Nodule-Inception-Like Protein (*NLP*) Gene Family Identified in *Physcomitrella Patens* Genome Responds to Variable Nitrogen Supply

Muhammad Irfan Fareed
University of Science and Technology of China
Sami Ullah Jan
University of Science and Technology of China
Tahmina Nazish
University of Science and Technology of China
Ayesha Liaqat
University of Science and Technology of China
Ayesha Javaid
University of Science and Technology of China
Mohsin Ali
University of Science and Technology of China
Farees Ud Din Mufti
University of Science and Technology of China
Muhammad Fayyaz
Kohat University of Science and Technology
Maha Rehman
Kohat University of Science and Technology
Muhammad Jamil
Kohat University of Science and Technology
Sarah Gul
International Islamic University
Faiza Munir
National University of Sciences and Technology
Rabia Amir
National University of Sciences and Technology
Alvina Gul (alvina_gul@asab.nust.edu.pk)
National University of Sciences and Technology

Research Article

Keywords: Nitrogen, Nitrate, NODULE-INCEPTION-Like Protein (NLP), RWP-RK Domain, PB1 Domain, Physcomitrella patens

Posted Date: June 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-554676/v1

Version of Record: A version of this preprint was published at Pakistan Journal of Botany on June 20th, 2022. See the published version at https://doi.org/10.30848/PJB2023-1(41).

Abstract

NODULE-INCEPTION-like proteins (NLPs) are plant specific transcription factors that play significant role in orchestrating nitrogen response. NLPs have been widely studied in vascular plants but very less is known about NLPs in non-vascular bryophytes till date. In current study, first, the *in silico* tools were employed for identification and characterization of NLPs in model bryophyte *Physcomitrella patens* genome. 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 possesses similarities in there iso-electric point and hydropathicity values with those of *AtNLPs* while gene lengths, protein lengths, and molecular weights were found higher in *PpNLPs* than *AtNLPs*. It was further observed that all *PpNLPs*, except *PpNLP6*, yield acidic hydrophilic proteins localized in 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 nitrate (5 and 10 mM) as sole nitrogen source while no significant increment was observed in the absence (0 mM) of nitrogen. The expression levels increased with increasing retention-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 Message

It is established in various vascular plants that *NODULE-INCEPTION-LIKE PROTEIN* (NLP) gene family is responsive to nitrogen. Our study aimed for genome-wide identification and characterization to identify NLP gene family in *Physcomitrella patens* genome. Furthermore, we found through expression analysis that NLPs are responsive to nitrogen supply in model bryophyte.

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 pollutions (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 118 Mt in 2019 (FAO 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 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 can 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* (Orsel et al. 2002) and *NRT2* (Orsel et al. 2002) across the channels including *CLC*: chloride channel (Zifarelli and 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 and 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 and 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 molecule 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 and 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 japonicas* which also contains RWP-RK domain (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 and Luo 2019). *NIN* and *NLPs* are found in both non-leguminous and leguminous plants (van Velzen et al. 2018; Yokota and 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 and Yanagisawa 2013). A key feature of *NLP* proteins is their response to N is nitrate-stimulated rapid nuclear translocation mechanism (Marchive et al. 2013). Overall, the *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, less is known about *NLPs* in non-vascular plants 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-wide 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). 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 various parts of *P. patens* 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.

RNA Extraction and qPCR

The total cellular RNA from selected three parts; rhizoid, stem, and phylloid, was extracted with column-based RNA extraction methods (Seçgin et al. 2020). 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 and 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 and 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 (plantCARE : 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).

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 their location in chromosomes. 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.

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)

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*. Figure 6 shows schematic model of all *PpNLPs* interacting with cellular proteins.

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 (Fig. 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.

Discussion

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 detail 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.

The study of evolutionary relationship among *AtNLPs* and *PpNLPs* clustered them into three distinct clades in a phylogenetic tree, as shown in Fig. 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).

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.

Declarations

Funding

Not applicable.

Conflict of interests

The authors declare no conflicts of interests.

Availability of data and material

All data and materials used in this study are attached.

Authors' contribution

The study conception and design were perceived by Sami Ullah Jan and Alvina Gul. The first

four authors did all the data analysis. The validation of results was done by Sami Ullah Jan.

Overall supervision and review of the manuscript were done by Alvina Gul.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Alfatih A, Wu J, Zhang Z-S, Xia J-Q, Jan SU, Yu L-H, Xiang C-B (2020) Rice NIN-LIKE PROTEIN 1 rapidly responds to nitrogen deficiency and improves yield and nitrogen use efficiency Journal of Experimental Botany 71:6032-6042 doi:10.1093/jxb/eraa292
2. Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite Nucleic acids research 43:W39-W49 doi:10.1093/nar/gkv416
3. Bustin S, Huggett J (2017) qPCR primer design revisited Biomolecular detection and quantification 14:19-28 doi:10.1016/j.bdq.2017.11.001
4. Chardin C, Girin T, Roudier F, Meyer C, Krapp A (2014) The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development Journal of Experimental Botany 65:5577-5587 doi:10.1093/jxb/eru261
5. Fagodiya RK, Pathak H, Kumar A, Bhatia A, Jain N (2017) Global temperature change potential of nitrogen use in agriculture: A 50-year assessment Scientific Reports 7:44928 doi:10.1038/srep44928

6. FAO (2019) World fertilizer trends and outlook to 2022.
7. Feng H, Fan X, Miller AJ, Xu G (2020) Plant nitrogen uptake and assimilation: regulation of cellular pH homeostasis Journal of Experimental Botany 71:4380-4392 doi:10.1093/jxb/eraa150
8. Ferris PJ, Goodenough UW (1997) Mating Type in Chlamydomonas Is Specified by mid, the Minus-Dominance Gene Genetics 146:859
9. Forde BG, Lea PJ (2007) Glutamate in plants: metabolism, regulation, and signalling Journal of Experimental Botany 58:2339-2358 doi:10.1093/jxb/erm121
10. Garnett T, Conn V, Kaiser BN (2009) Root based approaches to improving nitrogen use efficiency in plants Plant, Cell & Environment 32:1272-1283 doi:10.1111/j.1365-3040.2009.02011.x
11. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis Nucleic acids research 31:3784-3788 doi:10.1093/nar/gkg563
12. Ge M et al. (2018) Genome-wide analysis of maize NLP transcription factor family revealed the roles in nitrogen response Plant Growth Regulation 84:95-105 doi:10.1007/s10725-017-0324-x
13. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server Bioinformatics 31:1296-1297 doi:10.1093/bioinformatics/btu817
14. Jagadhesan B, Sathee L, Meena HS, Jha SK, Chinnusamy V, Kumar A, Kumar S (2020) Genome wide analysis of NLP transcription factors reveals their role in nitrogen stress tolerance of rice Scientific Reports 10:9368 doi:10.1038/s41598-020-66338-6
15. Kan C-C, Chung T-Y, Juo Y-A, Hsieh M-H (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 doi:10.1186/s12864-015-1892-7
16. Kant S, Bi Y-M, Rothstein SJ (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency Journal of Experimental Botany 62:1499-1509 doi:10.1093/jxb/erq297
17. Koduri PKH, Gordon GS, Barker EI, Colpitts CC, Ashton NW, Suh D-Y (2010) Genome-wide analysis of the chalcone synthase superfamily genes of Physcomitrella patens Plant Molecular Biology 72:247-263 doi:10.1007/s11103-009-9565-z
18. Konishi M, Yanagisawa S (2013) Arabidopsis NIN-like transcription factors have a central role in nitrate signalling Nature Communications 4:1617 doi:10.1038/ncomms2621
19. Li S et al. (2018) Modulating plant growth–metabolism coordination for sustainable agriculture Nature 560:595-600 doi:10.1038/s41586-018-0415-5
20. Liu M et al. (2018) Genome-Wide Identification and Characterization of NODULE-INCEPTION-Like Protein (NLP) Family Genes in Brassica napus International journal of molecular sciences 19:2270 doi:10.3390/ijms19082270
21. Marchive C et al. (2013) Nuclear retention of the transcription factor NLP7 orchestrates the early response to nitrate in plants Nature Communications 4:1713 doi:10.1038/ncomms2650
22. Masclaux-Daubresse C, Daniel-Vedele F, Dechorgnat J, Chardon F, Gaufichon L, Suzuki A (2010) Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture Annals of Botany 105:1141-1157 doi:10.1093/aob/mcq028
23. Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization1 Agronomy Journal 74:562-564 doi:10.2134/agronj1982.00021962007400030037x
24. Mu X, Luo J (2019) Evolutionary analyses of NIN-like proteins in plants and their roles in nitrate signaling 76:3753-3764 doi:10.1007/s00018-019-03164-8
25. Olas JJ, Wahl V (2019) Tissue-specific NIA1 and NIA2 expression in Arabidopsis thaliana Plant signaling & behavior 14:1656035-1656035 doi:10.1080/15592324.2019.1656035
26. Orioli T, Vihinen M (2019) Benchmarking subcellular localization and variant tolerance predictors on membrane proteins BMC Genomics 20:547 doi:10.1186/s12864-019-5865-0
27. Orsel M, Krapp A, Daniel-Vedele F (2002) Analysis of the NRT2 nitrate transporter family in Arabidopsis. Structure and gene expression Plant physiology 129:886-896 doi:10.1104/pp.005280
28. Qiu J, Henderson SW, Tester M, Roy SJ, Gilliham M (2016) SLAH1, a homologue of the slow type anion channel SLAC1, modulates shoot Cl- accumulation and salt tolerance in Arabidopsis thaliana Journal of experimental botany 67:4495-4505 doi:10.1093/jxb/erw237
29. Rensing SA et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants Science 319:64-69 doi:10.1126/science.1150646
30. Schauser L, Roussis A, Stiller J, Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules Nature 402:191-195 doi:10.1038/46058
31. Seçgin Z, Gökdemir G, Atabay ES, Kurt Kızıldoğan A, Kavas M (2020) Development of new total RNA isolation method for tissues with rich phenolic compounds Turkish Journal of Biochemistry 45:343-350 doi:doi:10.1515/tjb-2019-0375
32. Szklarczyk D et al. (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets Nucleic Acids Research 47:D607-D613 doi:10.1093/nar/gky1131
33. Szklarczyk D et al. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible Nucleic acids research 45:D362-D368 doi:10.1093/nar/gkw937

34. Takahashi M, Sasaki Y, Ida S, Morikawa H (2001) Nitrite reductase gene enrichment improves assimilation of NO(2) in Arabidopsis Plant physiology 126:731-741 doi:10.1104/pp.126.2.731
35. Tegeder M, Masclaux-Daubresse C (2018) Source and sink mechanisms of nitrogen transport and use New Phytologist 217:35-53 doi:10.1111/nph.14876
36. Unno H et al. (2006) Atomic structure of plant glutamine synthetase: a key enzyme for plant productivity J Biol Chem 281:29287-29296 doi:10.1074/jbc.M601497200
37. van Velzen R et al. (2018) Comparative genomics of the nonlegume Parasponia reveals insights into evolution of nitrogen-fixing rhizobium symbioses Proceedings of the National Academy of Sciences 115:E4700 doi:10.1073/pnas.1721395115
38. Verma G et al. (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 doi:10.1371/journal.pone.0176399
39. Wu J et al. (2020) Rice NIN-LIKE PROTEIN 4 plays a pivotal role in nitrogen use efficiency Plant Biotechnology Journal n/a doi:10.1111/pbi.13475
40. Yandell M et al. (2006) Large-Scale Trends in the Evolution of Gene Structures within 11 Animal Genomes PLOS Computational Biology 2:e15 doi:10.1371/journal.pcbi.0020015
41. Yin M, Zhang Z, Xuan M, Feng H, Ye W, Zheng X, Wang Y (2020) Conserved Subgroups of the Plant-Specific RWP-RK Transcription Factor Family Are Present in Oomycete Pathogens Frontiers in microbiology 11:1724-1724 doi:10.3389/fmicb.2020.01724
42. Yokota K, Hayashi M (2011) Function and evolution of nodulation genes in legumes Cell Mol Life Sci 68:1341-1351 doi:10.1007/s00018-011-0651-4
43. Zhang C, Gschwend AR, Ouyang Y, Long M (2014) Evolution of Gene Structural Complexity: An Alternative-Splicing-Based Model Accounts for Intron-Containing Retrogenes Plant Physiology 165:412 doi:10.1104/pp.113.231696
44. Zhang C, Kong N, Cao M, Wang D, Chen Y, Chen Q (2020) Evolutionary significance of amino acid permease transporters in 17 plants from Chlorophyta to Angiospermae BMC Genomics 21:391 doi:10.1186/s12864-020-6729-3
45. Zhao L, Liu F, Crawford NM, Wang Y (2018) Molecular Regulation of Nitrate Responses in Plants International journal of molecular sciences 19:2039 doi:10.3390/ijms19072039
46. Zifarelli G, Pusch M (2010) CLC transport proteins in plants FEBS Letters 584:2122-2127 doi:https://doi.org/10.1016/j.febslet.2009.12.042

Tables

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

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 Responsi					
	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
AtNLP1		1		2	1				3	2	1			1	1				3
AtNLP2	1			1					5	1	1			6	6		1	1	2
AtNLP3	3	1		1					2		1						1	1	5
AtNLP4	2	1	1						1		1							1	1
AtNLP5	5	1	1						3	1				1	1	1			2
AtNLP6	1	1					1		2		1	1		2	2		2		1
AtNLP7	6	2	1		1			1	4		3	1					1	2	2
AtNLP8	2		2					1	3	1	2	1		3	3				1
AtNLP9	4	1				1	1		2					2	2		1		5
PpNLP1			1			1				1	1	2		2	2				2
PpNLP2				2					2			1	1			1	1		1
PpNLP3	1							1	1					1	1				2
PpNLP4	6		1	2				3								1			
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

Figures

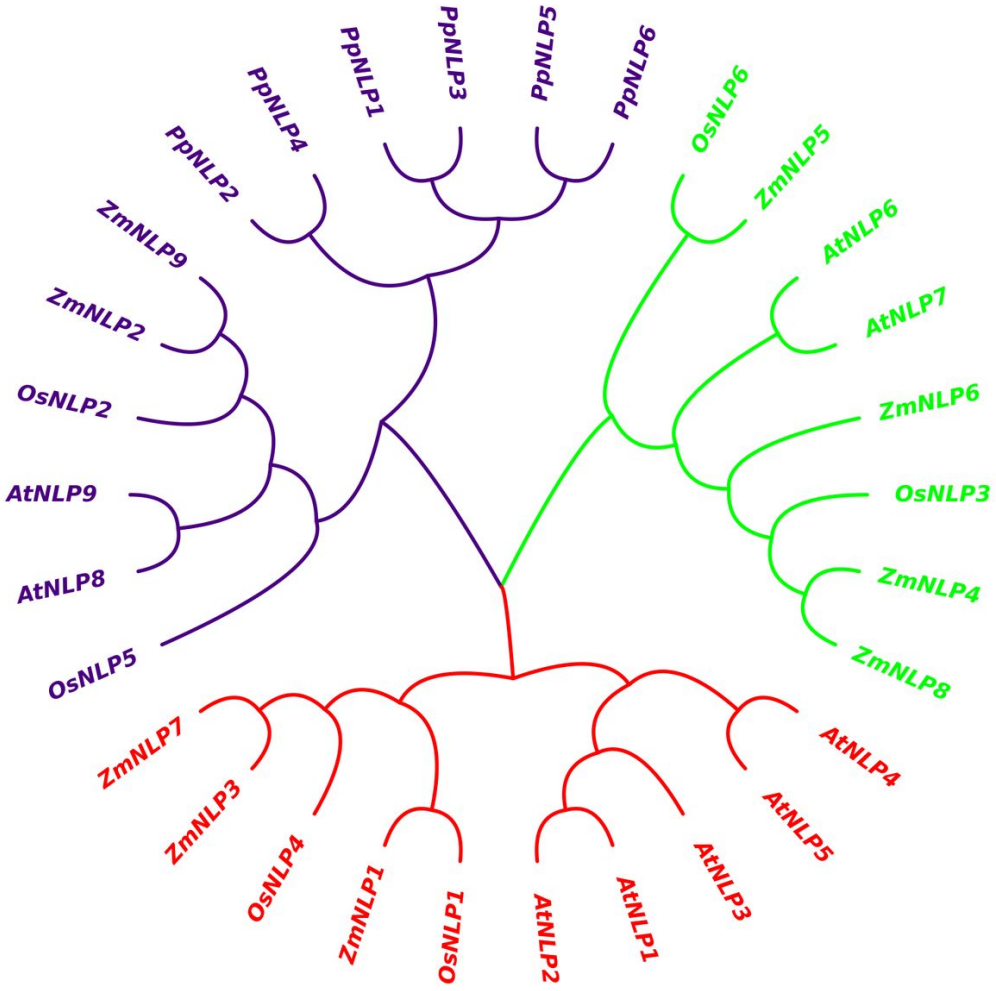


Figure 1

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 (Figure 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.

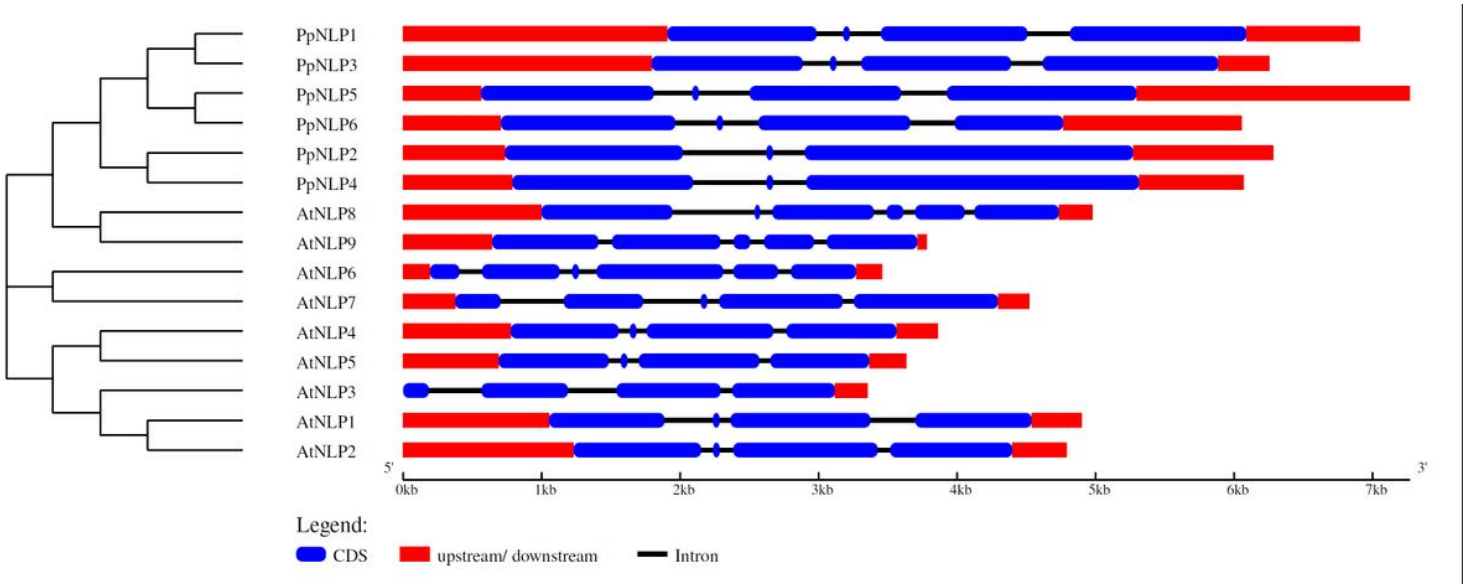


Figure 2

Structural components of AtNLPs and PpNLPs were analyzed using the gene and their coding sequences. Identification of introns, exons, and UTRs in genic region (Figure 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.

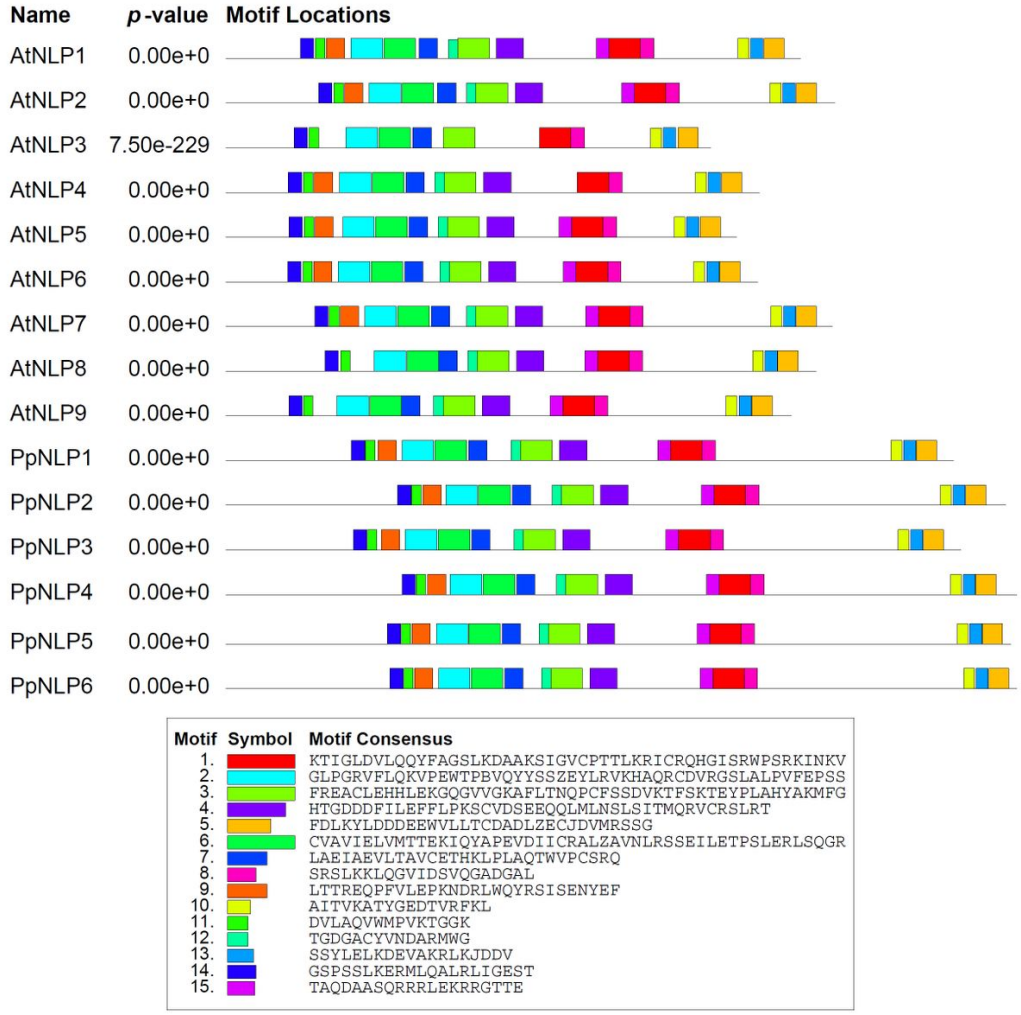


Figure 3

Up to 15 consensus motifs were figured out using MEME in PpNLP proteins (Figure 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.

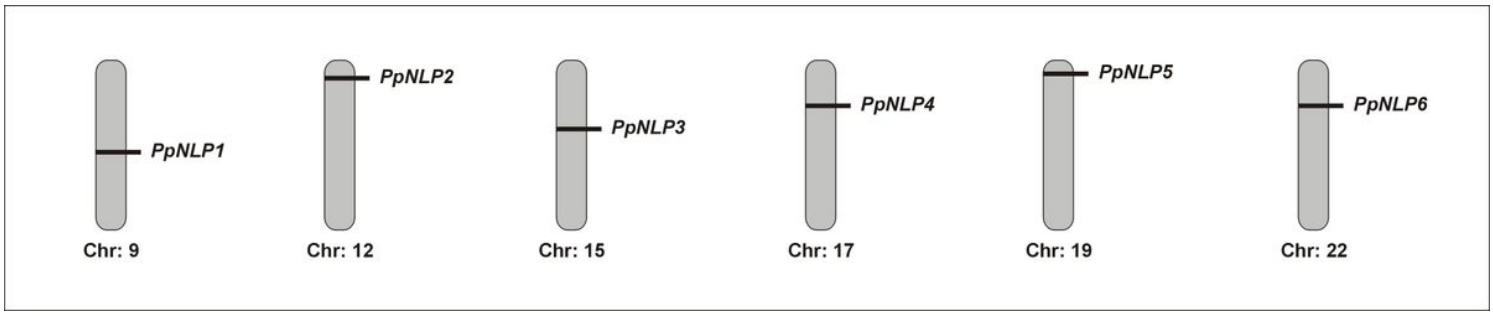


Figure 4

Appropriate localization of genes upon chromosome (Figure 4, Table S6) revealed that 6 PpNLPs are localized on different chromosomes (Chr. 9, 12, 15, 17, 19, 22)

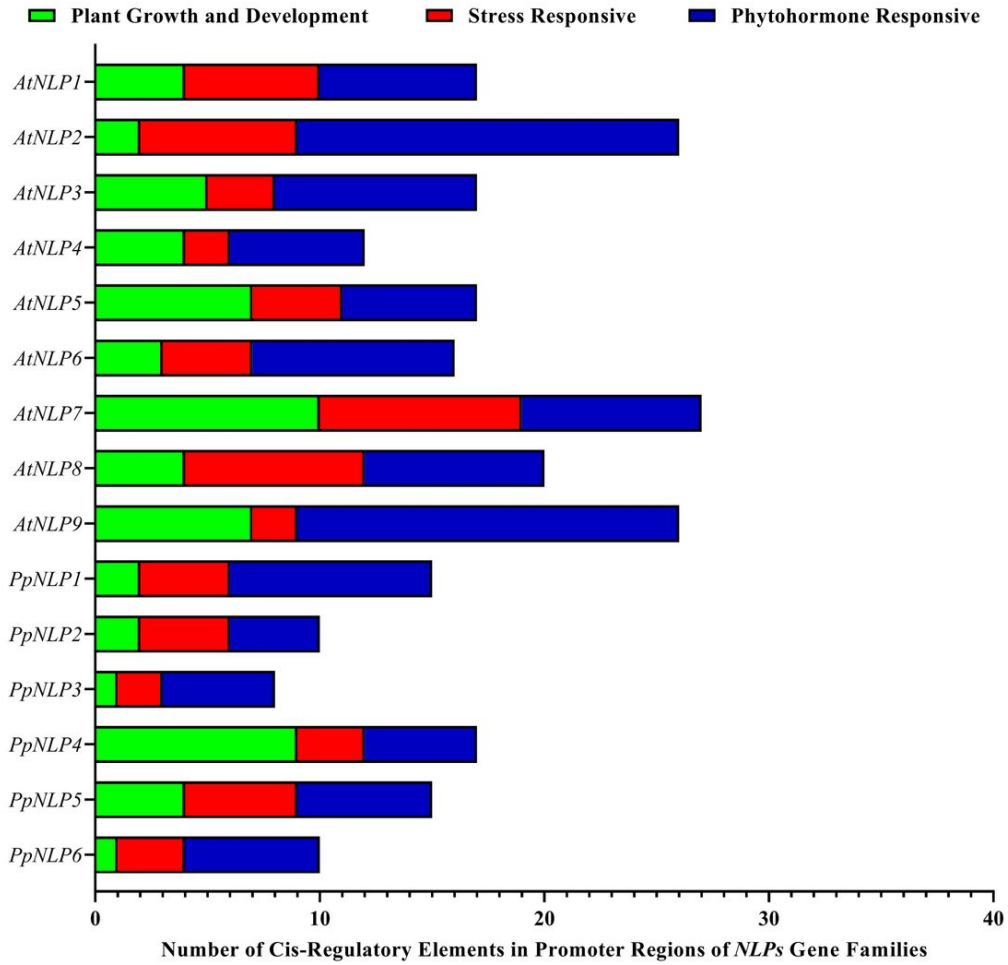
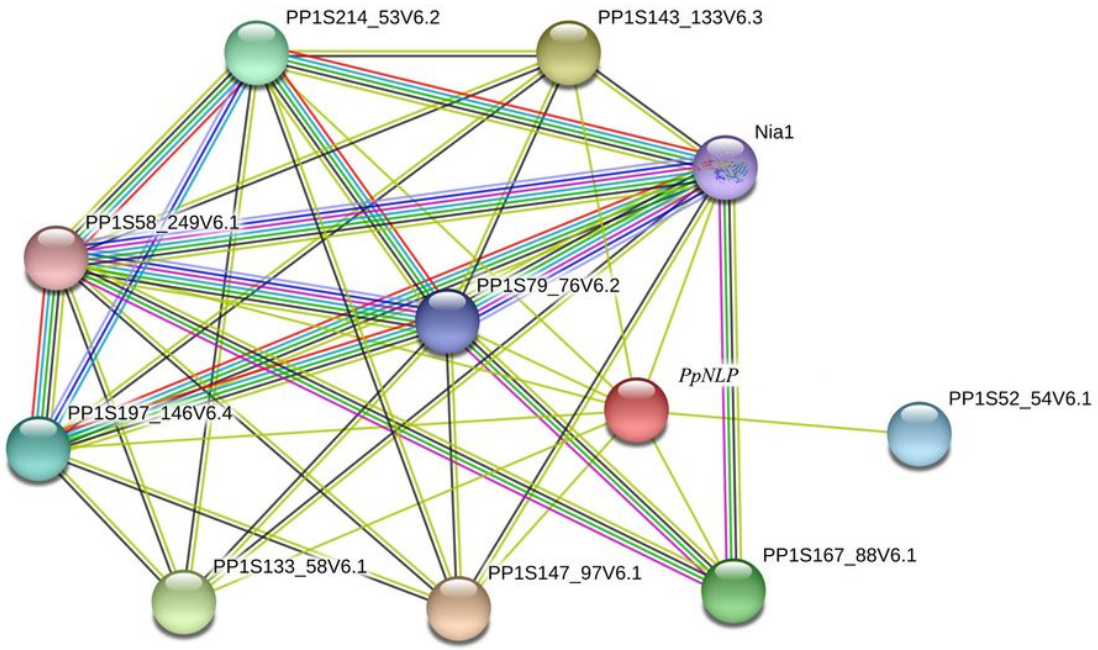


Figure 5

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 (Figure 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.



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

Figure 6

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. Figure 6 shows schematic model of all PpNLPs interacting with cellular proteins.

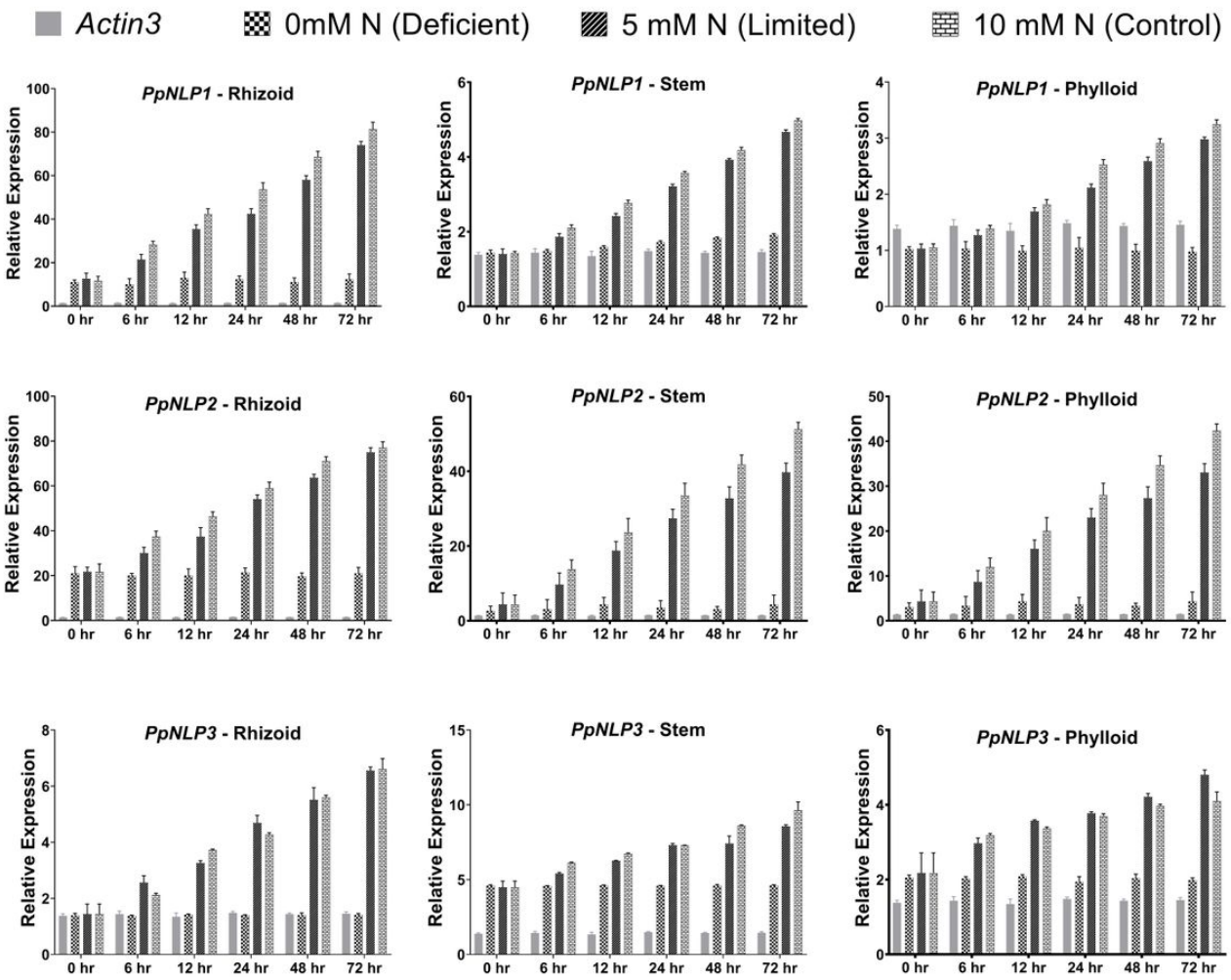


Figure 7

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 (Figure 7). 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.

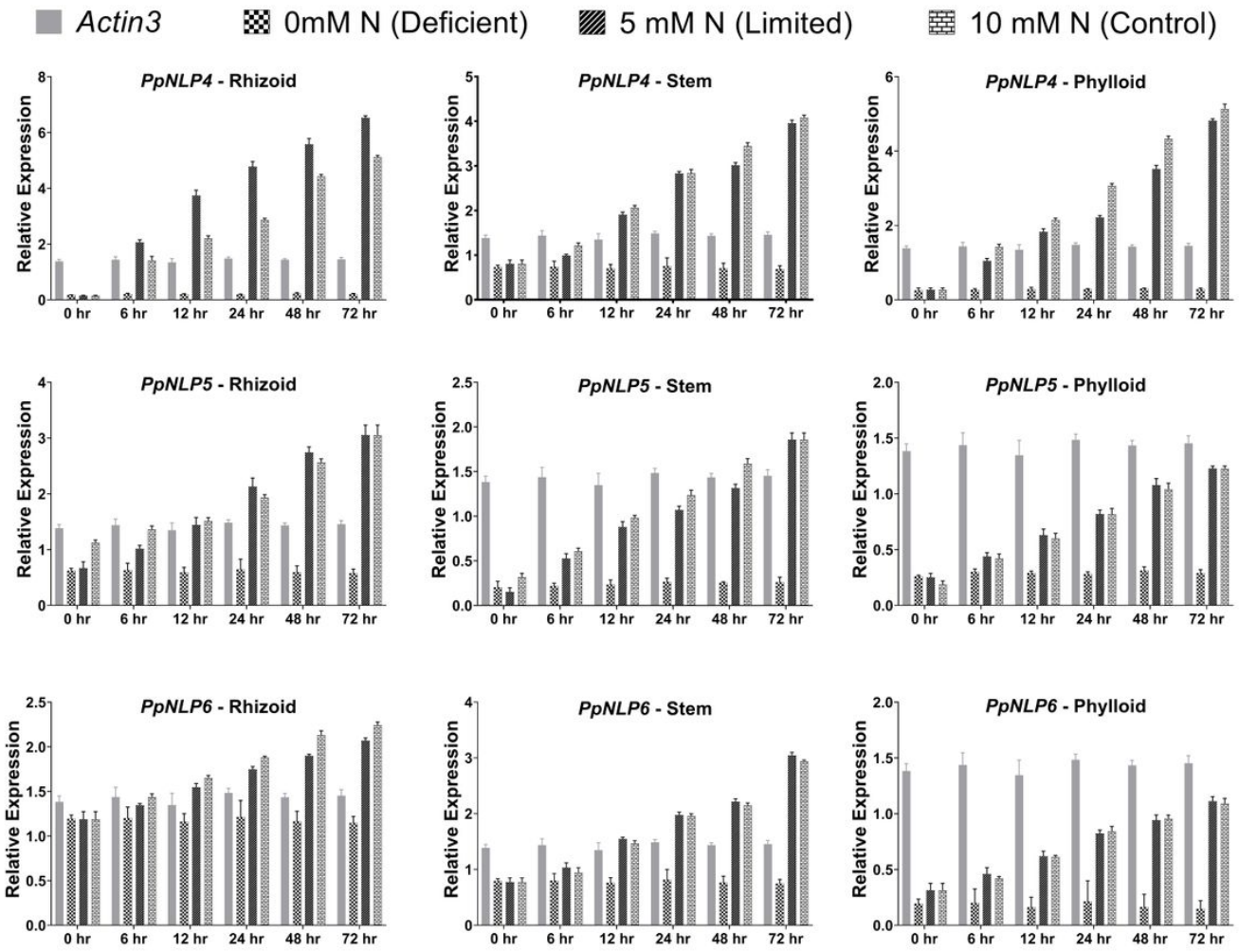


Figure 8

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 (Figure 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalInformationTables.xlsx