***Pak. J. Bot.*, 55(1): 113-128, 2023.** **DOI: http://dx.doi.org/10.30848/PJB2023-1(41)**

GENOME-WIDE STUDY AND EXPRESSION ANALYSIS OF *NODULE-INCEPTION-LIKE PROTEIN* (*NLP*) GENE FAMILY IN *PHYSCOMITRELLA PATENS* REVEALS ITS ROLE IN NITROGEN RESPONSE

SAMI ULLAH JAN[1*], MAHA REHMAN[2], TAHMINA NAZISH[3], SOHAIL AHMAD JAN[1], AYESHA LIAQAT[4], MAHMOUD MOUSTAFA[5, 6], ALVINA GUL[4], NOOR UL HUDA[4], SYEDA MARRIAM BAKHTIYAR[1], SARAH GUL[7,] FAREES UD DIN MUFTI[2] AND MUHAMMAD JAMIL[2]

[1]Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan
[2]Department of Biotechnology & Genetic Engineering, Kohat University of Science and Technology, Kohat 26000, KPK, Pakistan
[3]Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan
[4]Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Islamabad 44000, Pakistan
[5]Department of Biology, Faculty of Science, King Khalid University, Abha, Saudi Arabia
[6]Department of Botany and Microbiology, Faculty of Science, South Valley University, Qena, Egypt
[7]Department of Biological Sciences, Faculty of Basic and Applied Sciences, International Islamic University, Islamabad, Pakistan
[*]Corresponding author's email: samiullahjan@gmail.com

Abstract

NODULE-INCEPTION-LIKE Proteins (NLPs) are plant specific transcription factors that play a significant role in orchestrating nitrogen response. NLPs have been widely studied in vascular plants but they are not explicitly reported in non-vascular bryophytes till date. In the current study, *in silico* tools were employed for identification and characterization of NLPs in model bryophyte *Physcomitrella patens*. Furthermore, the expression profiles of *PpNLPs* were assessed under variable supply of nitrogen. A total of 6 *Physcomitrella patens NLP* genes (*PpNLPs*) were identified that shared resemblance in their physical and chemical attributes with *Arabidopsis thaliana NLPs* (*AtNLPs*). *PpNLP* genes possessed resemblances in their iso-electric point and hydropathicity values with those of *AtNLPs* while gene lengths, protein lengths, and molecular weights were found higher in *PpNLPs*. The online tools suggested that all *PpNLPs*, except *PpNLP6*, yield acidic hydrophilic proteins localized in the nucleus and share a significant degree of homology in their gene structures and protein motifs with *AtNLPs*. Phylogenetic analysis indicated that *PpNLPs* possess significant evolutionary linkage with *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays*. Protein-protein interaction analysis suggested that *PpNLPs* possess substantial coordination with nitrogen responsive genes like nitrate reductase. Expressions of all *PpNLPs* were up-regulated in the availability of nitrogen (5 and 10 mM) while no significant increment was observed in the absence (0 mM) of nitrogen. The expression levels increased with increasing time treatment of 0, 6, 12, 24, 48, and 72 hours. Results proposed that *NLPs* are responsive to as well as significantly regulated by nitrogen supply.

Key words: Nitrogen; Nodule-inception-like Protein; *Physcomitrella patens;* Transcription factor; Nitrogen use efficiency

Introduction

Nitrogen (N) is an essential macronutrient for plant growth and yield (Tegeder and Masclaux-Daubresse, 2018). Usable N are limited in soil therefore N fertilizers are supplemented in agriculture to achieve high crop yield (Li *et al*., 2018). However, plants absorb a fraction (30-40%) of applied N while more than half (60-70%) is lost in soil causing severe soil and water pollution (Garnett *et al*., 2009). Inefficient conversion and consumption of N fertilizer also induce emission of nitrous oxide hence elicit global warming (Fagodiya *et al*., 2017). Despite their potential threats to environment, global demands for N fertilizer in agriculture increases continuously. Approximately 112 million tons (Mt) of N fertilizer were applied worldwide in 2015 while it was recorded to be 118 Mt in 2019 (Anon., 2019). Such progressive increment in the demand for enormous fertilizer quantities elicits agricultural cost as well. Therefore, enhancing the plant's ability to use N efficiently can elevate crop yield with reduced fertilizers input, agricultural costs, and environmental pollution (Alfatih *et al*., 2020). The term NUE (N use efficiency) is referred to jointly delineate the processes of N-uptake efficiency (NUpE) and -utilization efficiency (NUtE) in plants. NUE has been precisely defined as the amount of crop biomass or grain yield achieved at per unit application of N (Moll *et al*., 1982). Crop NUE improvement is widely recognized as an economic, effective, and desirable way of reducing N-associated agricultural and environmental problems. It is estimated that increasing the crop's NUE by merely 1% can significantly enhance crop yield and possibly save up to 1.1 billion US dollars a year (Kant *et al*., 2011). However, the comprehensive molecular mechanisms regulating NUE are yet to be understood.

Plants are evolved with effective and highly coordinated molecular mechanisms of N acquisition, assimilation, transport, and metabolism, governed by several transcription factors (TFs) and gene families (Feng *et al*., 2020). Plants absorb predominant inorganic nitrate (NO_3^-) from soil and transport them with the help of nitrate transporters like *NRT1* and *NRT2* (Orsel *et al*., 2002) across the channels including *CLC*: chloride channel (Zifarelli & Pusch, 2010) and *SLAH*: slow anion channel associated homologues (Qiu *et al*., 2016) into the cell. The absorbed inorganic nitrate is then reduced to ammonium (NH_4^+) by nitrate reductases (*NIA1*, *NIA2*) (Olas & Wahl, 2019) and nitrite reductase (*NiR*) (Takahashi *et al*., 2001). Ammonium is further assimilated into organic amino acids like glutamate and glutamine with the help of *GOGAT*: glutamate synthase (Forde & Lea, 2007), and *GS*: glutamine synthetase (Unno *et al*., 2006), respectively. These assimilated amino acids serve as N donors in biosynthesis of plant biomolecules including nucleic acids, essential amino acids, and chlorophyll (Masclaux-Daubresse *et al*., 2010). Moreover, both the absorbed nitrate as well as assimilated amino acids also serves as signaling molecules in regulation of associated TFs and cellular processes (Kan *et al*., 2015; Zhao *et al*., 2018). These deliberations thus render the significances of N and N-responsive TFs in plant structure, function, and overall NUE.

In a study of *Chlamydomonas reinhardtii* under N starved conditions, it was concluded that differentiation of vegetative cells into gametes is regulated by a protein named *MID*: *minus dominant protein* which switches-on or -off the minus or plus gametic differentiation program, respectively, in response to N signals (Ferris & Goodenough, 1997). This *MID* contains a conserved sequence RWPYRK after leucine zipper motif, which went unnoticed initially; however later investigations identified it as first member of a new TF family named RWP-RK gene family (Yin *et al*., 2020). RWP-RK is plant-specific gene family found in slime molds, green algae, and all vascular plants. Later on, the first *NIN*: nodule inception gene was identified in leguminous plant *Lotus japonicus* which also contains RWP-RK domain and regulates N-mediated symbiotic root-nodule formation (Schauser *et al*., 1999). Comprehensive studies classified RWP-RK gene family into two sub-families (i) *RKD*: RWP-RK domain containing gene family, and (ii) *NLP*: *RKD* with an additional domain at C-terminus named Phox and Bem1 (PB1) (Chardin *et al*., 2014). Members of *NLPs* were found having structural similarities with *NIN* genes – thus named as *NIN-Like Proteins* (Mu & Luo, 2019). *NIN* is legume-specific while *NLPs* are found in both non-leguminous and leguminous plants (Yokota & Hayashi, 2011). PB1 domain (PF00564) of *NLPs* arbitrates in protein-protein interaction, RWP-RK (PF02042) serve in DNA-binding, while, N-terminal region functions in transcriptional activation of genes (Liu *et al*., 2018). *NLPs* act as transcriptional activator in expression of nitrate regulated genes by binding to nitrate responsive *cis* element (*NRE*) in their promoter region (Konishi & Yanagisawa, 2013). *NLP* gene family has demonstrated as effective regulator of N-responsive genes therefore could potentially enhance NUE (Alfatih *et al*., 2020; Wu *et al*., 2020). So far, genome-wide studies have identified 6 *NLP* genes in rice (Jagadhesan *et al*., 2020), 9 in maize (Ge *et al*., 2018), 18 in wheat (Ge *et al*., 2018), 31 in *Brassica napus* and 9 in *Arabidopsis thaliana* (Liu *et al*., 2018). However, similar study of *NLPs* in non-vascular plants has not been reported till date.

The moss *Physcomitrella patens* is an established model non-vascular bryophyte for modern plants because it lies at the base of evolutionary lineage of today's plants and algae. The similarities and dissimilarities between the mosses and modern plants must be eminent from their genomes. As the *Arabidopsis thaliana* and *Physcomitrella patens* genomes have been sequenced, the genome wise comparison of *A. thaliana* with *P. patens* for finding orthologous and paralogous genes seems plausible in finding the evolutionary linkage between these two model organisms (Rensing *et al*., 2008). Since identification, the study of comprehensive structural and functional characterization of *NLP* genes for NUE improvement have focused on vascular plants, thus, a vivid gap of similar study in non-vascular plants is comprehended. In this study, initially, we used *in silico* tools to identify *NLP* genes in *P. patens* genome databases. Subsequently, the expression patterns of *NLP* genes in response to varying N concentrations were also assessed. Our study provides a valuable ground to understand the evolutionary relationship among *NLPs* of model vascular and non-vascular plants which facilitates *in vivo* functional characterization of *PpNLPs* in future.

Materials and Methods

Physcomitrella patens growth conditions: The *Physcomitrella patens* growth conditions were optimized according to established protocol (Koduri *et al*., 2010). The gametophores of *P. patens* ecotype Gransden 2004 were axenically grown at 25 ± 1^{o}C in continuous light (intensity: 50 µmol m^{-2} s^{-2}) and sub-cultured for three weeks. Explants from pre-cultures were allowed to grow for a week followed by treating with variable supply of N on liquid BCDA medium (Table S1). The KNO_3 was used as sole N source in treating *P. patens* with N-deficient (0 mM), -limiting (5 mM) and -sufficient (10mM) conditions provided in BCDA medium. The grown *P. patens* were treated for 0, 6, 12, 24, 48, and 72 hours. The rhizoid, stem and phylloid were harvested and stored at -80^{o}C.

Table S1-A. BCDA medium composition.

Reagent	Quantity (for 1 L)	Final concentration
Solution B	10 ml	1 mM MgSO4
Solution C	10 ml	1.84 mM KH2PO4
Solution D	10 ml	10 mM KNO3
CaCl2	111 mg	1 mM
FeSO4.7H2O	12.5 mg	45 µM
Agar	7.5 g	0.75% (w/v)
Glucose	5 g	0.5% (w/v)
Hoagland's A-Z trace	1 ml	Trace element solution

Table S1-B. Composition of Hoagland's trace elements.

Reagent	Quantity (for 1 L)	Final concentration
Al2(SO4)3.K2SO4.24H2O	55 mg	0.006% (w/v)
CoCl2.6H2O	55 mg	0.006% (w/v)
CuSO4.5H2O	55 mg	0.006% (w/v)
H3BO3	614 mg	0.061% (w/v)
KBr	28 mg	0.003% (w/v)
KI	28 mg	0.003% (w/v)
LiCl	28 mg	0.003% (w/v)
MnCl2.4H2O	389 mg	0.039% (w/v)
SnCl2.2H2O	28 mg	0.003% (w/v)
ZnSO4.7H2O	55 mg	0.006% (w/v)

Table S2. List of primers used in study of expression pattern of *PpNLPs* gene family.

Gene	Primer	Length (bp)	Sequence (5' to 3')	Amplicon size
PpNLP1	Forward	58	ATATCAAGGTTCACCAGAGTGG	258
	Reverse	58	TAGAATGGGTTTTCACATCGGA	
PpNLP2	Forward	59	CTCTTCGGAGCAGGAGTTAAAG	97
	Reverse	57	ATTAGGAAGACACAGTAGAGGC	
PpNLP3	Forward	58	CCAGTAGCGATAATTGCTATGC	239
	Reverse	58	CACGTTTTCCATCGAGCTTAAA	
PpNLP4	Forward	58	CGAGAACTATGTATTTGCCGTG	100
	Reverse	58	GTAGAATTGCACATTCGGAGTC	
PpNLP5	Forward	58	CTGTACAGGAACATGACGGAG	132
	Reverse	57	GCTACTGTAATACTGCACGTTC	
PpNLP6	Forward	57	ATGGAACTTTTGAGGTCGAATC	186
	Reverse	59	CTCCATCAGATCCATCAACACC	

RNA extraction and qPCR: The total cellular RNA from selected three parts; rhizoid, stem, and phylloid, was extracted with TRIzol method (Xiao *et al*., 2011). The cDNA synthesis from extracted RNA was carried out through oligo-dT primers and reverse transcription (TaKaRa) as per supplier's protocol. The quantified cDNA was subjected to reverse transcription qPCR (Step One Plus Real Time PCR System) using *P. patens Actin3* gene as internal reference. Gene specific primers (Table S2) were obtained from qPrimerDB version 1.2 (Bustin & Huggett, 2017).

Screening of genome and transcription factors databases: The full-length gene, protein, and coding sequences of all members of *Arabidopsis thaliana NLP* (*AtNLP*) gene family were retrieved from *Arabidopsis* genome database (TAIR: http://arabidopsis.org/). In total, three genomes and one plant-TF databases were screened for identification of putative *PpNLPs*. First, the *AtNLP*s protein sequences were used as BLAST-query in screening NCBI (https://www.ncbi.nlm.nih.gov/). Second, both versions of Phytozome (v12: https://phytozome. jgi.doe.gov/pz/portal.html, and v13: https://phytozome-next.jgi.doe.gov/) were screened using accession numbers of RWP-RK (PF00564) and PB1 (PF02042) domains as keywords (Ge *et al*., 2018). Last, sequences of all members enlisted under RWP-RK in plant TF database (iTAK: http://itak.feilab.net/cgi-bin/itak/index.cgi) were downloaded. All the sequences were aligned to eliminate redundant as well as alternative spliced variants.

Physicochemical properties and conserved domains identification in *PpNLPs*: Among the retrieved sequences, potential *PpNLPs* were selected on the basis of conserved domains. Genes containing both RWP-RK and PB1 domains were selected. The physical as well as chemical properties including protein molecular weight (MW), hydropathicity (GRAVY) and theoretical isoelectric point (pI) of selected *PpNLPs* were examined online on Protparam Expasy (https://web.expasy.org/ protparam/) (Gasteiger *et al*., 2003) while subcellular localizations were predicted using CELLO (http://cello.life.nctu.edu.tw/) (Orioli & Vihinen, 2019).

Phylogenetics of *PpNLPs*: Sequences of finally selected *PpNLPs* protein sequences were aligned along with *NLP* gene families of *Arabidopsis thaliana*, *Oryza sativa* (Jagadhesan *et al*., 2020), and *Zea mays* (Ge *et al*., 2018) using MEGA-X v10.1.8 software followed by constructing a rooted phylogenetic tree with neighbor-joining (NJ) method, 1000 bootstrap replicates, and default parameters. The online Interactive Tree of Life v5 (iTOL: https://itol.embl.de/) was used for visualization of rooted phylogenetic tree.

Gene structure and motif composition in *PpNLP* gene family: The coding and full length gene sequences of *PpNLPs* were used to examine gene structural components using GSDS online server (Gene Structure Display Server: http://gsds.cbi.pku.edu.cn/) (Hu *et al*., 2015). The introns, exons, and un-translated regions (UTRs) were identified. Furthermore, occurrence of consensus motifs was elicited on MEME v5.1.1online tool (Multiple Em for Motif Elicitation http://meme-suite.org/tools/meme) with default parameters using 15 consensus motifs threshold (Bailey *et al*., 2015).

Putative cis-regulatory elements identification in *PpNLPs* homologues: Gene regulatory elements in promoter regions of *PpNLPs* were identified using upstream promoter region of *PpNLPs* (2000 bps) retrieved from web-based database Plant Ensembl (http://www. plants.ensembl.org/). Promoter regions were investigated for *cis*-regulatory elements online (plant CARE: http://bioinformatics.psb.ugent.be/webtools/ plantcare/ html/) (Verma *et al*., 2017).

Chromosomal locations of *PpNLPs*: Localization of *PpNLPs* genes on chromosomes of *Physcomitrella patens* were examined through genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/). Distribution and location of *PpNLP* genes were plotted using MapChart2 (https://mapchart.net/).

Protein-protein interaction of *PpNLPs*: The *PpNLPs* protein sequences were analyzed on SMAR (http://smart.embl-heidelberg.de/). The cellular proteins interacting with *PpNLPs* were predicted in STRING (https://www.expasy.org/resources/string) and compared with interacting proteins of *AtNLPs* (Szklarczyk *et al*., 2019; Szklarczyk *et al*., 2017).

Table S3. Conserved domains of *NLP* gene families in *Arabidopsis thaliana* and *Physcomitrella patens.*

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
Physcomitrella patens	*PpNLP1*	specific	1054	1133	1.85E-32	120.891	cd06407	PB1_NLP
		specific	705	753	8.76E-23	92.1628	pfam02042	RWP-RK
	PpNLP2	specific	1132	1212	3.69E-39	139.766	cd06407	PB1_NLP
		specific	774	822	1.28E-23	94.474	pfam02042	RWP-RK
	PpNLP3	superfamily	1065	1144	3.66E-29	111.261	cl02720	PB1 superfamily
		specific	718	766	1.31E-22	91.3924	pfam02042	RWP-RK
	PpNLP4	specific	1148	1228	2.11E-37	134.758	cd06407	PB1_NLP
		specific	782	830	1.99E-23	94.0888	pfam02042	RWP-RK
	PpNLP5	superfamily	1169	1247	2.39E-27	106.253	cl02720	PB1 superfamily
		specific	774	822	6.78E-23	92.548	pfam02042	RWP-RK
	PpNLP6	superfamily	1179	1257	3.19E-29	111.646	cl02720	PB1 superfamily
		specific	779	827	8.71E-23	92.1628	pfam02042	RWP-RK

Statistical analysis

The results were statistically validated with significance ($p<0.05$) and graphs were developed using GraphPad Prism 8.

Results

Genome wide identification and analysis of physcomitrella patens NLP homologues: In the present study, three genome databases (NCBI, Phytozome.v12, Phytozome.v13) and one plant TF database (iTAK) were screened to identify *NLPs* in *Physcomitrella patens* genome (Taxonomy ID: 3218) using *Arabidopsis thaliana NLPs* protein sequences as well as pfam accessions of RWP-RK (PF02042) and PB1 domain (PF00564) as queries. Initially, 62 sequences were obtained comprising 25 from NCBI, 24 from Phytozome, and 13 from iTAK. All the sequences and their information obtained from updated version of Phytozome (v13) were similar to those in v12 except their accession numbers. The spliced variants, repeated/redundant sequences, and short or incomplete fragments were excluded from retrieved sequences simultaneously validated through conserved domain identification. Finally, 6 *PpNLPs* were identified that contained both RWP-RK and PB1 domains (Table S3) and were labeled from 1 to 6 with respect to chromosome numbers. Accession numbers of same or redundant sequences found in selected databases are enlisted in (Table 1), while, the physical and chemical properties of *A. thaliana* and *P. patens NLP* gene families are summarized in (Table 2).

The gene lengths, protein lengths, and molecular weights (MW) of *PpNLPs* were found higher than *AtNLPs*, however, the pI and GRAVY values of both plants were close to each other. The average gene lengths of *AtNLPs* and *PpNLPs* were found 4141 and 6471 bp, respectively. Likewise, a significant difference was observed in protein lengths of *AtNLPs* and *PpNLPs* with average of 880 and 1218 amino acids, respectively. Average MW of *AtNLPs* was found 97357 Kilo Daltons (KDa) while *PpNLPs* had average 131511 KDa MW. All the *AtNLPs* (except *AtNLP3*) and *PpNLPs* (except *PpNLP6*) had pI values below 7 indicating them as acidic proteins while *AtNLP3* and *PpNLP6* with pI values 8.14 and 7.30, respectively, were suggested as basic proteins. The study of sub-cellular localization of both *A. thaliana* and *P. patens NLPs* proposed them to be localized in nucleus while all *NLPs* from both plants showed negative GRAVY values which showed *NLPs* as hydrophilic proteins.

Table 1. Accession numbers of Identified *NLPs* in *Physcomitrella patens* and their redundant accession numbers.

Given name	Phytozome accession number	Redundant sequences accession in databases			
		Phytozome.v12	Phytozome.v13	iTAK	NCBI
PpNLP1	Pp3c9_14600V3.1	Pp3c9_14600V3.1	Pp3c9_14600	Pp1s302_9V6	XP_024384005.1
PpNLP2	Pp3c12_2070V3.1	Pp3c12_2070V3.1	Pp3c12_2070	Pp1s128_79V6	XP_024391180.1
PpNLP3	Pp3c15_9180V3.1	Pp3c15_9180V3.1	Pp3c15_9180	Pp1s250_18V6	XP_024397374.1 XP_024397373.1
PpNLP4	Pp3c17_4370V3.1	Pp3c17_4375V3.1	Pp3c17_4375	Pp1s26_246V6	XP_024400585.1
		Pp3c17_4370V3.1	Pp3c17_4370		
PpNLP5	Pp3c19_2670V3.1	Pp3c19_2720V3.1	Pp3c19_2720	Pp1s109_79V6	XP_024404168.1
		Pp3c19_2670V3.1	Pp3c19_2670		XP_024356825.1 PNR33779.1
PpNLP6	Pp3c22_6370V3.1	Pp3c22_6370V3.1	Pp3c22_6370	Pp1s12_320V6	XP_024361132.1
		Pp3c22_6360V3.1	Pp3c22_6360	Pp1s12_321V6	XP_024361103.1

Table 2. Physical and chemical properties of *NLP* gene families of *Arabidopsis thaliana* and *Physcomitrella patens.*

Plant	Gene name	Chr	Position	Gene length (bp)	Protein length (aa)	Molecular weight	Iso-electric point	GRAVY	Localization
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	-0.420	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
Physcomitrella patens	*PpNLP1*	9	9756164 - 9763070	6907	1151	125929.48	5.55	-0.516	Nucleus
	PpNLP2	12	1717318 - 1723598	6281	1233	132885.98	5.66	-0.486	Nucleus
	PpNLP3	15	6095352 - 6101605	6254	1162	126229.88	5.51	-0.477	Nucleus
	PpNLP4	17	3527404 - 3533715	6068	1251	135591.7	5.51	-0.518	Nucleus
	PpNLP5	19	1514672 - 1521939	7268	1252	133420.05	6.53	-0.374	Nucleus
	PpNLP6	22	3740778 - 3746829	6052	1262	135010.09	7.30	-0.396	Nucleus

Sequence alignment and phylogenetic relationship of *PpNLPs* gene family: The percent similarities of *PpNLPs* and *AtNLPs* were matched to confirm the appropriate selection as well as singularity of each identified *PpNLP* gene used for further analysis (Table S4). All the *AtNLPs* and *PpNLPs* shared less than 78% similarity in their protein sequences which assured the uniqueness of each gene as well as evolutionary diversity among members of *PpNLP* gene family. The alignment output of *PpNLP* gene family along with *NLP* gene families of *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays* was used to construct a rooted neighbor-joining phylogenetic tree in MEGA-X v10.1.8 with default parameters and 1000 bootstrap replicates (Fig. 1). The phylogenetic evolutionary relationship among *NLP* gene families of selected four plants were clustered in three clades. The *NLP* gene family of non-vascular *P. patens* showed evolutionary divergence from other three vascular plants. The *AtNLP8*, *-9*, *OsNLP2*, *-5*, *ZmNLP2*, and *-9* were closest members in the clade of *PpNLP* gene family. This distribution of *NLP* gene families established substantial evolutionary divergence among vascular tracheophytes and non-vascular bryophytes.

Gene structure, consensus motifs and chromosomal distribution of *PpNLPs*: Structural components of *AtNLPs* and *PpNLPs* were analyzed using the gene and their coding sequences. Identification of introns, exons, and UTRs in genic region (Fig. 2) shows that *PpNLP2*, and *-4* contains 3 exons while remaining *PpNLPs* possess 4 exons in each gene. The number of exons range between 4 and 6 in *AtNLPs*, while, *AtNLP3* do not a 5'UTR. Up to 15 consensus motifs were figured out using MEME in *PpNLP* proteins (Fig. 3, Table S5) compared with *AtNLPs*. All the sequences contained significantly conserved motifs in both *A. thaliana* and *P. patens* proteins. All the *AtNLPs* and all *PpNLPs* contained all motifs except *AtNLP4*, *-8*, and *-9* that contain 14 motifs while *AtNLP3* has 11 motifs. Appropriate localization of genes upon chromosome (Fig. 4, Table S6) revealed that 6 *PpNLPs* are localized on different chromosomes (Chr. 9, 12, 15, 17, 19, 22).

Table S4. Sequence similarity among *NLP* proteins of *Arabidopsis thaliana* and *Physcomitrella patens* (similarity above 60% are marked with red font color).

	AtNLP1	*AtNLP2*	*AtNLP3*	*AtNLP4*	*AtNLP5*	*AtNLP6*	*AtNLP7*	*AtNLP8*	*AtNLP9*	*PpNLP1*	*PpNLP2*	*PpNLP3*	*PpNLP4*	*PpNLP5*	*PpNLP6*	
AtNLP1	100	69.08	44.13	41.35	45.04	39.63	37.27	37.5	37.08	37.22	37.1	37.28	38.66	36.91	38.69	*AtNLP1*
AtNLP2	68.66	100	45.43	41.08	43.24	39.58	40.28	35.98	36.54	38.49	37.88	38.22	37.83	37.93	38.22	*AtNLP2*
AtNLP3	44.13	45.88	100	43.94	41.25	34.1	34.99	33.86	32.94	33.68	34.11	33.51	32.56	32.83	31.28	*AtNLP3*
AtNLP4	46.39	40.8	39.17	100	69.33	34.32	42.75	40.52	38.53	41.3	42	43.31	42.9	45.35	44.35	*AtNLP4*
AtNLP5	41.26	39.31	40.58	70.57	100	35.59	44.19	39.53	39.67	40.92	42.45	40	43.67	43.67	42.34	*AtNLP5*
AtNLP6	35.88	35.22	31.53	35.58	36.3	100	60.21	35.08	40.26	40.69	43.2	41.18	44.04	42.47	40.6	*AtNLP6*
AtNLP7	36.54	40.28	33.99	42.39	44.19	60.14	100	39.61	40.37	41.05	42.21	43.58	39.15	43.92	42.39	*AtNLP7*
AtNLP8	37.5	35.98	33.64	41.26	39.53	39.93	39.61	100	66.12	40.77	40.26	43.94	42.05	41.51	42.15	*AtNLP8*
AtNLP9	37.08	36.54	33.65	39.26	39.67	40.26	40.37	65.67	100	39.59	41.85	38.87	45.72	41.16	44.63	*AtNLP9*
PpNLP1	38.02	38.16	33.93	41.81	41.05	40.69	39.97	40.77	43.62	100	53.39	74.23	51.29	47.31	54.77	*PpNLP1*
PpNLP2	38.31	37.88	33.85	40.18	43.13	43.35	43.33	40.55	41.55	52.74	100	42.61	75.14	53.08	45.04	*PpNLP2*
PpNLP3	38.25	38.03	32.7	40.36	42.46	41.34	41.01	43.94	39.87	74.23	45.05	100	45.27	47.91	47.67	*PpNLP3*
PpNLP4	36.65	38.37	33.25	43.33	44.12	43.72	35.54	42.05	44.68	47.41	74.74	44.72	100	44.09	54.92	*PpNLP4*
PpNLP5	36.62	36.95	32.83	45.81	43.67	42.08	44.47	41.35	41.22	42.81	53.18	43.75	53.58	100	77.11	*PpNLP5*
PpNLP6	37.98	37.67	32.12	44.11	42.18	40.33	42.94	42.15	44.6	47.59	53.53	44.19	53.66	76.8	100	*PpNLP6*
	AtNLP1	*AtNLP2*	*AtNLP3*	*AtNLP4*	*AtNLP5*	*AtNLP6*	*AtNLP7*	*AtNLP8*	*AtNLP9*	*PpNLP1*	*PpNLP2*	*PpNLP3*	*PpNLP4*	*PpNLP5*	*PpNLP6*	

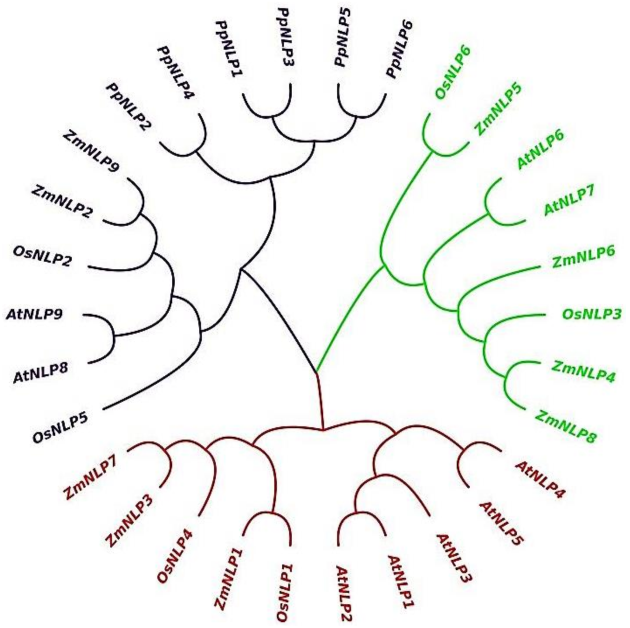


Fig. 1. Phylogenetic analysis of *PpNLPs* through neighbor joining method using MEGA-X.

Identification of cis-regulatory elements in promoter regions of *PpNLPs*: The recognition of *cis*-regulatory elements in upstream promoter regions (2000 bp) is a significant approach in proposing the gene function and regulation. Three categories of *cis*-regulatory elements in promoter regions of both *AtNLPs* and *PpNLPs* were devised to categorize the identified *cis*-elements in three groups including phytohormone (PR), stress (SR), and plant growth and development (PGD), shown in (Table 3). Comparatively, *AtNLPs* possess higher number of regulatory elements than *PpNLPs*. Highest total number of *cis*-elements (87) identified in *AtNLPs* were responsive to phytohormones, while, total numbers of *AtNLPs cis*-elements responsive to SR and PGD were 45 and 46, respectively (Fig. 5). All *AtNLPs* contained higher number of PR *cis*-elements except *AtNLP7* whose number of PGD responsive cis-elements were higher than SR and PR. Likewise, in *PpNLPs*, *PpNLP4* possess higher number of PGD responsive cis-elements while remaining *PpNLPs* have higher number of cis-elements in PR group. The total number of PGD, SR, and PR cis-elements identified in *PpNLPs* are 19, 21, and 35, respectively.

Protein-protein interaction of *PpNLPs*: The interacting *NLP* proteins networks were predicted online through STRING (Table S7). All the *PpNLP* proteins were suggested to interact with plethora of N related genes. Among them, 10 genes were commonly interacting with all *PpNLP* proteins. Most of these 10 genes are un-annotated predicted proteins, however, three *NIA*: nitrate reductases genes (PP1S58_252V6.1, PP1S58_249V6.1, and PP1S79_76V6.2) have been identified as significant putative N related genes interacting with *PpNLPs*. (Fig. 6) shows schematic model of all *PpNLPs* interacting with cellular proteins.

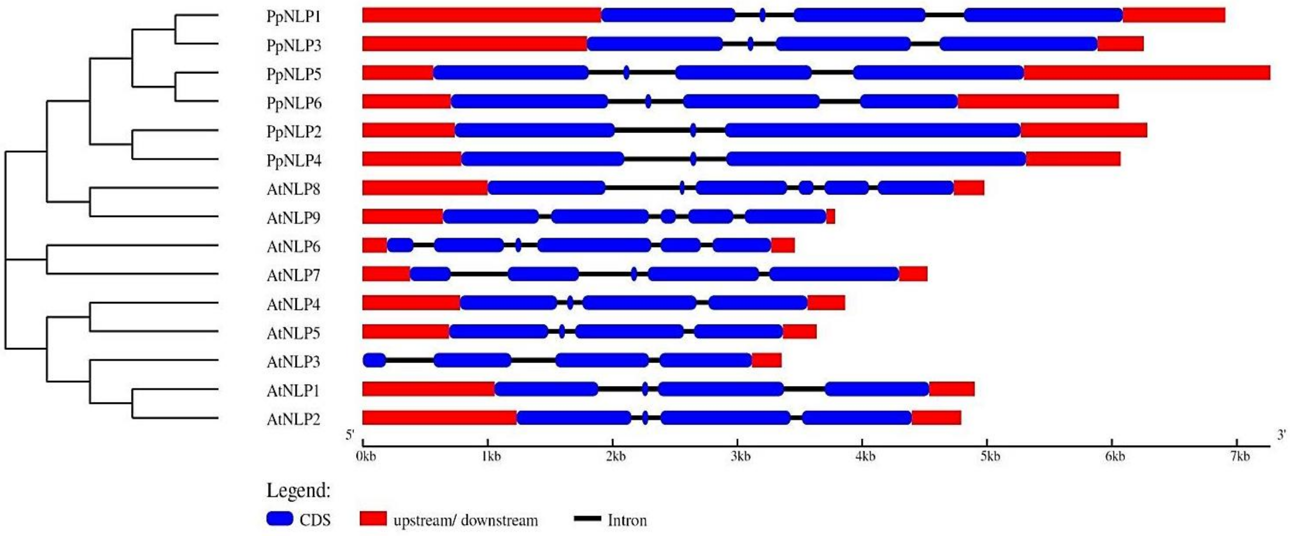
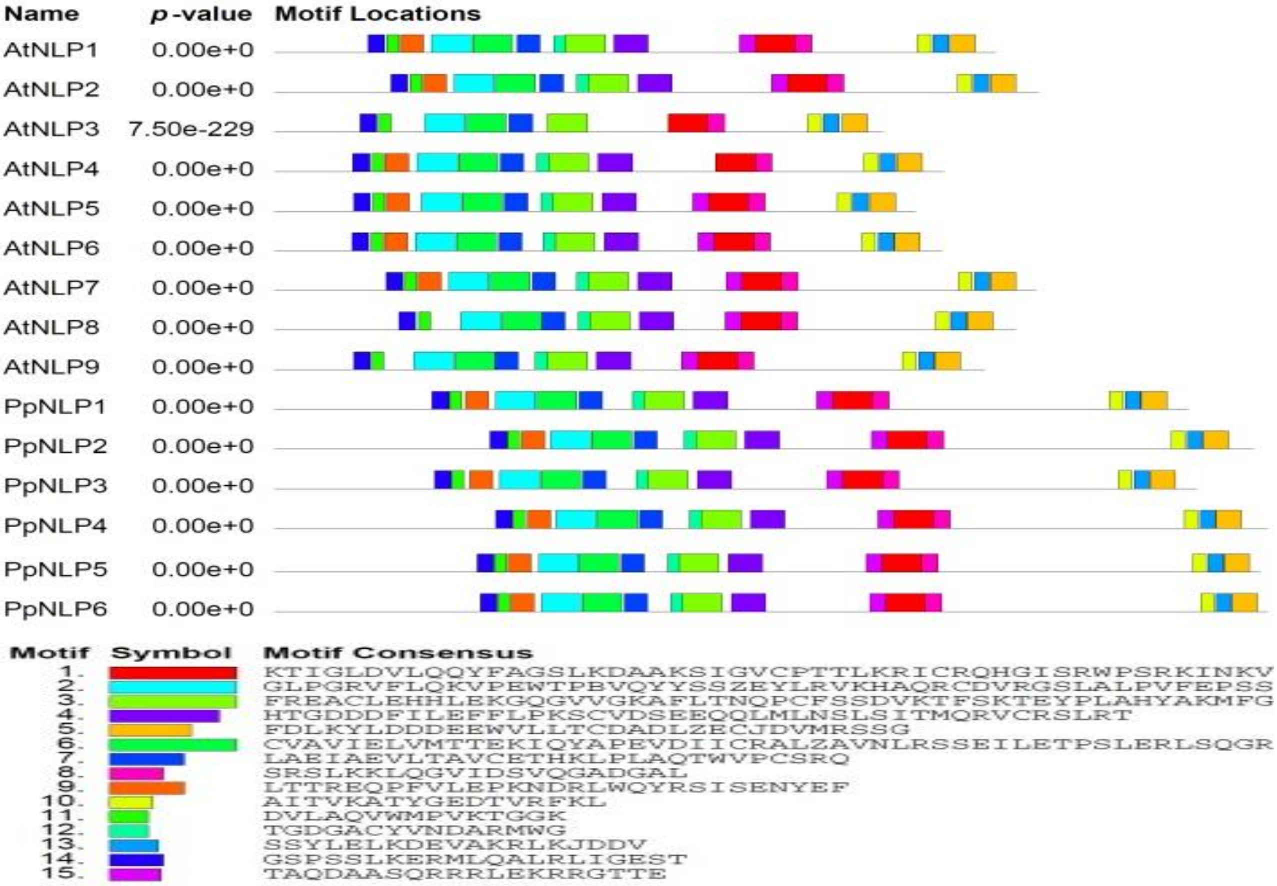


Fig. 2. Phylogenetic relationship and exon-intron structure of *AtNLPs* and *PpNLPs*.



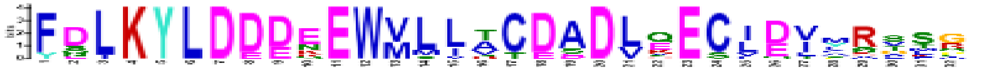
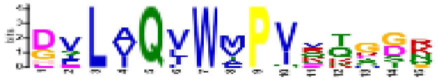
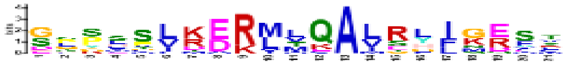
Name	*p*-value
AtNLP1	0.00e+0
AtNLP2	0.00e+0
AtNLP3	7.50e-229
AtNLP4	0.00e+0
AtNLP5	0.00e+0
AtNLP6	0.00e+0
AtNLP7	0.00e+0
AtNLP8	0.00e+0
AtNLP9	0.00e+0
PpNLP1	0.00e+0
PpNLP2	0.00e+0
PpNLP3	0.00e+0
PpNLP4	0.00e+0
PpNLP5	0.00e+0
PpNLP6	0.00e+0

Motif	Symbol	Motif Consensus
1.	🟥	KTIGLDVLQQYFAGSLKDAAKSIGVCPTTLKRICRQHGISRWPSRKINKV
2.	🟦	GLPGRVFLQKVPEWTPBVQYYSSZEYLRVKHAQRCDVRGSLALPVFEPSS
3.	🟩	FREACLEHHLEKGQGVVGKAFLTNQPCFSSDVKTFSKTEYPLAHYAKMFG
4.	🟪	HTGDDDFILEFFLPKSCVDSEEQQLMLNSLSITMQRVCRSLRT
5.	🟧	FDLKYLDDDEEWVLLTCDADLZECJDVMRSSG
6.	🟩	CVAVIELVMTTEKIQYAPEVDIICRALZAVNLRSSEILETPSLERLSQGR
7.	🟦	LAEIAEVLTAVCETHKLPLAQTWVPCSRQ
8.	🟪	SRSLKKLQGVIDSVQGADGAL
9.	🟧	LTTREQPFVLEPKNDRLWQYRSISENYEF
10.	🟨	AITVKATYGEDTVRFKL
11.	🟩	DVLAQVWMPVKTGGK
12.	🟩	TGDGACYVNDARMWG
13.	🟦	SSYLELKDEVAKRLKJDDV
14.	🟦	GSPSSLKERMLQALRLIGEST
15.	🟪	TAQDAASQRRRLEKRRGTTE

Fig. 3. Consensus Motifs in *AtNLPs* and *PpNLPs* gene families.

Expression pattern of *PpNLPs* gene family: The real time quantitative PCR was executed to assess the expression level of *PpNLP* in rhizoid, stem, and phylloids of *P. patens* while *Actin3* was taken as internal control. Three N treatments 0 (deficient), 5 (limiting), and 10 mM (sufficient) were provided for 0, 6, 12, 24, 48, 72 hours. Results indicated a significant differential pattern common in all *PpNLPs* in rhizoid, stem and phylloids (Figs. 7, 8). Expression of *PpNLPs* increased with increasing time of treatment from 6 to 72 hours under limiting (5 mM) and sufficient (10 mM) N supply, while no changes were observed in N deficient (0 mM) conditions. Thus, indicated that *PpNLPs* are highly regulated with N availability. The overall expression pattern showed significant up-regulation of all *PpNLPs* with immediate response due to expression increment within 0 to 6 hours in all three plant parts.

Table S5. List of consensus motifs and their sequences in *Arabidopsis thaliana* and *Physcomitrella patens NLP* gene family proteins.

Motif	Sequence	Logo
Motif 1	KTIGLDVLQQYFAGSLKDAAKSIGVCPTTLKRICRQHGISRWPSRKINKV	
Motif 2	GLPGRVFLQKVPEWTPBVQYYSSZEYLRVKHAQRCDVRGSLALPVFEPSS	
Motif 3	FREACLEHHLEKGQGVVGKAFLTNQPCFSSDVKTFSKTEYPLAHYAKMFG	
Motif 4	HTGDDDFILEFFLPKSCVDSEEQQLMLNSLSITMQRVCRSLRT	
Motif 5	FDLKYLDDDEEWVLLTCDADLZECJDVMRSSG	
Motif 6	CVAVIELVMTTEKIQYAPEVDIICRALZAVNLRSSEILETPSLERLSQGR	
Motif 7	LAEIAEVLTAVCETHKLPLAQTWVPCSRQ	
Motif 8	SRSLKKLQGVIDSVQGADGAL	
Motif 9	LTTREQPFVLEPKNDRLWQYRSISENYEF	
Motif 10	AITVKATYGEDTVRFKL	
Motif 11	DVLAQVWMPVKTGGK	
Motif 12	TGDGACYVNDARMWG	
Motif 13	SSYLELKDEVAKRLKJDDV	
Motif 14	GSPSSLKERMLQALRLIGEST	
Motif 15	TAQDAASQRRRLEKRRGTTE	

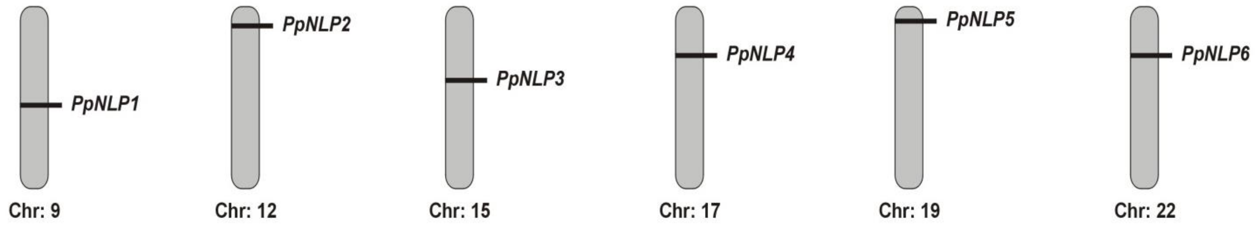


Fig. 4. Chromosomal distribution of *NLPs* genes in *Physcomitrella patens* genome.

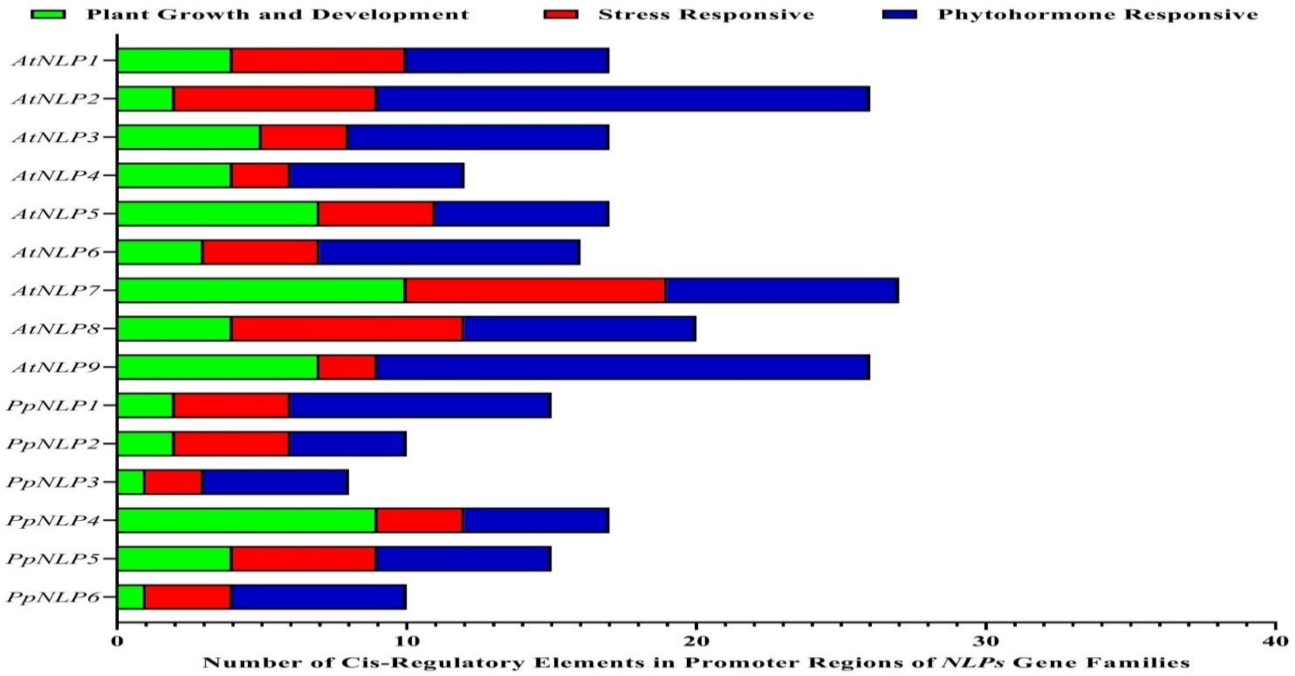


Fig. 5. Category wise presentation of total number of cis-elements in promoter regions of *NLP* gene families of *A. thaliana* and *P. patens*.

Discussion

Plant Transcription factors play a key role in plant growth and development in both biotic and abiotic stresses (Shah *et al*., 2016; Khurshid *et al*., 2018; Shinwari *et al*., 2020; Liping *et al*., 2021; Jan *et al*., 2022). The NODULE-INCEPTION-Like Proteins (NLPs) constitute an important group of plant specific transcription factors (Liu *et al*., 2018). Former studies have demonstrated an established significant role of NLPs in N uptake, assimilation, and transport regulated by N availability (Alfatih *et al*., 2020; Wu *et al*., 2020). It is well understood that expression of *NLPs* is not induced by availability of N (Masclaux-Daubresse *et al*., 2010) however, the *NLPs* directs initial response to N by nuclear-retention mechanism to localize *NLPs* (Marchive *et al*., 2013) therefore, N availability cause higher accumulation of *NLPs* proteins which ultimately enhances expression of N responsive genes enabling plants to utilize larger quantities of N. Although such studies have sought to encompass detailed characterization of NLPs in several vascular plants, yet, NLPs have not been explicitly studied in non-vascular bryophytes. Our findings suggest that the same phenomenon is conserved in non-vascular *P. patens*. The expression pattern of all *PpNLPs* remained unchanged with the passage of time in N deficient (0 mM) condition. It is more likely due to the reason that *P. patens* initially grown on normal BCDA contained N which expressed *PpNLPs* but, later on, shifting plants to N deficient environment could not over-express the *PpNLPs*. On the other hand, expression of *PpNLPs* increased with increasing N supply as well as treatment duration from 0, 6, 12, 24, 48, and 72 hours under both N-limiting (5 mM) and -sufficient (10 mM) conditions. The normal BCDA medium contains 10 mM N concentration thus our experiment provided three levels of N concentrations; the absent or deficient (0 mM), half or limiting (5 mM), and normal or sufficient (10 mM). It is preliminarily evident from this experiment that *PpNLP* orchestrates response to N availability. Developing over-expression as well as mutant *ppnlps* could further attest to these mechanisms.

The whole-genome sequence of the first as well as model bryophyte (Physcomitrella patens) published in 2008 (Rensing *et al*., 2008) provided the opportunity to study *PpNLPs* in the current study. Although genome-wide studies do not confirm the actual detailed molecular mechanisms happening inside a cell, however, such studies are significantly effective in mining a genome database for initial identification as well as preliminary indication of structural and functional attributes of a particular gene. Such genome-wide studies directed before remained helpful as well as are validated through detailed investigations comprehended later on (Ge *et al*., 2018;

Jagadhesan *et al*., 2020). In current study, we identified 6 *NLPs* genes through genome-wide *in silico* analysis in *P. patens* genome-databases and compared their attributes with *NLPs* of *A. thaliana*. The *in silico* studies are largely based on comparison algorithms, therefore, the similarities observed in comparing genomic information can be used to predict function of a gene. We observed that gene lengths, protein lengths, and molecular weights of *PpNLPs* were found higher as compared to *AtNLPs*, however, the pI and GRAVY values of both gene families were found in proximity indicating putative functional homology among the members of both gene families.

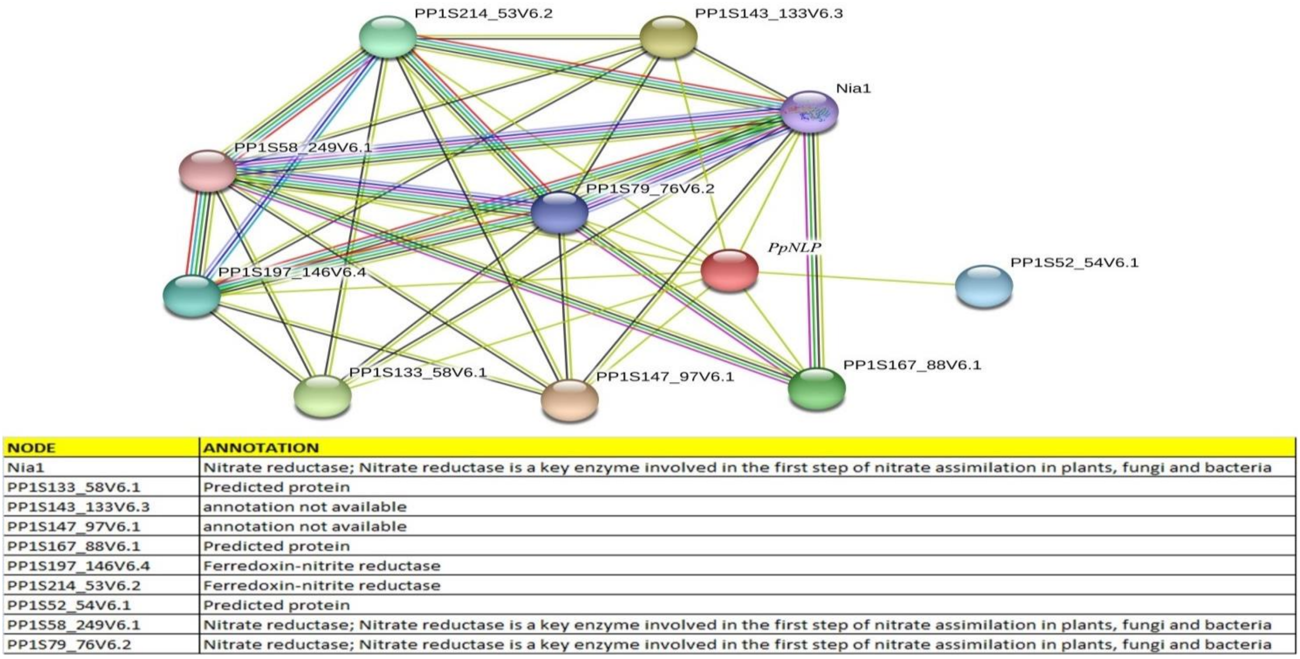
Table S6. Chromosomal distribution of *PpNLPs* gene family.

Gene	Coordinates	Locus
Gene1 PpNLP1 Chr9	9,756,113 - 9,763,070	LOC112286382
Gene2 PpNLP2 Chr12	1,717,295 - 1,723,429	LOC112289804
Gene3 PpNLP3 Chr15	6,095,293 - 6,101,583	LOC112292789
Gene4 PpNLP4 Chr17	3,527,588 - 3,532,842	LOC112294425
Gene5 PpNLP5 Chr19	1,515,069 - 1,520,361	LOC112296151
Gene6 PpNLP6 Chr22	3,740,757 - 3,746,564	LOC112275200

The study of evolutionary relationship among *AtNLPs* and *PpNLPs* clustered them into three distinct clades in a phylogenetic tree, as shown in Figure 1. All the *PpNLPs* were clustered in a sub-clade while sister-group contained 6 members with 2 members from each of *A. thaliana* (*AtNLP8, 9*), *O. sativa* (*OsNLP2, 5*), and *Z. mays* (*ZmNLP2, 9*). Two logical explanations can be inferred from this phylogenetic relationship. First, all the *PpNLPs* are grouped in a separate sub-cluster which may be due to the evolutionary lineage among vascular and non-vascular plants. Second, the presence of *PpNLPs* in close relationship with *NLPs* from vascular plants in sister-group confirms the ancestral lineage of NLPs among bryophytes and vascular plants. In a relevant study of assessing the significance of evolution in amino acid permeases (AAPs) gene families of 17 plants confirmed that bryophytes and vascular plants had common ancestor and gene duplications occurred in evolutionary phases (Zhang *et al*., 2020). The evolutionary relationship can also be linked with the properties of NLPs genes and protein sequence (Yandell *et al*., 2006). The gene structure analysis (Fig. 2) showed that members of *PpNLPs* had 3-4 introns while it varied between 4 and 6 among members of *AtNLPs*. It is evident from previous reports that gene structure evolution is suggested by loss or gain of introns (Zhang *et al*., 2014). Our findings entail higher phylogenetic divergence with higher ancestral linkage among members of vascular and non-vascular NLPs. Presence of one or both of the two protein domains (RWP-RK, and PB1) also explicates the evolutionary relationship among members of *AtNLPs* and *PpNLPs*. Likewise, presence of consensus protein motifs among all the *PpNLPs* further confirms both the ancestral relationship as well as evolutionary divergence of NLP gene families in bryophytes and vascular plants.

Identification of cis-elements in promoter region of a gene is an effective parameter in proposing the role and regulation of a gene. It was observed in our study that *PpNLPs* have higher frequency of cis-elements responsive to plant growth and development that can be related with the growth and development of plant affected by N supply and regulation. The results suggested that more the number of cis-elements - higher will be the associated function. Although it is purely suggested through *in silico* tools from our study that all *PpNLPs* are primarily involved in plant growth development mechanisms while stress as well as phytohormone responses may be their secondary role, however, this statement can be confirmed through detailed investigations led by advance molecular techniques.

Analysis of predicting proteins interacting with a gene family is yet another preliminary procedure in directing functional characterization. Comparing with expression profiles suggest that the predicted proteins enlisted might have conserved function in N uptake, transport, and assimilation. As demonstrated in previous studies, functional characterization of NLP genes in rice showed that they are responsive to N and are significant in improving overall NUE (Alfatih *et al*., 2020; Wu *et al*., 2020).

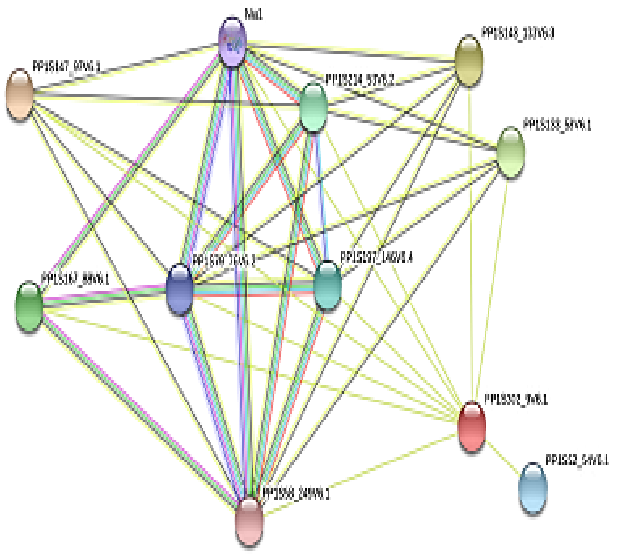


NODE	ANNOTATION
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria

Fig. 6. Schematic representation of common proteins interacting with all members of *PpNLPs*.

Table S7. Protein-protein suggested interaction of *PpNLP* gene family.

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria



***PpNLP2* (PP1S128_79V6.1)**

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria

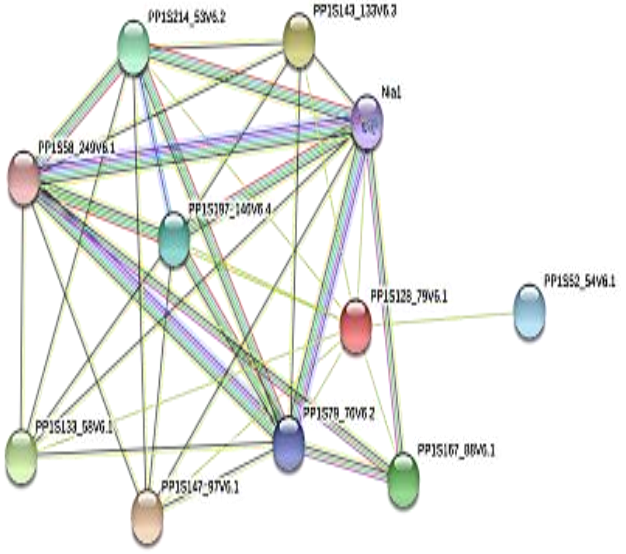
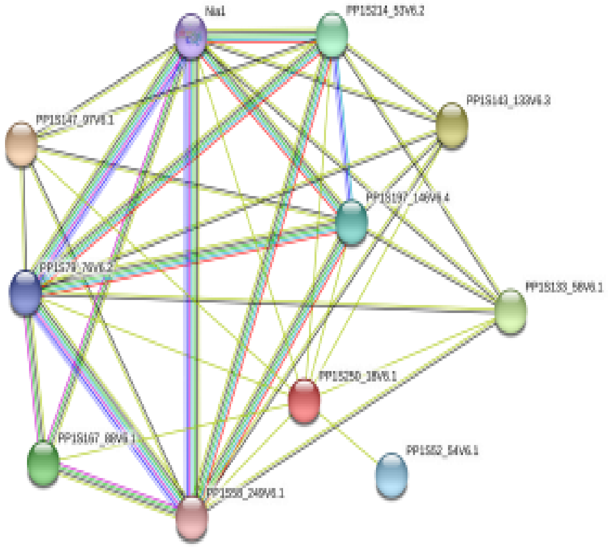


Table S7. (Cont'd.).

***PpNLP3* (PP1S250_18V6.1)**

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria



***PpNLP4* (PP1S26_246V6.1)**

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria

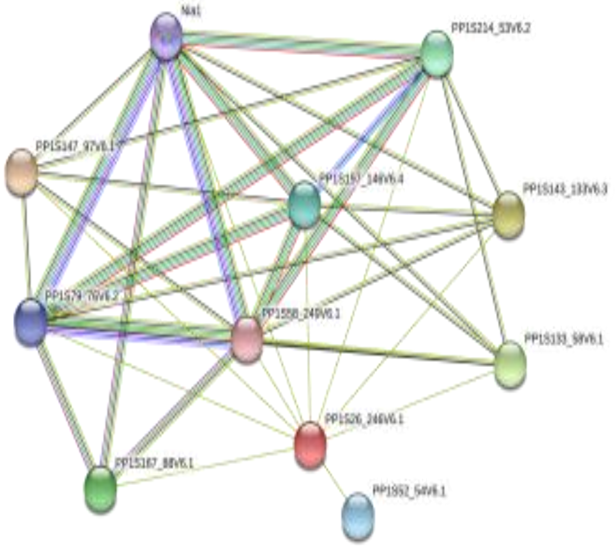
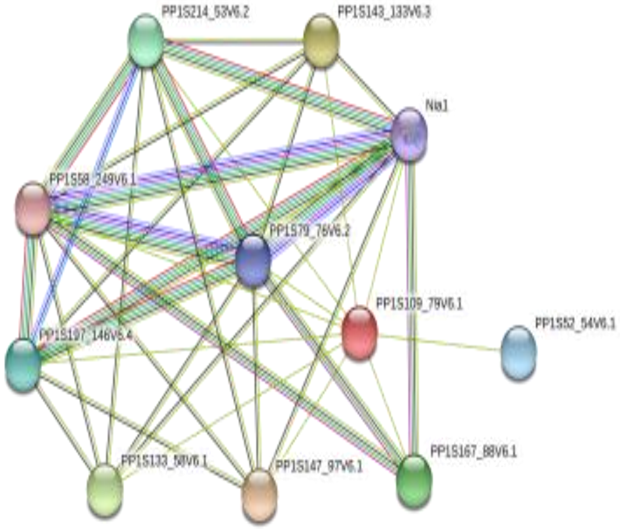


Table S7. (Cont'd.).

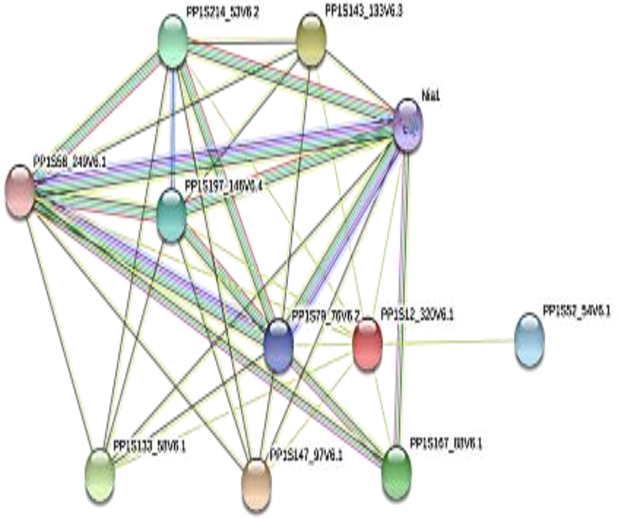
***PpNLP5* (PP1S109_79V6.1)**

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria



***PpNLP6* (PP1S12_320V6.1)**

Node	Annotation
Nia1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S133_58V6.1	Predicted protein
PP1S143_133V6.3	annotation not available
PP1S147_97V6.1	annotation not available
PP1S167_88V6.1	Predicted protein
PP1S197_146V6.4	Ferredoxin-nitrite reductase
PP1S214_53V6.2	Ferredoxin-nitrite reductase
PP1S52_54V6.1	Predicted protein
PP1S58_249V6.1	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria
PP1S79_76V6.2	Nitrate reductase; Nitrate reductase is a key enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria



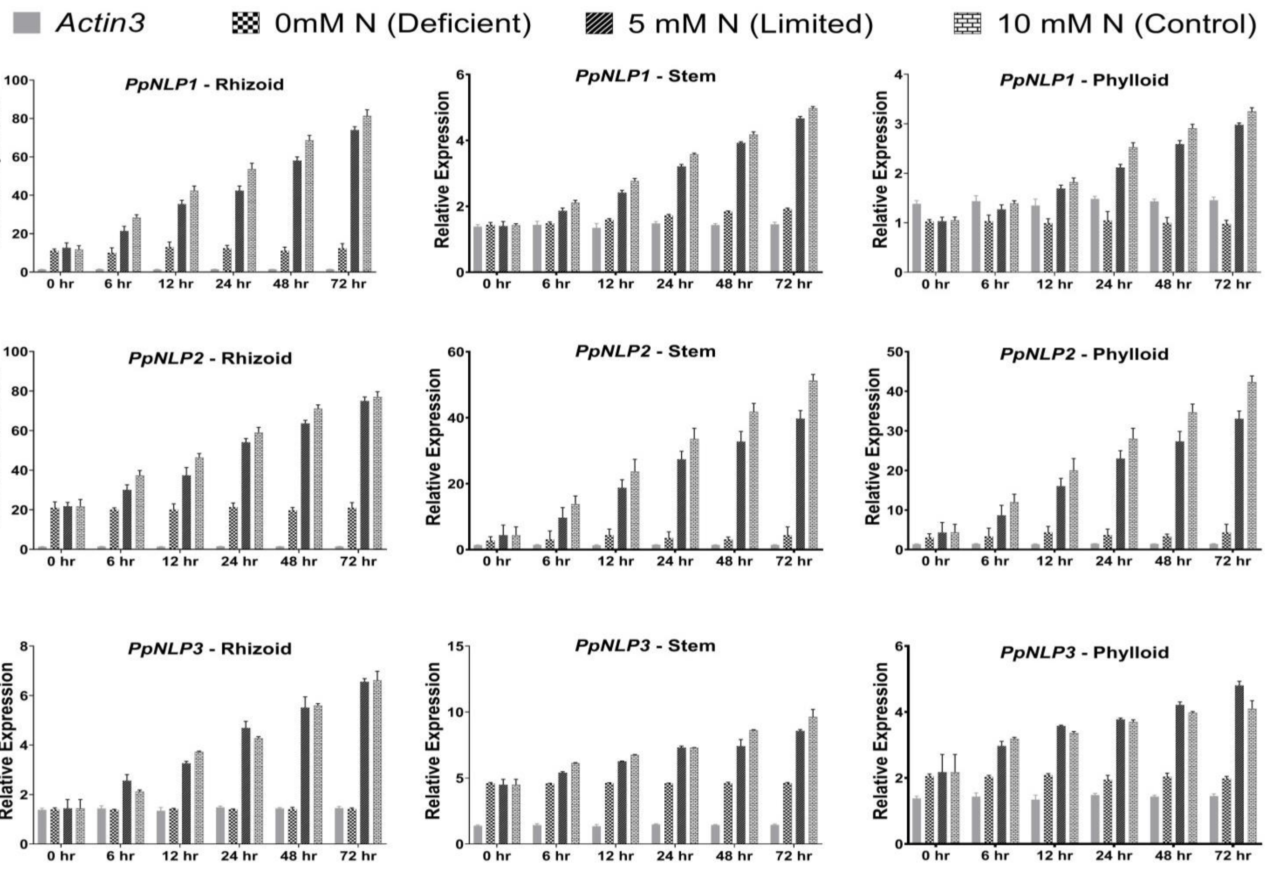


Fig. 7. Expression profile of *PpNLP1*, *2*, & *3* in response to variable nitrogen supply.

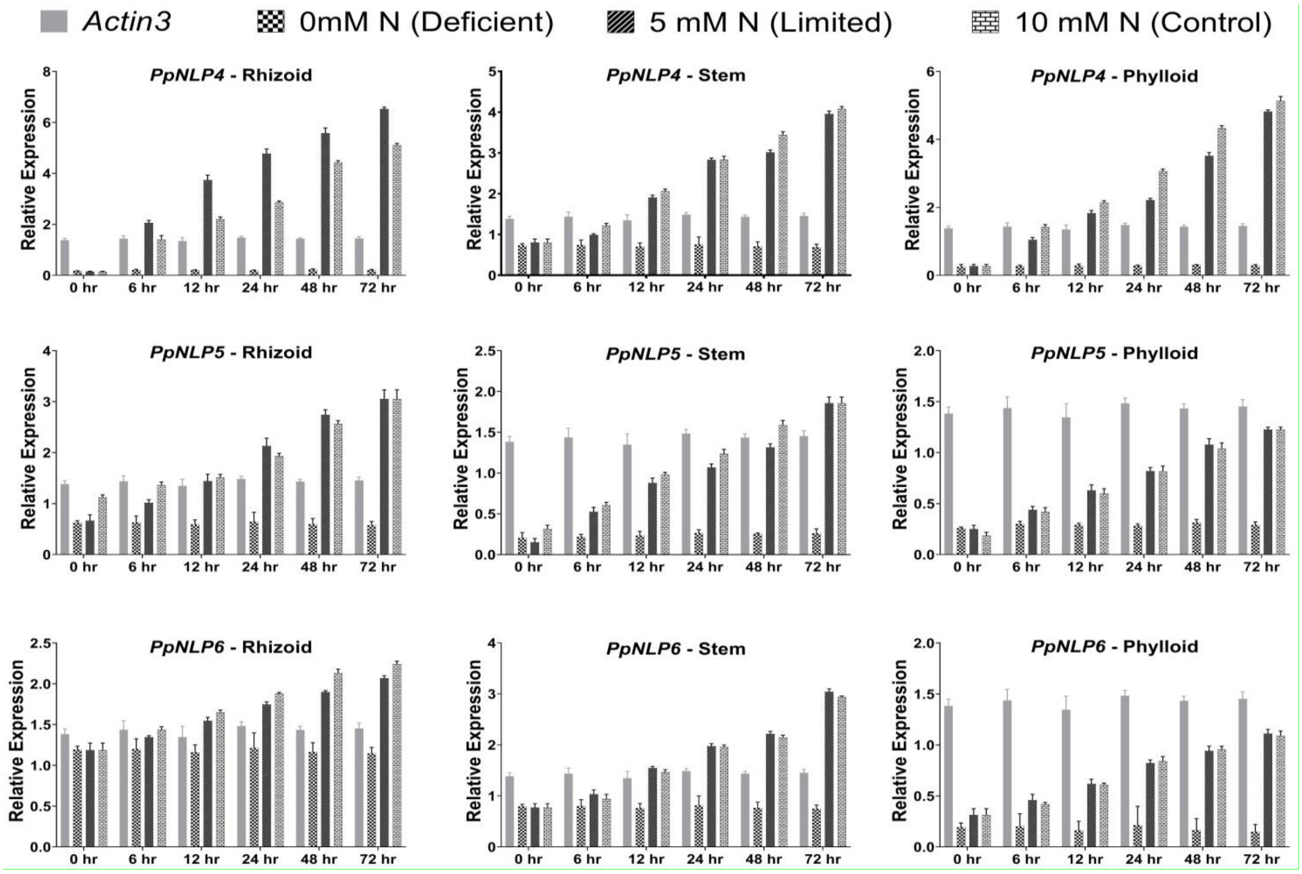


Fig. 8. Expression profile of *PpNLP4*, *5*, & *6* in response to variable nitrogen supply.

Table 3. Number of *cis*-regulatory elements identified in promoter regions of *AtNLPs* and *PpNLPs* Gene Families.

Gene	Plant growth and development							Stress responsive						Phytohormone responsive								
	Box 4	MRE	CAT-Box	O2-site	circadian	GCN4_Motif	MSA-Like	WUN-Motif	ARE	MBS	TC-rich repeats	LTR	GC-motif	CGTCA-Motif	TGACG-Motif	GARE-Motif	P-Box	TATC-Box	ABRE	ERE	TGA-element	TCA-element
AtNLP1		1		2	1				3	2	1			1	1				3		1	1
AtNLP2	1			1					5	1	1			6	6		1	1	2		1	
AtNLP3	3	1		1					2		1						1	1	5		1	1
AtNLP4	2	1	1						1		1							1	1	3		1
AtNLP5	5	1	1						3	1				1	1	1			2			1
AtNLP6	1	1					1		2		1	1		2	2		2		1		1	1
AtNLP7	6	2	1		1			1	4		3	1					1	2	2	1	2	
AtNLP8	2		2					1	3	1	2	1		3	3				1			1
AtNLP9	4	1				1	1		2					2	2		1		5	2	1	4
PpNLP1			1			1				1	1	2		2	2				2	2		1
PpNLP2				2					2			1	1			1	1		1			1
PpNLP3	1							1	1					1	1				2	1		
PpNLP4	6		1	2				3								1				4		
PpNLP5	2		1			1		1	2	1		1		1	1	1			3			
PpNLP6					1				2	1				2	2			1	1			

Number of *cis*-elements indicated by distinguished colors

0	1	2	3	4	5	6

Conclusion

It is concluded on the basis of our findings in this study compared with those reported earlier, that *PpNLPs* are responsive to as well as are significantly regulated by N availability. NLPs are promising group of transcription factors that could potentially contribute in improving crop's N use efficiency (NUE). Our study provides only a hypothetical basis for the study of NLPs thus highlights questions for further detailed investigations. First, detailed structural and functional characterization by employing mutant studies can truly speck their molecular attributes. Our aim in studying NLPs in *Physcomitrella patens* was to fill the gap due to lack of relevant reports. *Physcomitrella patens* shall be focused for such studies, particularly for N transport, because it lies on the border-line of algae and vascular plants – thus can be promising for exploiting detailed mechanisms and key factors involved in N regulation for improving crop's NUE.

References

Alfatih, A., J. Wu, Z.S. Zhang, J.Q. Xia, S.U. Jan, L.H. Yu and C.B. Xiang. 2020. Rice NIN-LIKE PROTEIN 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency. *J. Exp. Bot.*, 71: 6032-6042.

Anonymous. 2019. World fertilizer trends and outlook to 2022.

Bailey, T.L, J. Johnson, C.E. Grant and W.S. Noble. 2015. The MEME Suite. *Nuc. Acids Res*., 43: 39-49.

Bustin, S. and J. Huggett. 2017. qPCR primer design revisited. *Biomol. Detect. Quantif.,* 14: 19-28.

Chardin, C., T. Girin, F. Roudier, C. Meyer and A. Krapp. 2014. The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development. *J.Exp. Bot*., 65: 5577-5587 doi:10.1093/jxb/eru261.

Fagodiya, R.K., H. Pathak, A. Kumar, A. Bhatia and N. Jain. 2017. Global temperature change potential of nitrogen use in agriculture: A 50-year assessment. *Sci. Rep.*, 7: 44928.

Feng, H., X. Fan, A.J. Miller and G. Xu. 2020. Plant nitrogen uptake and assimilation: regulation of cellular pH homeostasis. *J. Exp. Bot.*, 71: 4380-4392.

Ferris, P.J and U.W. Goodenough. 1997. Mating type in chlamydomonas is specified by mid, the Minus-dominance gene. *Genetics*, 146: 859.

Forde, B.G. and P.J. Lea. 2007. Glutamate in plants: metabolism, regulation, and signalling. *J. Exp. Bot*., 58: 2339-2358.

Garnett, T., V. Conn and B.N. Kaiser. 2009. Root based approaches to improving nitrogen use efficiency in plants. *Plant Cell Environ.,* 32: 1272-1283.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nuc. Acids Res.,* 31: 3784-3788.

Ge, M., Y. Liu and L. Jiang. 2018. Genome-wide analysis of maize NLP transcription factor family revealed the roles in nitrogen response. *Plant Growth Regul.,* 84: 95-105.

Hu, B., J. Jin, A.Y. Guo, H. Zhang, J. Luo and G. Gao. 2015. GSDS 2.0: an upgraded gene feature visualization. *Server Bioinfor.,* 31: 1296-1297.

Jagadhesan, B., L. Sathee, H.S. Meena, S.K. Jha, V. Chinnusamy, A. Kumar and S. Kumar. 2020. Genome wide analysis of NLP transcription factors reveals their role in nitrogen stress tolerance of rice. *Sci. Rep.*, 10: 9368.

Jan, S.A., Z.K. Shinwari, I. Khan, S. Khan, A. Iqbal and H. Khurshid. 2022. CRISPR-CAS9 mediated genome editing in plants against viruses: an updated review. *Pak. J. Bot.*, 54(4): 1575-1578.

Kan, C.C., T.Y. Chung, Y.A. Juo and M.H. Hsieh. 2015. Glutamine rapidly induces the expression of key transcription factor genes involved in nitrogen and stress responses in rice roots. *BMC Genomics.,* 16: 731-731.

Kant, S., Y.M. Bi and S.J. Rothstein. 2011. Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J. Exp. Bot*., 62: 1499-1509.

Khurshid, H., S.A. Jan, Z.K. Shinwari, M. Jamal and S.H. Shah. 2018. An era of CRISPR/ Cas9 mediated plant genome editing. *Curr. Issue. Mol. Biol*., 26: 47-55.

Koduri, P.K.H., G.S. Gordon, E.I. Barker, C.C. Colpitts, N.W. Ashton and D.Y. Suh. 2010. Genome-wide analysis of the chalcone synthase superfamily genes of Physcomitrella patens. *Plant Mol. Biol*., 72: 247-263.

Konishi, M. and S. Yanagisawa. 2013. Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. *Nat. Commun*., 4: 1617.

Li, S., Y. Tian, K. Wu and Y. Ye. 2018. Modulating plant growth–metabolism coordination for sustainable agriculture. *Nat.,* 560: 595-600.

Liping, R., Z. Jinbo, C. Xiaohan, W. Wenyang, Y. Dandan and S. Xiaohui. 2021. Transcriptome-wide identification and functional characterization of BBX transcription factor family in Toona sinensis. *Pak. J. Bot*., 53(3): 915-921.

Liu, M., W. Chang and Y. Fan. 2018. Genome-wide identification and characterization of nodule-inception-like protein (NLP) family genes in *Brassica napus*. *Int. J. Mol. Sci.,* 19: 2270.

Marchive, C., F. Roudier and L. Castaings. 2013. Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants. *Nat. Commun.,* 4: 1713.

Masclaux-Daubresse, C., F. Daniel-Vedele, J. Dechorgnat, F. Chardon, L. Gaufichon and A. Suzuki. 2010. Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Annal. Bot.,* 105: 1141-1157.

Moll, R.H., E.J. Kamprath and W.A. Jackson. 1982. Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization1. *Agron. J*., 74: 562-564.

Mu, X. and J. Luo. 2019. Evolutionary analyses of NIN-like proteins in plants and their roles in nitrate. *Signaling.,* 76: 3753-3764.

Olas, J.J. and V. Wahl. 2019. Tissue-specific NIA1 and NIA2 expression in *Arabidopsis thaliana*. *Plant Signal Behav*., 14: 1656035-1656035.

Orioli, T. and M. Vihinen. 2019. Benchmarking subcellular localization and variant tolerance predictors on membrane proteins. *B.M.C. Genomics*., 20: 547.

Orsel, M., A. Krapp and F. Daniel-Vedele. 2002. Analysis of the NRT2 nitrate transporter family in Arabidopsis. Structure gene expression. *Plant Physiol*., 129: 886-896.

Qiu, J., S.W. Henderson, M. Tester, S.J. Roy and M. Gilliham. 2016. SLAH1, a homologue of the slow type anion channel SLAC1, modulates shoot Cl- accumulation and salt tolerance in Arabidopsis thaliana. *J. Exp. Bot*., 67: 4495-4505.

Rensing, S.A., D. Lang and A. Salamov. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Sci.,* 319: 64-69.

Schauser, L., A. Roussis, J. Stiller and J. Stougaard. 1999. A plant regulator controlling development of symbiotic root nodules. *Nat.,* 402: 191-195.

Shah, S.H., S. Ali, Z. Hussain and G.M. Ali. 2016. Genetic improvement of tomato (*Solanum lycopersicum*) with *AtDREB1A* gene for cold stress tolerance using optimized *Agrobacterium*-mediated transformation system. *Int. J. Agric. Biol*., 18: 471-482.

Shinwari, Z.K., S.A. Jan, K. Nakashima and K. Yamaguchi-Shinozaki. 2020. Genetic engineering approaches to understanding drought tolerance in plants. *Plant Biotechnol. Rep.,* 14(2): 151-162.

Szklarczyk, D., A.L. Gable and A. Junge. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nuc. Acids Res.*, 47: 607-613.

Szklarczyk, D., J.H. Morris and M. Kuhn. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nuc. Acids Res*., 45: 362-368.

Takahashi, M., Y. Sasaki, S. Ida and H. Morikawa. 2001. Nitrite reductase gene enrichment improves assimilation of NO(2) in Arabidopsis. *Plant Physiol*., 126: 731-741.

Tegeder, M. and C. Masclaux-Daubresse. 2018. Source and sink mechanisms of nitrogen transport and use *New Phytol*., 217: 35-53.

Unno, H., T. Uchida and G. Kurisu. 2006. Atomic structure of plant glutamine synthetase: a key enzyme for plant productivity. *J. Biol. Chem*., 281: 29287-29296.

Verma, G., Y.V. Dhar and M. Kidwai. 2017. Genome-wide analysis of rice dehydrin gene family: Its evolutionary conservedness and expression pattern in response to PEG induced dehydration stress. *PLOS ONE.,* 12: e0176399.

Wu, J., Z.S. Zhang and J.Q. Xia. 2020. Rice NIN-LIKE PROTEIN 4 plays a pivotal role in nitrogen use efficiency *Plant Biotechnol. J*., 19(03): 448-461.

Xiao, L., H. Wang, P. Wan, T. Kuang and Y. He. 2011. Genome-wide transcriptome analysis of gametophyte development in Physcomitrella patens. *BMC Plant Biol*., 11: 177.

Yandell, M., C.J. Mungall and C. Smith. 2006. Large-scale trends in the evolution of gene structures within 11 animal genomes. *PLOS Comp. Biol*., 2: e15.

Yin, M., Z. Zhang, M. Xuan, H. Feng, W. Ye, X. Zheng and Y. Wang. 2020. Conserved subgroups of the plant-specific RWP-RK transcription factor family are present in oomycete pathogens. *Front. Microbiol*., 11: 1724-1724.

Yokota, K. and M. Hayashi. 2011. Function and evolution of nodulation genes in legumes. *Cellular Mol. Life Sci*., 68: 1341-1351.

Zhang, C., A.R. Gschwend, Y. Ouyang and M. Long. 2014. Evolution of gene structural complexity: An alternative-splicing-based model accounts for intron-containing retrogenes. *Plant Physiol*., 165: 412.

Zhang, C., N. Kong, M. Cao, D. Wang, Y. Chen and Q. Chen. 2020. Evolutionary significance of amino acid permease transporters in 17 plants from Chlorophyta to Angiospermae. *B.M.C. Genomics*., 21: 391.

Zhao, L., F. Liu, N.M. Crawford and Y. Wang. 2018. Molecular regulation of nitrate responses in plants. *Int. J. Mol. Sci*., 19: 2039.

Zifarelli, G. and M. Pusch. 2010. CLC transport proteins in plants. *FEBS Lett*., 584: 2122-2127.

(Received for publication 12 July 2021)