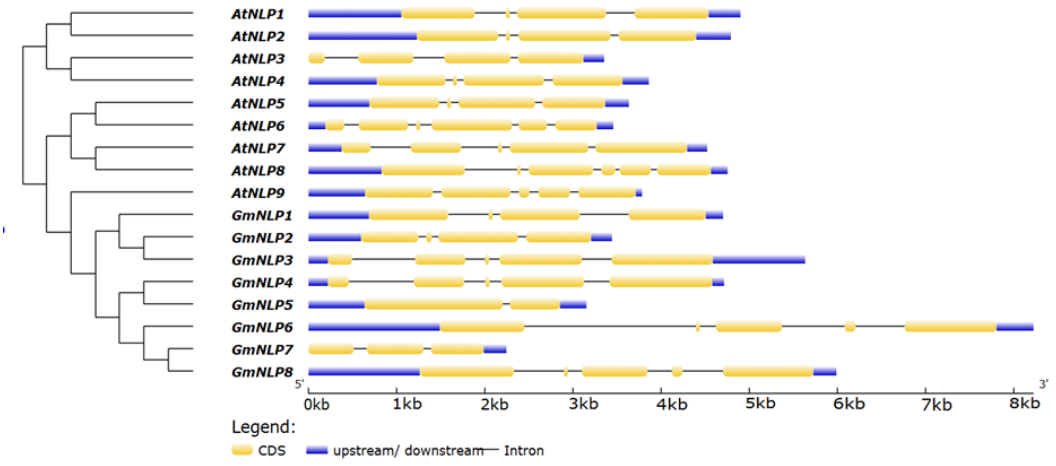

FIGURE 4.1: Gene Structure Determination of *AtNLP* and *GmNLP*

4.5 Consensus Motifs Composition

MEME was used to identify up to 15 conserved motifs in *GmNLP* proteins when compared to *AtNLPs*. Significantly conserved motifs were present in all the sequences of both the *G. max* and *A. thaliana* proteins.

GmNLP1,3,4,6 and 8 contain 13 motifs, *GmNLP2* and 5 contain 9 motifs, while *GmNLP7* contain 8 motifs. Similarly, *AtNLP1* contain 12 motifs, *AtNLP2*,4,5,6,7,8 and 9 contain 13 motifs, while *AtNLP3* contain only 10 motifs.

No protein had all 15 motifs present in them, indicating their uniqueness and specificity while the presence of the majority of motifs in all *GmNLPs* indicate that they indeed belong to the same family and are clued to be similar in function. Figure 4.2 shows all of the 15 motifs and their sequences.

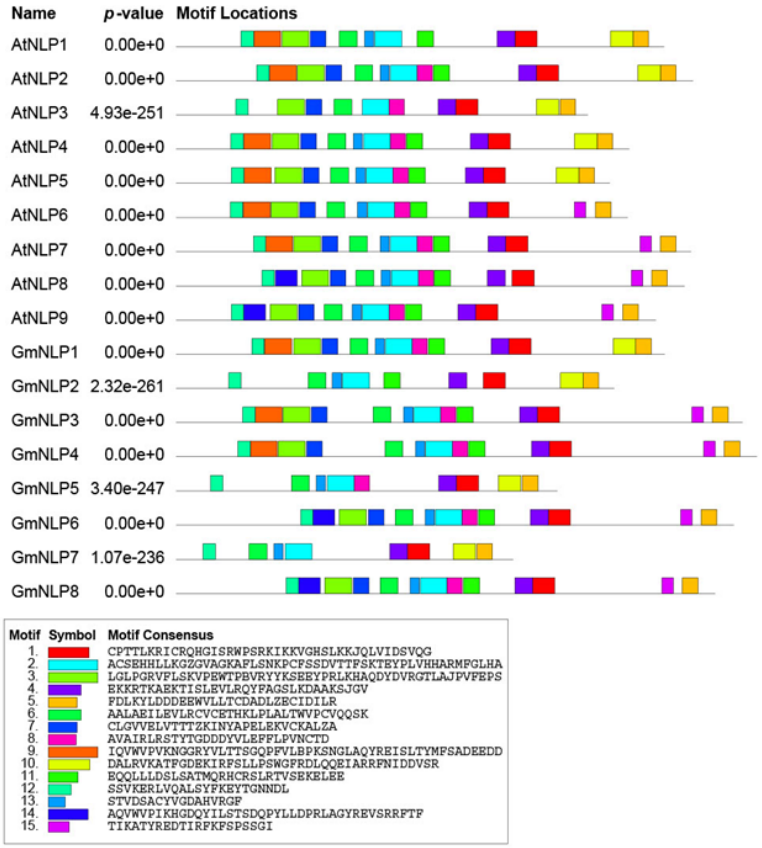

FIGURE 4.2: Identification of Consensus Motifs of *AtNLPs* and *GmNLPs*

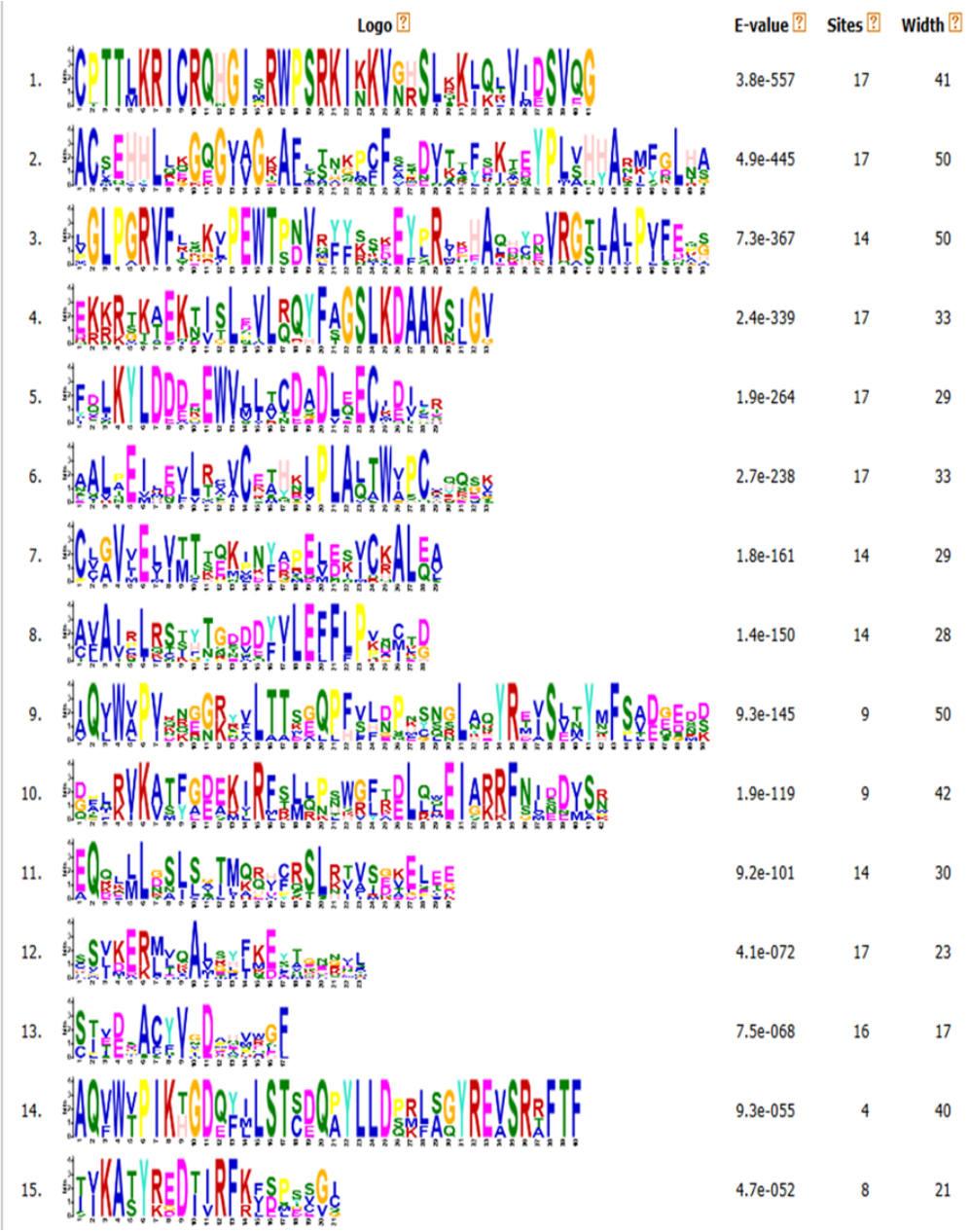

FIGURE 4.3: Logos of Consensus Motifs of *AtNLP* and *GmNLP*

4.6 Protein-Protein Interactions

The STRING database (https://string-db.org/) was used to study PPI, and the *NLP* interacting protein network was predicted using it. It was proposed that all of the *GmNLP* proteins interacted with a number of genes associated to nitrogen.

Two GRAS domain proteins, NODULATION SIGNALING PATHWAY 1 (*GMNSP1*) and (*GMNSP2*) have been identified. NSP1 and NSP2 are necessary for the development of nodules and Nod-factor signaling. GMARF: Additionally, auxin response factor has been discovered, which is implicated in DNA binding and contributes to signal transduction, growth, and development. *GmNLP4* had two interacting proteins, both of which were J domain-containing proteins. *GmNLP2*, *GmNLP3* and *GmNLP5* contained several uncharacterized proteins. The colour lines shows known and predicted interactions. The balls that are in different colours shows nodes, while the coloured lines are the edges means protein sequences. In PPI, the proteins interacting with *NLP* are those involved in N uptake, assimilation and transport; thus, they predict the role in overall NUE. It has been proposed that all *GmNLP* proteins interact with several N-related genes.

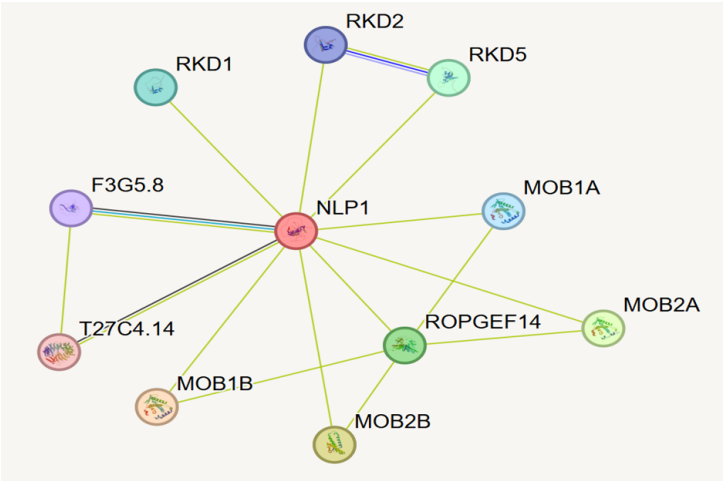

FIGURE 4.4: Results from protein-protein interaction analysis of *AtNLP1*

TABLE 4.3: Protein-protein interaction of *AtNLP1*

Node	**Annotation**
MOB1B	Activator-like MOB kinase 1B; a member of the MOB1/phocein family.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
ROPGEF14	Guanine-nucleotide exchange factor 14, also known as GEF, is a guanine-nucleotide exchange factor that promotes the exchange of GDP for GTP. It is an activator of Rop GTPases, which are responsible for the Rho of plants.
RKD5	Protein RKD5; Putative transcription factor.
RKD1	Protein RKD1; Putative transcription factor.

Table 4.3 continued from previous page

Node	Annotation
MOB1A	MOB kinase activator-like 1A; is essential for controlling the division and growth of cells.
RKD2	Protein RKD2; Putative transcription factor.
F3G5.8	Ypt/Rab-GAP domain of the protein belonging to the gyp1p superfamily.
T27C4.14	Exportin-4 protein.

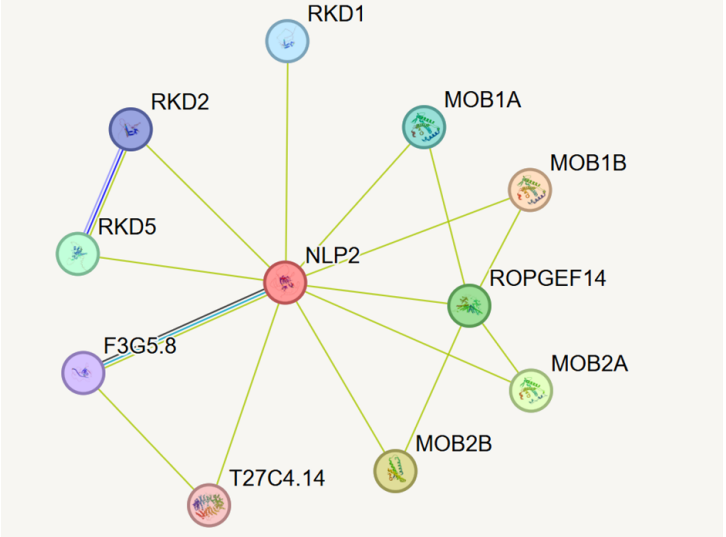

FIGURE 4.5: Results from protein-protein interaction analysis of *AtNLP2*

TABLE 4.4: Protein-protein interaction of *AtNLP2*

Node	Annotation
MOB1B	MOB kinase activator-like 1B; Belongs to the MOB1/phocein family.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
ROPGEF14	Rop guanine nucleotide exchange factor 14 is a guanine-nucleotide exchange factor (GEF) that promotes the exchange of GDP for GTP with the purpose of acting as an activator of Rop (Rho of plants) GTPases.
RKD5	Protein RKD5; Putative transcription factor.
MOB1A	MOB kinase activator-like 1A; is essential for controlling the division and growth of cells.
RKD1	Protein RKD1; Putative transcription factor.
RKD2	Protein RKD2; Putative transcription factor.
F3G5.8	Ypt/Rab-GAP domain of the protein belonging to the gyp1p superfamily.
T27C4.14	Exportin-4 protein.

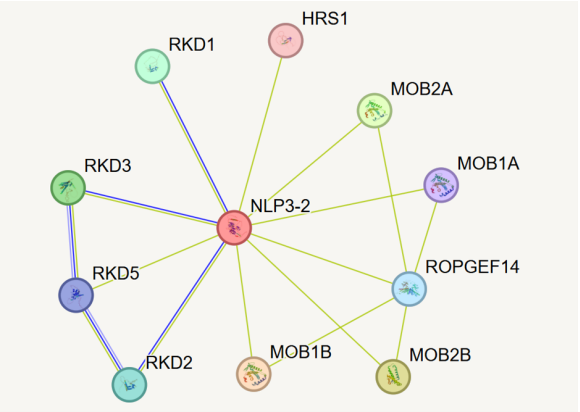

FIGURE 4.6: Results from protein-protein interaction analysis of *AtNLP3*

TABLE 4.5: Protein-protein interaction of *AtNLP3*

Node	**Annotation**
MOB1B	The MOB1/phocein family includes MOB kinase activator-like 1B.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
RKD3	Protein RKD3; Putative transcription factor.
RKD1	Protein RKD1; Putative transcription factor.
RKD2	Protein RKD2; Putative transcription factor.
ROPGEF14	Guanine-nucleotide exchange factor 14, also known as GEF
RKD5	Protein RKD5; Putative transcription factor.
MOB1A	MOB kinase activator-like 1A; plays a crucial part in controlling cell division and growth.
HRS1	Transcription factor HRS1; Transcription factor that plays a role in roots' nitrate and phosphate signalling.

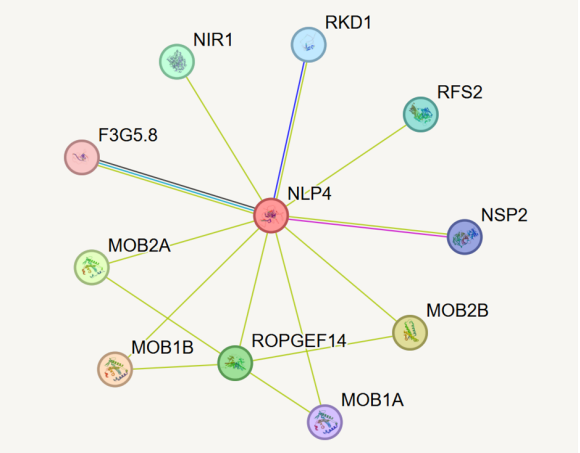

FIGURE 4.7: Results from protein-protein interaction analysis of *AtNLP4*

TABLE 4.6: Protein-protein interaction of *AtNLP4*

Node	Annotation
MOB1B	Activator-like MOB kinase 1B is a member of the MOB1/phocein family.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
ROPGEF14	By encouraging the conversion of GDP into GTP, Rop guanine nucleotide exchange factor 14—also known as Guanine-nucleotide exchange factor (GEF)—activates Rop (plant Rho) GTPases.
NIR1	Ferredoxin is a chloroplastic nitrite reductase that catalyses the conversion of nitrite to ammonium by six electrons.
RFS2	Most likely, transglycosidase functions by a ping-pong reaction mechanism, as does galactinol—sucrose galactosyltransferase 2.
RKD1	Protein RKD1; Putative transcription factor.
NSP2	Nitrile-specifier protein 2; Promotes simple nitriles, but not epithionitrile or thiocyanate formation.
MOB1A	MOB kinase activator-like 1A; is essential for controlling the division and growth of cells.
F3G5.8	Ypt/Rab-GAP domain of gyp1p superfamily protein.

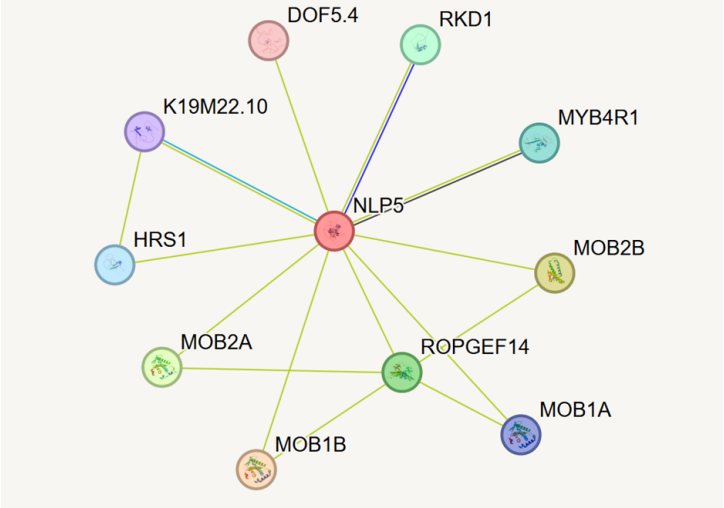

FIGURE 4.8: Results from protein-protein interaction analysis of *AtNLP5*

TABLE 4.7: Protein-protein interaction of *AtNLP5*

Node	Annotation
MOB1B	Activator-like MOB kinase 1B is a member of the MOB1/phocein family.
MOB2B	Putative MOB kinase activator-like 2B.

Table 4.7 continued from previous page

Node	Annotation
MOB2A	MOB kinase activator-like 2A.
ROPGEF14	Guanine-nucleotide exchange factor 14, also known as GEF, is a guanine-nucleotide exchange factor that promotes the exchange of GDP for GTP. It is an activator of Rop GTPases, which are responsible for the Rho of plants.
RKD1	Protein RKD1; Putative transcription factor.
MYB4R1	Putative transcription factor.
HRS1	Transcription factor HRS1; A transcription factor that has a role in the signalling systems of nitrate and phosphate in roots.
MOB1A	MOB kinase activator-like 1A; performs an essential function in the control of the growth and division of cells.
K19M22.10	Similarity to Myb-related transcription factor.
DOF5.4	A transcription factor known as Dof zinc finger protein DOF5.4 selectively binds to a consensus core sequence of 5'-AA[AG]G-3'.

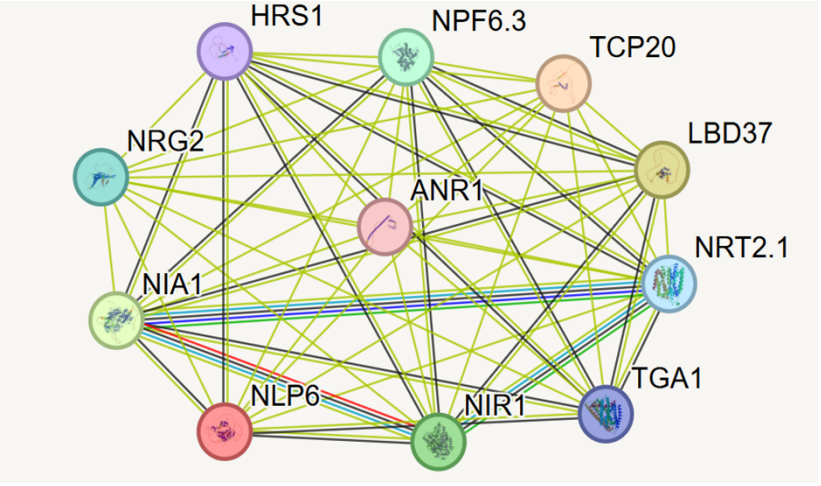

FIGURE 4.9: Results from protein-protein interaction analysis of *AtNLP6*

TABLE 4.8: Protein-protein interaction of AtNLP6

Node	Annotation
TCP20	TCP20 is a transcription factor that attaches itself to the site II motif (3'-TGGGCC/T-5') in the PCNA-2 promoter and to the 3'-GCCCG/A-5' elements in the cyclin CYCB1-1 and ribosomal protein gene promoters.
NIA1	Nitrate reductase [NADH] 1 is an essential enzyme implicated in the initial stage of nitrate uptake in bacteria, fungi, and plants.
NIR1	Nitrite reductase, a chloroplastic, is catalysed by ferredoxin to reduce nitrite to ammonium by six electrons.

Table 4.8 continued from previous page

Node	Annotation
NPF6.3	Protein NRT1/PTR Family 6.3: Nitrate transporter with dual affinity.
NRG2	Nitrate regulatory gene2 protein; Required for nitrate signaling.
NRT2.1	Nitrate transport is mediated by high-affinity nitrate transporter 2.1, however it does not appear to be capable of mediating transport on its own.
TGA1	The transcription factor TGA1 is a transcriptional activator that binds to the 5'-TGACG-3' DNA sequence.
HRS1	Transcription factor HRS1; Transcription factor that plays a role in roots' nitrate and phosphate signalling.
ANR1	Probable transcription factor; MADS-box transcription factor ANR1. necessary for the nitrate-responsive plasticity of roots.

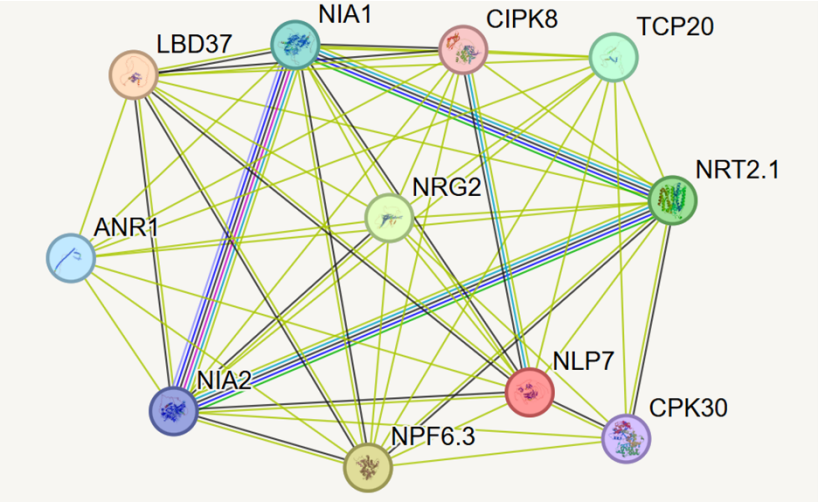

FIGURE 4.10: Results from protein-protein interaction analysis of *AtNLP7*

TABLE 4.9: Protein-protein interaction of *AtNLP7*

Node	Annotation
LBD37	LOB domain-containing protein 37.
NPF6.3	Protein NRT1/PTR Family 6.3: Nitrate transporter with dual affinity. involved in controlling the nitrate transporter NRT2.1 and proton-dependent nitrate absorption.
NRG2	Nitrate regulatory gene2 protein; Required for nitrate signaling.
NRT2.1	Nitrate transport is mediated by high-affinity nitrate transporter 2.1, however it does not appear to be capable of mediating transport on its own.

Table 4.9 continued from previous page

Node	Annotation
TCP20	The transcription factor is capable of binding to the site II motif (3'-TGGGCC/T-5') in the PCNA-2 promoter and to the 3'-GCCCG/A-5' elements in the cyclin CYCB1-1 and ribosomal protein gene promoters.
NIA1	One important enzyme in the initial stage of nitrate assimilation is nitrate reductase [NADH] 1.
ANR1	Probable transcription factor; MADS-box transcription factor ANR1. necessary for the flexibility of roots in reaction to nitrate
NIA2	One important enzyme in the initial stage of nitrate assimilation is nitrate reductase [NADH] 2.
CPK30	It is possible that calcium-dependent protein kinase 30 is involved in signal transduction pathways where calcium is used as a second messenger.
CIPK8	CBL proteins interact with CIPK serine-threonine protein kinases and CBL-interacting serine/threonine-protein kinase 8.

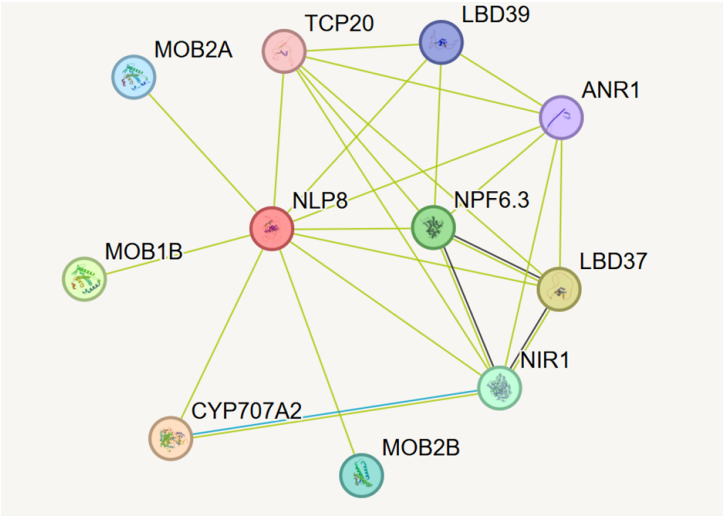

FIGURE 4.11: Results from protein-protein interaction analysis of *AtNLP8*

TABLE 4.10: Protein-protein interaction of *AtNLP8*

Node	Annotation
CYP707A2	Abscisic acid 8'-hydroxylase 2: involved in abscisic acid's oxidative breakdown
LBD37	LOB domain-containing protein 37.
MOB1B	Activator-like MOB kinase 1B is a member of the MOB1/phocein family.
NPF6.3	The dual affinity nitrate transporter, also known as the protein NRT1/PTR family 6.3. The nitrate transporter NRT2.1 is regulated by this protein, which is also involved in the proton-dependent absorption of nitrate.

Table 4.10 continued from previous page

Node	Annotation
NIR1	Ferredoxin–nitrite reductase, chloroplastic; catalyses the conversion of nitrite to ammonium by six electrons.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
LBD39	LOB domain-containing protein 39.
ANR1	Probable transcription factor; MADS-box transcription factor ANR1. necessary for the flexibility of roots in reaction to nitrate
TCP20	The transcription factor is capable of binding to the site II motif (3'-TGGGCC/T-5') in the PCNA-2 promoter and to the 3'-GCCCG/A-5' elements in the cyclin CYCB1-1 and ribosomal protein gene promoters.

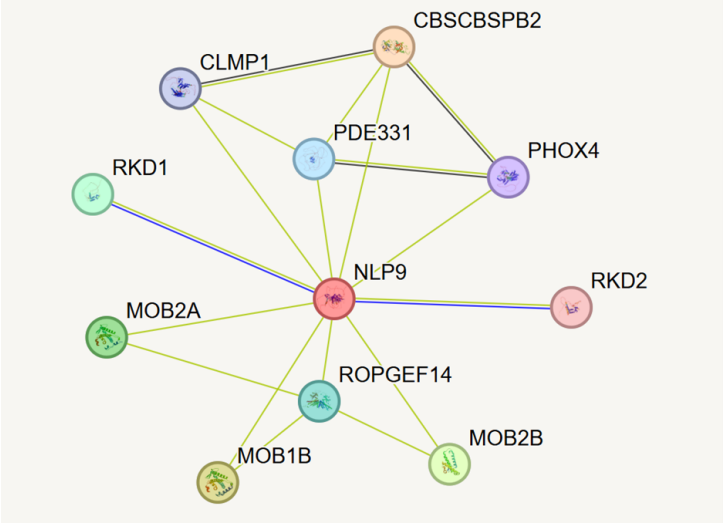

FIGURE 4.12: Results from protein-protein interaction analysis of *AtNLP9*

TABLE 4.11: Protein-protein interaction of *AtNLP9*

Node	Annotation
CBSCBSPB2	CBS domain-containing protein CBSCBSPB2.
MOB1B	Activator-like MOB kinase 1B is a member of the MOB1/phocein family.
MOB2B	Putative MOB kinase activator-like 2B.
MOB2A	MOB kinase activator-like 2A.
RKD1	Protein RKD1; Putative transcription factor.
ROPGEF14	By encouraging the conversion of GDP to GTP, Rop guanine nucleotide exchange factor 14—also known as Guanine-nucleotide exchange factor (GEF)—activates Rop (plant Rho) GTPases.
PDE331	Octicosapeptide/Phox/Bem1p family protein.

Table 4.11 continued from previous page

Node	Annotation
CLMP1	Protein CLMP1; Required for plastid separation and partitioning during cell division.
RKD2	Protein RKD2; Putative transcription factor.
PHOX4	Protein PHOX4 is a carboxylate clamp-type tetratricopeptide repeat protein that has the potential to function as a co-chaperone for Hsp90 and Hsp70. The polar development of root hairs is influenced by this factor.

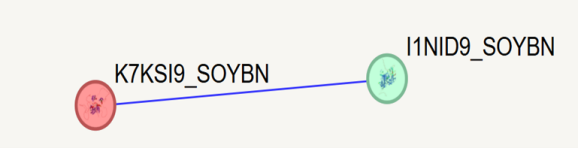

FIGURE 4.13: Results from protein-protein interaction analysis of *GmNLP1*

TABLE 4.12: Protein-protein interaction of *GmNLP1*

Node	Annotation
1NID9_SOYBN	PHD domain-containing protein.

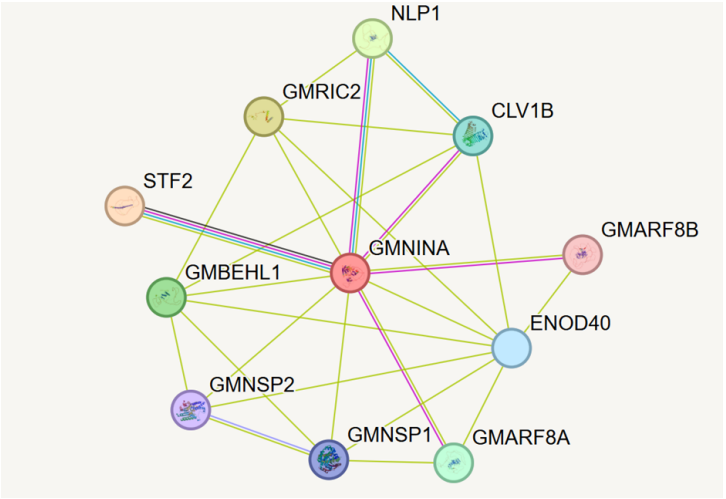

FIGURE 4.14: Results from protein-protein interaction analysis of *GmNLP2*

TABLE 4.13: Protein-protein interaction of *GmNLP2*

Node	Annotation
STF2	Uncharacterized protein.
GMRIC2	CLE35 protein.

Table 4.13 continued from previous page

Node	Annotation
NLP1	NAC domain-containing protein
GMBEHL1	BES1_N domain-containing protein.
GMARF8A	Auxin response factor; transcriptional factors known as auxin response factors (ARFs) bind selectively to the 5'-TGTCTC-3' DNA sequence present in auxin-responsive promoter elements (AuxREs).
CLV1B	Root nodule organogenesis and root and shoot development are regulated by the leucine-rich repeat receptor-like kinase protein CLV1B. actively participates in long-distance nodulation signalling.
ENOD40	Early nodulin-40; Modifies auxin's effect and may operate as a regulator of plant development, changing the reactions of phytohormones.
GMNSP1	GRAS domain-containing protein; A member of the GRAS family.
GMNSP2	GRAS domain-containing protein; A member of the GRAS family.
GMARF8B	DNA sequence 5'-TGTCTC-3', which is present in auxin-responsive promoter elements (AuxREs), is the particular binding site for auxin response factors (ARFs), which are transcriptional factors.

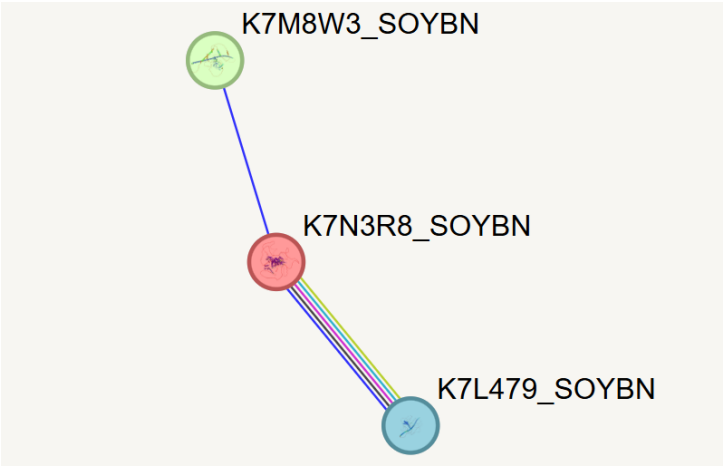

FIGURE 4.15: Results from protein-protein interaction analysis of *GmNLP3*

TABLE 4.14: Protein-protein interaction of *GmNLP3*

Node	Annotation
K7M8W3_SOYBN	Uncharacterized protein.
K7L479_SOYBN	J domain-containing protein

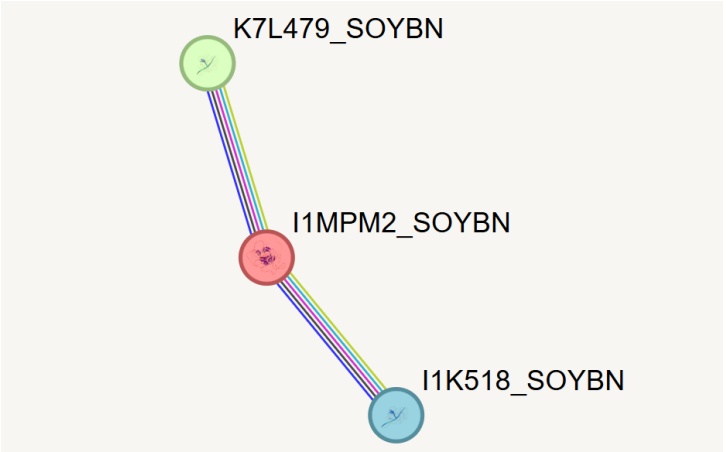

FIGURE 4.16: Results from protein-protein interaction analysis of *GmNLP4*

TABLE 4.15: Protein-protein interaction of *GmNLP4*

Node	**Annotation**
K7L479_SOYBN	J domain-containing protein.
I1K518_SOYBN	J domain-containing protein

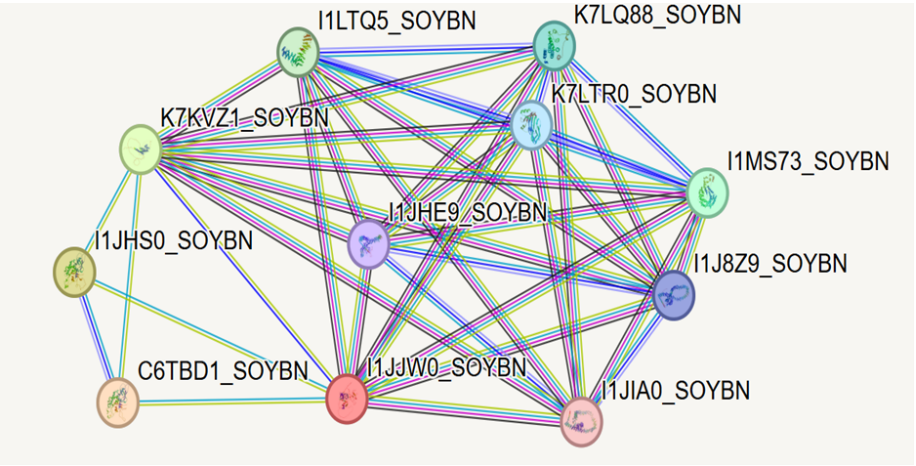

FIGURE 4.17: Results from protein-protein interaction analysis of *GmNLP5*

TABLE 4.16: Protein-protein interaction of *GmNLP5*

Node	**Annotation**
C6TBD1_SOYBN	Autophagy protein 5; Required for autophagy
I1JHS0_SOYBN	Autophagy protein 5; Required for autophagy.
K7KVZ1_SOYBN	RWP-RK domain-containing protein

Table 4.16 continued from previous page

Node	Annotation
1LTQ5_SOYBN	Alpha component of the AP-2 complexes; adaptor protein complex 2 subunit (AP-2). In several membrane traffic routes, adaptor protein complexes facilitate protein transport through transport vesicles.
I1MS73_SOYBN	Alpha component of the AP-2 complexes; adaptor protein complex 2 (AP-2) subunit. In several membrane traffic routes, adaptor protein complexes facilitate protein transport through transport vesicles.
K7LQ88_SOYBN	Uncharacterized protein
K7LTR0_SOYBN	Alpha_adaptinC2 domain-containing protein.
I1J8Z9_SOYBN	Clathrin heavy chain; Clathrin is the primary protein that make up the polyhedral coat of coated pits and vesicles; Clathrin is a member of the family of clathrin heavy chain proteins.
1JHE9_SOYBN	Belonging to the clathrin heavy chain family, clathrin is the main protein of the polyhedral coat of coated pits and vesicles.
I1JIA0_SOYBN	Clathrin heavy chain; a member of the clathrin heavy chain family; the main protein of the polyhedral coat of coated pits and vesicles.

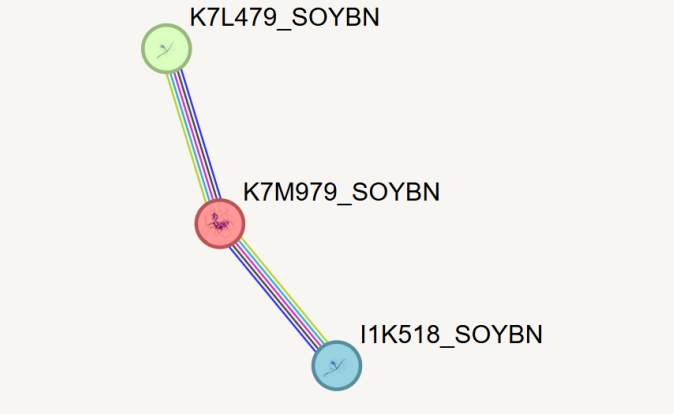

FIGURE 4.18: Results from protein-protein interaction analysis of *GmNLP6*

TABLE 4.17: Protein-protein interaction of *GmNLP6*

Node	Annotation
K7L479_SOYBN	J domain-containing protein
I1K518_SOYBN	J domain-containing protein

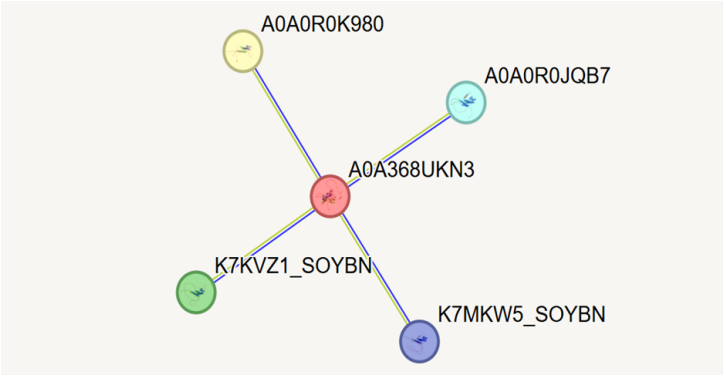

FIGURE 4.19: Results from protein-protein interaction analysis of *GmNLP7*

TABLE 4.18: Protein-protein interaction of GmNLP7

Node	**Annotation**
A0A0R0K980	RWP-RK domain-containing protein
K7KVZ1_SOYBN	RWP-RK domain-containing protein.
A0A0R0JQB7	RWP-RK domain-containing protein.
K7MKW5_SOYBN	RWP-RK domain-containing protein.

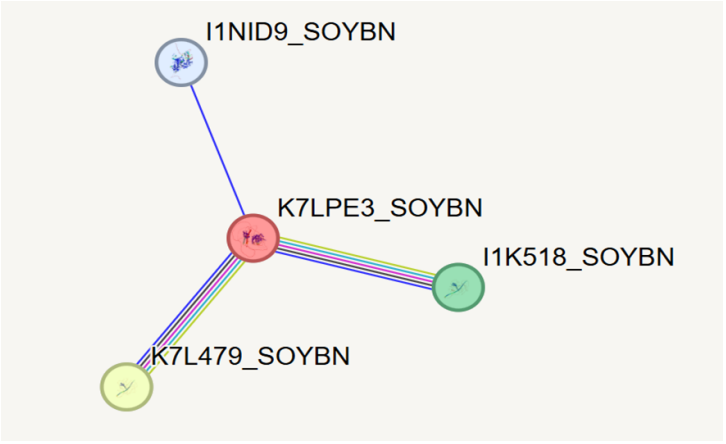

FIGURE 4.20: Results from protein-protein interaction analysis of *GmNLP8*

TABLE 4.19: Protein-protein interaction of *GmNLP8*

Node	**Annotation**
K7L479_SOYBN	Protein with J domain.
I1K518_SOYBN	Protein with J domain.
I1NID9_SOYBN	PHD domain-containing protein.

4.7 Sequence Alignment and Phylogenetic Analysis of *Glycine max*

The percentage similarities between *GmNLPs* and *AtNLPs* were compared to ensure the right selection and singularity of each found *GmNLP* gene for further study. All *AtNLPs* and *GmNLPs* have protein sequences with less than 70% similarity, ensuring each gene's uniqueness in addition to evolutionary diversity among *GmNLP* gene family members. Then sequence of different strains of NLP genes were taken from different plants: *A. thaliana, Physcomitrella patens* (spreading earthmoss) *Oryza sativa* (Rice), *Sorghum bicolor* (great millet) *Zea mays, Solanum lycopersicum*(Tomato), *Brassica napus* (rapeseed), *Populus trichocarpa* (black cottonwood), *Triticum aestivum* (Wheat).

Sequences were downloaded and then converted into FASTA file. MEGA-X v12 software is used to build the phylogenetic tree using default settings, 1000 bootstrap replicates, and the neighbor joining (NJ) technique, iTol (https://itol.embl.de/) is then used to recreate the tree. The evolutionary relationship of *GmNLP* with different species having NLP gene families were analyzed by the phylogenetic tree. Figure 4.21 shows the phylogenetic relationship between *G. max* and other plants. The NLP gene family of *Glycine max* showed evolutionary relationship with the other selected plants and their structural and functional characteristics are predicted to be more similar with *Populus trichocarpa*, *Triticum aestivum* and *Solanum lycopersicum*. The distribution of *G. max* NLP proteins across distinct clades indicates that these genes have undergone many duplication and divergence events. Their existence in distinct evolutionary branches adds to the likelihood of functional specialization. Proteins that cluster together (e.g., *GmNLP3* and *GmNLP4*) may have comparable regulatory activities, whereas proteins from different clades (e.g., *GmNLP1* vs. *GmNLP6*) may operate under different environmental circumstances or developmental stages.

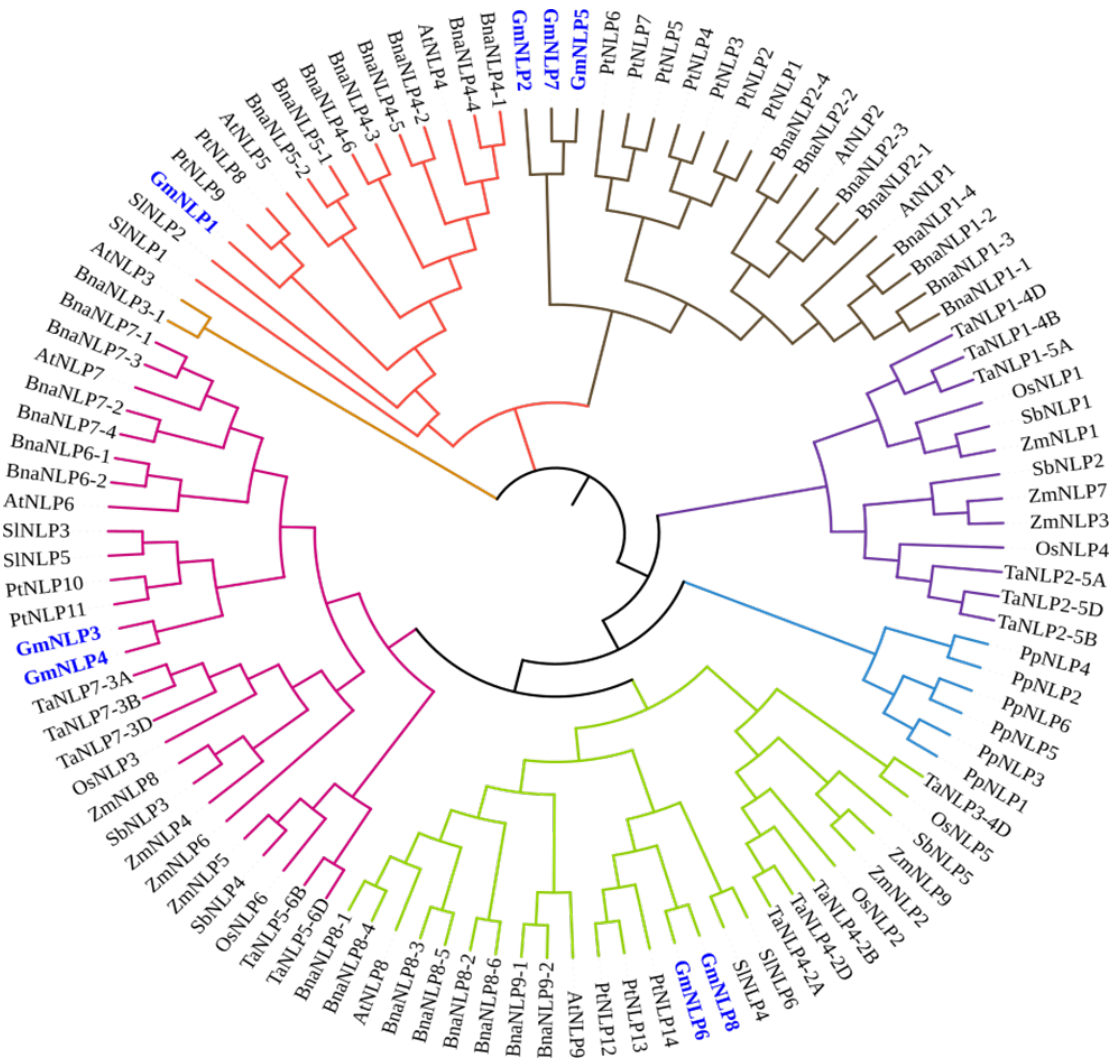

FIGURE 4.21: Phylogenetic Analysis of *GmNLPs* through neighbor joining method. The different colored arcs indicate different groups of *NLP* domains.

	AtNLP1	AtNLP2	AtNLP3	AtNLP4	AtNLP5	AtNLP6	AtNLP7	AtNLP8	AtNLP9	GmNLP1	GmNLP2	GmNLP3	GmNLP4	GmNLP5	GmNLP6	GmNLP7	GmNLP8
AtNLP1	100	68.66	44.13	46.39	41.26	35.88	36.54	37.5	37.08	42.93	40.94	33.77	34.62	41.53	37.11	38.09	36.8
AtNLP2	69.08	100	45.88	40.8	39.31	35.22	40.28	35.98	36.54	42.38	40.98	35.39	37.16	44.23	35.75	38.99	36.29
AtNLP3	44.13	45.43	100	39.17	40.58	31.53	33.99	32.66	33.65	39.72	35.87	31.71	34.32	31.07	34.72	28.07	34.03
AtNLP4	41.35	41.08	43.94	100	70.57	35.58	42.39	40.11	39.26	50	40.02	38.13	39.89	39.97	38.83	37.39	39.4
AtNLP5	45.04	43.24	41.25	69.33	100	36.3	44.19	38.46	39.67	47.11	38.72	41.88	42.34	39.97	41.32	37.12	39.16
AtNLP6	39.63	39.58	34.1	34.32	35.59	100	60.14	38.83	40.26	43.46	32.7	63.47	61.39	30.98	41.85	28.3	42.8
AtNLP7	37.27	40.28	34.99	42.75	44.19	60.21	100	38.53	40.37	42.42	35.52	48.51	48.73	32.09	37.58	29.76	41.44
AtNLP8	37.5	35.98	32.82	39.38	38.46	34.4	38.53	100	64.69	38.77	29.88	34.97	35.68	30.05	48.67	31.32	49.38
AtNLP9	37.08	36.54	32.94	38.53	39.67	40.26	40.37	65.14	100	38.49	35.43	34.65	37.5	32.68	51.94	35.78	48.9
GmNLP1	43.15	42.65	40.06	48.89	47.03	35.53	37.44	38.92	38.46	100	37.57	38.82	40.46	38.82	34.72	37.2	38.88
GmNLP2	40.62	40.53	34.89	39.19	39.14	32.7	35.36	30.17	36.23	37.2	100	38.92	38.91	50.38	31.72	49.56	31.72
GmNLP3	33.77	35.39	34.75	39.44	41.88	63.47	48.31	34.97	34.65	39.59	38.92	100	66.27	38.06	35.28	35.95	32.81
GmNLP4	34.62	36.83	34.46	39.84	42.34	61.39	48.87	35.68	37.5	40.46	38.58	66.85	100	68.48	36.95	69.23	37.91
GmNLP5	38.57	44.82	31.28	34.9	35.37	30.34	31.93	30.25	32.13	37.33	48.23	35.64	36.55	100	30.39	68.84	31.97
GmNLP6	37.11	35.75	34.99	40.47	41.32	41.85	37.98	48.84	51.91	39.63	31.4	35.28	36.95	27.58	100	36.44	64.76
GmNLP7	39.07	38.95	31.26	33.16	34.95	28.3	29.81	31.63	33.62	36.29	46.67	36.08	36.15	69.11	37.47	100	39.5
GmNLP8	36.35	36.56	34.29	40.36	39.16	42.8	41.44	49.69	49.16	38.79	32.07	32.81	37.91	31.24	64.36	39.5	100

FIGURE 4.22: Similarities in sequence between the NLP proteins of *G.max* and *A. thaliana*

4.8 Recognition of *GmNLP* Promoter 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). Three categories of cis-regulatory elements were created in the promoter regions of both *AtNLPs* and *GmNLPs* in order to classify the identified cis-regulatory elements into three groups: phytohormone (PR), plant growth and development (PGD) and stress (SR) (Table 4.20).

In comparison, *AtNLPs* have more regulatory components than *GmNLPs*. *AtNLPs* had more cis-elements (87) that reacted to phytohormones than *GmNLPs*, However, a total of 46 and 45 *AtNLPs* cis-elements, respectively, reacted to PGD and SR. All *AtNLPs* had more PR cis-elements than SR and PGD, with the exception of *AtNLP7*, which had more PGD responsive cis-elements. Similarly, *GmNLPs* have more PGD-responsive cis-elements in *GmNLP3* and *GmNLP5*, more stress-responsive cis-elements in *GmNLP4*, and more cis elements in the PR group in the remaining *GmNLPs*. The overall plant growth and development, stress responsive, and phytohormone responsive cis-elements discovered in *GmNLPs* are 28, 16, and 65, respectively.

TABLE 4.20: Cis-regulatory elements found in the promoter regions of *AtNLPs* and *GmNLPs*

Gene	Plant Growth & Development								Stress Responsive							Phytohormone Responsive					
	Box 4	MRE	O2-site	Circadian	GCN4_Motif	MSA-Like	WUN-Motif	ARE	MBS	TC-rich	LTR	GC-motif	CGTCA-Motif	TGACG-Motif	GARE-Motif	P-Box	TATC-Box	ABRE	ERE	TGA-element	TCA-element
AtNLP1	0	1	2	1	0	0	0	3	2	1	0	0	1	1	0	0	0	3	0	1	1
AtNLP2	1	0	1	0	0	0	0	5	1	1	0	0	6	6	0	1	1	2	0	1	0
AtNLP3	3	1	1	0	0	0	0	2	0	1	0	0	0	0	0	1	1	5	0	1	1
AtNLP4	2	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	3	0	1
AtNLP5	5	1	0	0	0	0	0	3	1	0	0	0	1	1	1	0	0	2	0	0	1
AtNLP6	1	1	0	0	0	1	0	2	0	1	1	0	2	2	0	2	0	1	0	1	1
AtNLP7	6	2	0	1	0	0	1	4	0	3	1	0	0	0	0	1	2	2	1	2	0
AtNLP8	2	0	0	0	0	0	1	3	1	2	1	0	3	3	0	0	0	1	0	0	1
AtNLP9	4	1	0	0	1	1	0	2	0	0	0	0	2	2	0	1	0	5	2	1	4
GmNLP1	2	0	0	0	0	0	2	1	0	0	1	0	1	1	0	0	0	3	3	1	1
GmNLP2	6	0	0	0	0	0	1	2	0	1	0	0	1	1	0	0	0	0	4	0	1
GmNLP3	11	0	0	0	0	0	0	1	1	0	0	0	0	3	3	1	1	0	1	1	0
GmNLP4	0	0	1	0	0	0	0	2	1	2	0	0	1	1	2	0	0	0	0	0	1
GmNLP5	6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1
GmNLP6	2	0	0	0	0	0	0	4	1	0	0	0	1	1	0	0	0	1	2	0	1
GmNLP7	6	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	4	5	2	2
GmNLP8	3	0	0	0	0	0	1	1	0	2	0	0	1	1	0	1	1	1	3	0	0

Chapter 5

Discussion

In the plant body, Transcription factors are essential for regulating the expression of genes as well as the stimulation of plant growth, the cell cycle, signaling, and the response to stress. Higher plants have been revealed to include more than 60 transcription factor families [51]. Numerous transcription factor families related to hormone control, seed development, and environmental adaptability in soybean plants have been investigated such as *LACS* (long-chain acyl-CoA synthetase Gene Family), *BTB* gene family, *MGT* (Magnesium Transporter gene family) and other transcription factor families.

The Nin like Proteins (*NLPs*) are a vital family of transcription factors exclusive to plants. Earlier studies have revealed that *NLPs* play a crucial role in the processes of absorbing, assimilating, and transporting nitrogen, which is controlled by nitrogen availability. It is widely recognized that the availability of nitrogen does not trigger the expression of *NLPs*; nevertheless, *NLPs* guide the first response to N via the nuclear-retention mechanism to localize *NLPs*. Consequently, the availability of nitrogen causes an increase in the accumulation of *NLPs*, which ultimately leads to an increase in the expression of genes that are sensitive to nitrogen, so enabling plants to ingest greater quantities of nitrogen [5]. *NLPs* are crucial TFs in the signaling of nitrates. Reports indicate that nitrate phosphorylates *NLPs* via triggering nitrate-coupled CPK signaling and nitrate CPK

(Ca2+-sensor protein kinase)-*NLP* signaling. These pathways combine transcription, metabolism, transport, and plant development systems and are critical for nitrate signaling [52].

The whole genome sequencing of *G.max* was done in 2010 [53] that provided foundational ground for this study to be conducted and the *GmNLP* genes to be studied both structurally and functionally. While genome-wide study pipelines cannot accurately confirm the existing molecular mechanisms or structures present within the cell, such studies provide a foundational background for future, more detailed in-vitro studies and may provide insight into important structural and functional characteristics of a specific gene family. These studies are also useful in identifying and analyzing a specific gene family at a low cost, as well as accurately predicting the associations of gene families with certain concepts that may be important for future academic research and industrial and commercial applications. For example, a previous study conducted in 2015, which examined the relationships between blooming time, maturation date, and height of plants in *G.max* over the whole genome, concluded that the chromosomal regions and loci identified during the study may serve as promising targets for future molecular mechanism studies [54].

However, no previous reports have provided thorough information regarding the *NLPs* TF family in soybean. In this study, we examined the soybean genome annotation data utilizing bioinformatics analysis and other relevant technologies in order to identify members of the *NLP* transcription factor family. The current work used multiple genome-wide computational techniques to identify 8 *NLPs* from the *Glycine max* genome. These Gm*NLPs* were matched to *Arabidopsis thaliana's NLP* gene family. Because *in-silico* research relies on comparison algorithms, the similarities discovered during this comparison can be used to predict a gene's function. Previous investigations have demonstrated that the conserved domains for the family of *NLP* genes are PB1 and RWP-RK, and the existence of these domains alludes to an evolutionary link between *AtNLPs* and *GmNLPs*.

Our research entails a thorough examination of soybean *NLP* members, including physiochemical properties, gene structure, evolutionary links, PPI, conserved motifs, and chromosomal locations and the identification of cis-regulatory elements in

promoter regions, among other things. Slightly lower pI values and longer protein sequences are notable characteristics of *NLP* genes in a variety of species, which is consistent with our findings. Based on these findings, transcription factors (TFs) appear to be more active in acidic conditions. The two important domains of the GmNLP proteins found in this study are PB1 and RWP-RK. The RWP-RK domain has an amphipathic leucine zipper and a helix-turn-helix motif, both capable of binding DNA. Further studies have shown that its role is independent of nitrate signaling. The signaling mechanism and transcriptional activation domain are both impacted by the N-terminal sections of *NLPs*. A combined five-stranded sheet and two helices make up the PB1 domain, which is thought to be essential for protein binding. The PB1 domain is thought to act as a trigger, possibly controlling when the binding process starts or stops [55].

All *NLPs* from *A. thaliana* and *Glycine max* plants have negative GRAVY values, indicating that *NLPs* are hydrophilic proteins. The isoelectric point of the *NLP* protein was low in both species, with values below 7. This means that NLP transcription factors are turned on when the pH is low. Except for *GmNLP5* and GmNLP7, soybean NLP genes have a conservative gene structure with 4 to 5 exons. All *NLP* proteins were projected into the nucleus. Because the nucleus regulates cellular activity, this might explain why NLP improves plants' capacity to withstand low nitrogen under nitrogen-deficient circumstances. *GmNLP* protein-protein interactions reveal the presence of two GRAS domain proteins, *GMNSP1* and *GMNSP*2, which are required for Nod-factor signaling and nodule formation. We discovered several cis-acting elements in the promoters of the *GmNLP* gene family members that are linked to hormone and stress responsive elements. This suggests that these cis-acting elements may control *GmNLP* genes throughout soybean growth and in reaction to phytohormones and stress.

In the case of *Glycine max*, preliminary information about *NLPs* may be useful for various genetic engineering tools and techniques that can be implemented to improve crop yield and production, as well as Nitrogen-Use-Efficiency (NUE), which can lead to promising results for both the environment and the agriculture sector.

Chapter 6

Conclusion and Future Work

In short, this study looked at all of the *NLP* genes in the genome of *Glycine max* (soybean). This study identified eight members of the NLP gene family in the soybean genome, which are distributed across eight different chromosomes. All *GmNLP* proteins showed acidic pI values and negative GRAVY scores, indicating their hydrophilic nature and potential role in transcriptional regulation and stress responses. Phylogenetic analysis indicated that the members of *Glycine max NLPs* exhibited the closest evolutionary relationship with those of *Populus trichocarpa*, and *Solanum lycopersicum.* Each participant has a typical conservative domain. The promoter region of RWP-RK and PB1 includes a number of cis-acting elements related to hormone and stress response. Determining the molecular mechanism of nitrogen utilization and enhancing NUE in soybeans are thought to need a thorough grasp of the functions of *GmNLP* under varying nutrition situations. Significantly, Protein-protein interaction analysis revealed that *GmNLPs* strongly interact with nitrogen-responsive genes such as NSP and ARF families, which are essential for root and nodulation development and may indicate more general functions in symbiosis and plant growth.

The complete characterization of the *NLP* gene family in *Glycine max* has allowed us to learn valuable new information about its structural and functional characteristics in nitrogen signaling and use. This study identified eight *GmNLP* genes, as well as their conserved domains (PB1 and RWP-RK), evolutionary relationships,

and potential regulatory mechanisms. The fact that their promoters contain cis-elements that are sensitive to both stress and hormones further supports their role in NUE, suggesting a role in phytohormone signalling and abiotic stress adaption. Protein-protein interactions were analyzed and shown to be linked to genes involved in nodulation and nitrogen absorption, emphasizing their role in soybean symbiotic nitrogen fixation. The results of this study offer fresh perspectives on the development, diversity, and functional significance of the *NLP* gene family in soybean in addition to providing new insights. To improve the efficiency with which nitrogen is used and to promote environmentally friendly farming methods, the genes that have been identified as *GmNLP* have the potential to serve as excellent targets for future genetic and functional investigations. Exploring the functional diversity and regulatory roles of *NLP* genes could provide valuable insights for biotechnological approaches aimed at developing high-yielding soybean cultivars with improved NUE. Such advancements would be critical for reducing dependency on synthetic nitrogen fertilizers, therefore promoting environmentally friendly farming techniques and sustainable agriculture.

6.1 Future Recommendation

Future studies should focus on experimental validation of the identified *GmNLP* genes to confirm their predicted roles. Advanced techniques such as CRISPR/-Cas9 genome editing or transgenic methods can be applied to enhance nitrogen use efficiency in soybean, contributing to more sustainable and productive cultivation. Moreover, exploring their expression under multiple environmental stresses could provide broader insights into their regulatory functions. Collectively, these efforts can guide the development of soybean varieties with improved yield, resilience, and environmental sustainability, making a significant contribution to future agricultural practices.

Bibliography

[1] M. Liu et al., "Genome-Wide Identification and Characterization of NODULE-INCEPTION-Like Protein (NLP) Family Genes in Brassica napus," International Journal of Molecular Sciences, vol. 19, no. 8, p. 2270, 2018, doi: https://doi.org/10.3390/ijms19082270.

[2] X. Lian et al., "Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray," Plant Molecular Biology, vol. 60, pp. 617-631, 2006, doi: https://doi.org/10.1007/s11103-005-5441-7.

[3] K. A. Congreves, O. Otchere, D. Ferland, S. Farzadfar, S. Williams, and M. M. Arcand, "Nitrogen use efficiency definitions of today and tomorrow," Frontiers in Plant Science, vol. 12, p. 637108, 2021, doi: https://doi.org/10.3389/fpls.2021.637108.

[4] P. Govindasamy et al., "Nitrogen use efficiency—a key to enhance crop productivity under a changing climate," (in English), Frontiers in Plant Science, Review vol. 14, 2023-April-18 2023, doi: https://doi.org/10.3389/fpls.2023.1121073.

[5] S. U. Jan et al., "Genome-wide study and expression analysis of Nodule-inception-like protein (NLP) gene family in Physcomitrella patens reveal its role in nitrogen response," Pakistan Journal of Botany, vol. 55, 02/15 2023, doi: http://.www.doi.org/10.30848/PJB2023-1(41).

[6] L. Schauser, W. Wieloch, and J. Stougaard, "Evolution of NIN-like proteins in *Arabidopsis,* rice, and Lotus japonicus," Journal of Molecular Evolution, vol. 60, pp. 229-237, 2005, doi: https://doi.org/10.1007/s00239-004-0144-2.

[7] C. Chardin, T. Girin, F. Roudier, C. Meyer, and A. Krapp, "The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development," Journal of Experimental Botany, vol. 65, no. 19, pp. 5577-5587, 2014, doi: 10.1093/jxb/eru261.

[8] L. Castaings et al., "The nodule inception-like protein 7 modulates nitrate sensing and metabolism in *Arabidopsis,*" The Plant Journal : for cell and molecular biology, vol. 57, pp. 426-35, 10/01 2008, doi: https://doi.org/10.1111/j.1365-313X.2008.03695.x.

[9] P. Guan et al., "Interacting TCP and NLP transcription factors control plant responses to nitrate availability," Proceedings of the National Academy of Sciences, vol. 114, p. 201615676, 02/15 2017, doi: 10.1073/pnas.1615676114.

[10] C. Marchive et al., "Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants," Nature Communications, vol. 4, no. 1, p. 1713, 2013/04/16 2013, doi: https://doi.org/10.1038/ncomms2650.

[11] 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," Physiology and Molecular Biology of Plants, vol. 30, no. 7, pp. 1209-1223, 2024/07/01 2024, doi: http://.www.doi.10.1007/s12298-024-01485-y.

[12] S. Asad, M. Wahid, F. Shaheen, A. Raza, and M. Farooq, "Soybean production in Pakistan: experiences, challenges and prospects," International Journal of Agriculture and Biology, 10/10 2020, doi: 10.17957/IJAB/15.1526.

[13] R. Mishra, M. Tripathi, R. S. Sikarwar, Y. Singh, and N. Tripathi, "Soybean (Glycine max L. Merrill): A Multipurpose Legume Shaping Our World," Plant Cell Biotechnology and Molecular Biology, vol. 25, no. 3-4, pp. 17-37, 04/15 2024, doi: 10.56557/PCBMB/2024/v25i3-48643.

[14] R. F. Wilson, "Soybean: Market Driven Research Needs," in Genetics and Genomics of Soybean, G. Stacey Ed. New York, NY: Springer New York, 2008, pp. 3-15.

[15] G. Dukariya, S. Shah, G. Singh, and A. Kumar, "Soybean and Its Products: Nutritional and Health Benefits," Journal of Nutritional Science and Healthy Die, vol. 1, no. 2, pp. 22-29, 07/07 2020.

[16] C. Fang and F. Kong, "Soybean," Current Biology, vol. 32, no. 17, pp. R902-R904, 2022, doi: http://.www.doi.10.1016/j.cub.2022.06.054.

[17] D. K. Agarwal, S. D. Billore, A. N. Sharma, B. U. Dupare, and S. K. Srivastava, "Soybean: Introduction, Improvement, and Utilization in India—Problems and Prospects," Agricultural Research, vol. 2, no. 4, pp. 293-300, 2013/12/01 2013, doi: https://doi.org/10.1007/s40003-013-0088-0.

[18] R. M. Nair et al., "Global Status of Vegetable Soybean," Plants, vol. 12, no. 3, p. 609, 2023, doi: https://.www.doi.org/10.3390/plants12030609.

[19] R. Nova, A. Hoemardani, and M. Louisa, "Potential of herbal medicines in cancer therapy," The Indonesia Journal of Cancer Control, vol. 1, no. 1, pp. 32-42, 05/10 2021, doi: http://.www.doi.10.52830/inajcc.v1i1.44.

[20] R. Singh, "Botany and Cytogenetics of Soybean," in The Soybean Genome, 2017, pp. 11-40.

[21] S. Jan and A. fatima, "Approaches for sustainable production of soybean under current climate change condition," vol. 08, no. 1, pp. 27-31, 04/03 2023, doi: https://.doi.10.15406/mojbm.2023.08.00179.

[22] M. A. Hussain, A. Wahab, U. Ahmad, and Z. Arshad, "Current Scenario of Oilseeds in Pakistan," vol. 1, no. 1, pp. 29-36, 08/15 2024, doi: https://doi.org/10.62460/TBPS/2023.006.

[23] H. Khurshid and S. Jan, "Miracle Crop: The Present and Future of Soybean Production in Pakistan," MOJ Biology and Medicine, vol. 2, 10/28 2017, doi: 10.15406/mojbm.2017.02.00042.

[24] J. Medic, C. Atkinson, and C. R. Hurburgh, "Current Knowledge in Soybean Composition," Journal of the American Oil Chemists' Society, vol. 91, no. 3, pp. 363-384, 2014/03/01 2014, doi: https://doi.org/10.1007/s11746-013-2407-9.

[25] M. Messina, "Soy and Health Update: Evaluation of the Clinical and Epidemiologic Literature," Nutrients, vol. 8, no. 12, p. 754, 2016, doi: https://doi.org/10.3390/nu8120754.

[26] K.-I. Chen, M.-H. Erh, N.-W. Su, W.-H. Liu, C.-C. Chou, and K.-C. Cheng, "Soyfoods and soybean products: from traditional use to modern applications," Applied Microbiology and Biotechnology, vol. 96, no. 1, pp. 9-22, 2012/10/01 2012, doi: http://www.doi.10.1007/s00253-012-4330-7.

[27] J. H. Kang, Z. Dong, and S. H. Shin, "Benefits of Soybean in the Era of Precision Medicine: A Review of Clinical Evidence," (in eng), J Microbiol Biotechnol, vol. 33, no. 12, pp. 1552-1562, Dec 28 2023, doi: http://www.doi.10.4014/jmb.2308.08016.

[28] K.-I. Chen, M.-H. Erh, N.-W. Su, W.-H. Liu, C.-C. Chou, and k.-C. Cheng, "Soyfoods and soybean products: From traditional use to modern applications," Applied Microbiology and Biotechnology, vol. 96, pp. 9-22, 08/08 2012, doi: 10.1007/s00253-012-4330-7.

[29] M. Jayachandran and B. Xu, "An insight into the health benefits of fermented soy products," Food Chemistry, vol. 271, pp. 362-371, 2019/01/15/ 2019, doi: https://doi.org/10.1016/j.foodchem.2018.07.158.

[30] G. Dukariya, S. Shah, G. Singh, and A. Kumar, "Soybean and Its Products: Nutritional and Health Benefits," vol. 1, no. 2, pp. 22-29, 07/07 2020.

[31] J. Y. Ye, W. H. Tian, and C. W. Jin, "Nitrogen in plants: from nutrition to the modulation of abiotic stress adaptation," Stress Biology, vol. 2, no. 1, p. 4, 2022/01/07 2022, doi: https://doi.org/10.1007/s44154-021-00030-1.

[32] A. Fathi, "Role of nitrogen (N) in plant growth, photosynthesis pigments, and N use efficiency: A review," Agrisost, vol. 28, pp. 1-8, 10/04 2022, doi: http://.www.doi.10.5281/zenodo.7143588.

[33] J. D. Murray, C.-W. Liu, Y. Chen, and A. J. Miller, "Nitrogen sensing in legumes," Journal of Experimental Botany, vol. 68, no. 8, pp. 1919-1926, 2017, doi: https://doi.org/10.1093/jxb/erw405.

[34] Y.-Y. Wang, Y.-H. Cheng, K.-E. Chen, and Y.-F. Tsay, "Nitrate transport, signaling, and use efficiency," Annual Review of Plant Biology, vol. 69, no. 1, pp. 85-122, 2018, doi: https://doi.org/10.1146/annurev-arplant-042817-040056.

[35] W. Wang, B. Hu, A. Li, and C. Chu, "NRT1. 1s in plants: functions beyond nitrate transport," Journal of Experimental Botany, vol. 71, no. 15, pp. 4373-4379, 2020, doi: https://doi.org/10.1093/jxb/erz554.

[36] C. Xiao et al., "Nitrate transporter NRT1. 1 and anion channel SLAH3 form a functional unit to regulate nitrate-dependent alleviation of ammonium toxicity," Journal of Integrative Plant Biology, vol. 64, no. 4, pp. 942-957, 2022, doi: https://doi.org/10.1111/jipb.13239.

[37] 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, doi: https://doi.org/10.3390/f15071191.

[38] J. Qin et al., "Responses of nitrogen metabolism, photosynthetic parameter and growth to nitrogen fertilization in Cyclocarya paliurus," Forest Ecology and Management, vol. 502, p. 119715, 2021, doi: https://doi.org/10.1016/j.foreco.2021.119715.

[39] Q. Wang, C. Liu, D. Huang, Q. Dong, P. Li, and F. Ma, "High-efficient utilization and uptake of N contribute to higher NUE of 'Qinguan'apple under drought and N-deficient conditions compared with 'Honeycrisp'," Tree physiology, vol. 39, no. 11, pp. 1880-1895, 2019, doi: https://doi.org/10.1093/treephys/tpz093.

[40] N. Sandhu, M. Sethi, A. Kumar, D. Dang, J. Singh, and P. Chhuneja, "Biochemical and Genetic Approaches Improving Nitrogen Use Efficiency in Cereal Crops: A Review," (in English), Frontiers in Plant Science, Review vol. Volume 12 - 2021, no. 1, pp. 1-45, 2021-June-04 2021, doi: http://.www.doi.10.3389/fpls.2021.657629.

[41] S. Suthar, P. Bishnoi, S. Singh, P. K. Mutiyar, A. K. Nema, and N. S. Patil, "Nitrate contamination in groundwater of some rural areas of Rajasthan, India," Journal of Hazardous Materials, vol. 171, no. 1-3, pp. 189-199, 2009, doi: https://doi.org/10.1016/j.jhazmat.2009.05.111.

[42] A. J. Franzluebbers, J. Sawchik, and M. A. Taboada, "Agronomic and environmental impacts of pasture–crop rotations in temperate North and South America," Agriculture, Ecosystems & Environment, vol. 190, pp. 18-26, 2014, doi: https://doi.org/10.1016/j.agee.2013.09.017.

[43] D. L. Zuluaga, D. De Paola, M. Janni, P. L. Curci, and G. Sonnante, "Durum wheat miRNAs in response to nitrogen starvation at the grain filling stage," PLoS One, vol. 12, no. 8, p. e0183253, 2017, doi: https://doi.org/10.1371/journal.pone.0183253.

[44] G. Yuan et al., "Genome-wide identification and expression analysis of NIN-like Protein (NLP) genes reveals their potential roles in the response to nitrate signaling in watermelon," Horticultural Plant Journal, vol. 8, no. 5, pp. 602-614, 2022/09/01/ 2022, doi: https://doi.org/10.1016/j.hpj.2022.06.010.

[45] K.-H. Liu et al., "NIN-like protein 7 transcription factor is a plant nitrate sensor," Science, vol. 377, no. 6613, pp. 1419-1425, 2022, doi: https://doi.org/10.1126/science.add1104.

[46] M. Konishi, T. Okitsu, and S. Yanagisawa, "Nitrate-responsive NIN-like protein transcription factors perform unique and redundant roles in *Arabidopsis,*" Journal of Experimental Botany, vol. 72, no. 15, pp. 5735-5750, 2021, doi: http://.www.doi.org10.1093/jxb/erab246.

[47] M. Durand et al., "The Arabidopsis transcription factor NLP2 regulates early nitrate responses and integrates nitrate assimilation with energy and carbon skeleton supply," The plant cell, vol. 35, no. 5, pp. 1429-1454, 2023, doi: https://doi.org/10.1093/plcell/koad025.

[48] K.-h. Liu et al., "Discovery of nitrate–CPK–NLP signalling in central nutrient–growth networks," Nature, vol. 545, no. 7654, pp. 311-316, 2017/05/01 2017, doi: https://doi.org/10.1038/nature22077.

[49] Y. Ueda, M. Konishi, and S. Yanagisawa, "Molecular basis of the nitrogen response in plants," Soil Science and Plant Nutrition, vol. 63, no. 4, pp. 1-13, 08/01 2017, doi: https://.www.doi.10.1080/00380768.2017.1360128.

[50] B. Jagadhesan et al., "Genome wide analysis of NLP transcription factors reveals their role in nitrogen stress tolerance of rice," (in eng), Sci Rep, vol. 10, no. 1, p. 9368, Jun 10 2020, doi: http://.www.doi.org10.1038/s41598-020-66338-6.

[51] Y. Wu et al., "Identification and expression characteristics of NLP (NIN-like protein) gene family in pepper (Capsicum annuum L.)," Molecular Biology Reports, vol. 50, no. 8, pp. 6655-6668, 2023/08/01 2023, doi: http://doi.org/10.1007/s11033-023-08587-y.

[52] 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," (in eng), BMC Genomics, vol. 24, no. 1, p. 320, Jun 13 2023, doi: http://.www.doi.org10.1186/s12864-023-09418-x.

[53] J. Schmutz et al., "Genome sequence of the palaeopolyploid soybean," Nature, vol. 463, no. 7278, pp. 178-183, 2010/01/01 2010, doi: http://.www.doi.10.1038/nature08670.

[54] J. Zhang et al., "Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm," BMC Genomics, vol. 16, no. 1, p. 217, 2015/03/20 2015, doi: http://.www.doi.org10.1186/s12864-015-1441-4.

[55] 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.)," Current Plant Biology, vol. 42, p. 100463, 2025/06/01/ 2025, doi: https://doi.org/10.1016/j.cpb.2025.100463.