

RESEARCH ARTICLE

# Computational Identification Raises a Riddle for Distribution of Putative NACHT NTPases in the Genome of Early Green Plants

Preeti Arya<sup>1,2</sup>, Vishal Acharya<sup>1,2\*</sup>

**1** Functional Genomics and Complex System Lab, Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Council of Scientific and Industrial Research, Palampur- 176061, Himachal Pradesh, India, **2** Academy of Scientific and Innovative Research (AcSIR), CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT) Campus, Palampur, Himachal Pradesh, India

\* [vishal@ihbt.res.in](mailto:vishal@ihbt.res.in); [acharya.vishalacharya@gmail.com](mailto:acharya.vishalacharya@gmail.com)



**OPEN ACCESS**

**Citation:** Arya P, Acharya V (2016) Computational Identification Raises a Riddle for Distribution of Putative NACHT NTPases in the Genome of Early Green Plants. PLoS ONE 11(3): e0150634. doi:10.1371/journal.pone.0150634

**Editor:** Thomas A Kufer, University Hohenheim, GERMANY

**Received:** December 25, 2015

**Accepted:** February 17, 2016

**Published:** March 1, 2016

**Copyright:** © 2016 Arya, Acharya. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was financially supported by Council of Scientific and Industrial Research (CSIR), New Delhi, India in the form of CSIR network project "Computational Systems and Network Biology". Authors are thankful to Department of Biotechnology, Government of India for infrastructural support in the form of Bioinformatics Infrastructure Facility (BIF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

NACHT NTPases and AP-ATPases belongs to STAND (signal transduction ATPases with numerous domain) P-loop NTPase class, which are known to be involved in defense signaling pathways and apoptosis regulation. The AP-ATPases (also known as NB-ARC) and NACHT NTPases are widely spread throughout all kingdoms of life except in plants, where only AP-ATPases have been extensively studied in the scenario of plant defense response against pathogen invasion and in hypersensitive response (HR). In the present study, we have employed a genome-wide survey (using stringent computational analysis) of 67 diverse organisms viz., archaeobacteria, cyanobacteria, fungi, animalia and plantae to revisit the evolutionary history of these two STAND P-loop NTPases. This analysis divulged the presence of NACHT NTPases in the early green plants (green algae and the lycophyte) which had not been previously reported. These NACHT NTPases were known to be involved in diverse functional activities such as transcription regulation in addition to the defense signaling cascades depending on the domain association. In *Chlamydomonas reinhardtii*, a green algae, WD40 repeats found to be at the carboxyl-terminus of NACHT NTPases suggest probable role in apoptosis regulation. Moreover, the genome of *Selaginella moellendorffii*, an extant lycophyte, intriguingly shows the considerable number of both AP-ATPases and NACHT NTPases in contrast to a large repertoire of AP-ATPases in plants and emerge as an important node in the evolutionary tree of life. The large complement of AP-ATPases overtakes the function of NACHT NTPases and plausible reason behind the absence of the later in the plant lineages. The presence of NACHT NTPases in the early green plants and phyletic patterns results from this study raises a quandary for the distribution of this STAND P-loop NTPase with the apparent horizontal gene transfer from cyanobacteria.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

About 470 to 500 million years ago (Ordovician period), the ancestors of modern flowering plants conquered and spread to almost of all the terrestrial habitat [1–4]; this time event led to the three major groups of land plants evolved: bryophytes (liverworts, hornworts and mosses), pteridophytes (lycophytes and monilophytes) and spermatophytes (seed-producing plants) with the latter dominating the nowadays habitats. Despite differences in the natural habitat and system complexity, a large number of similar features were found (as evidenced by similarities in pigmentation, cell-wall chemistry, biochemistry, and method of cell division) between early green plants (lycophyte, bryophyte and green algae) and modern land plants that suggested they shared common ancestor [5]. Besides, *Selaginella moellendorffii*, a spikemoss (having certain characteristic features that are also shared among land plants) is considered as a model organism for the comparative genomic studies [6] and considered as an important node in the evolution of plant life. The evolution is itself very interesting, associated with change, for the adaptation or survival of entities against the shifting environmental conditions. Therefore, with a rapid advancement in the next generation sequencing data, leading to reconstruction of the phylogenetic relationships among biological macromolecules, will be essential to identifying the innovations underlying the diversity of life.

The AP-ATPases and NACHT-NTPases are the members of signal transduction ATPases with numerous domain (STAND) P-loop NTPase class, play a crucial role in the signaling cascade of apoptosis and defense response against pathogens [7,8]. The STAND P-loop NTPase, a class of P-loop NTPases (phosphate-binding proteins), firstly defined by Leipe et al. [8], contain multiple domains including enzymatic (engaged in several signaling cascades), DNA or protein binding and superstructure-forming repeats (viz Toll-like Interleukin (TIR), Leucine Rich Repeat (LRR)) at carboxy-terminus or amino-terminus. The AP-ATPases are shared by animal apoptosis regulators Apaf-1 (human regulator of apoptosis), CED-4 (cell death protein 4 of *Ceanorhabditis elegans*) and plant disease resistance (R) proteins. These AP-ATPases are known to be involved in binding and hydrolysis of ATP as molecular function in defense response pathways and abbreviated as NB-ARC where NB stands for nucleotide binding; ARC for Apaf1, resistance gene (R) of plants and CED-4 [8,9]. In NACHT NTPases, the NACHT stands for NAIP (Neuronal apoptosis inhibitor protein), C2TA (MHC class 2 transcription activator), HET-E (Incompatibility locus protein from *Podospora anserina*) and TP1 (telomerase-associated protein) while the “NTPases” term used for nucleotide binding protein fold [10]. The NACHT NTPase can bind to either GTP or ATP, with preference for GTP over ATP [8]. In addition to defense response and apoptosis regulation, NACHT NTPases, are also engaged in the diverse functional activities such as transcription regulation in bacteria [8] and are widely spread in all kingdoms of life except plants.

The structure topology of STAND P-loop NTPase class was characterized by a central  $\alpha\beta$  fold containing regulatory units [11,12]. This fold contains two well-known conserved motifs: Walker A/P-loop/kinase-1a and Walker B/kinase-2 motif involved in the nucleotide-binding and hydrolysis. Both AP-ATPase and NACHT NTPase also contains this nucleotide-binding region that consists of several conserved motifs including Walker-A and Walker-B motif (ATPase/GTPase specific P-loop and  $Mg^{++}$ -binding site respectively). However, the region of Walker-B motif was distinct in those two sister families that might be responsible for their differential biochemical functional activity [13].

The AP-ATPase (NB-ARC) is known to be present in diverse organisms ranging from bacteria to eukaryotes whereas NACHT NTPase is well reported in bacteria, fungi and animalia; however, it is absent in plants. In this study, for the first time, our stringent computational analysis has shown the presence of putative NACHT NTPases in early green plants: lycophyte

(*Selaginella moellendorffii*) and green algae (*Chlamydomonas reinhardtii* and *Coccomyxa subellipsoidea*). With these observations, we get insight into the evolutionary history of both STAND P-loop NTPases involved in defense response and apoptosis regulation.

## Materials and Methods

### Sequence Retrieval

The whole protein sequence dataset of 40 different plant species, including 32 land plants and six green algae were retrieved from publicly available Phytozome v9.0 (<http://www.phytozome.net/>, last accessed: Feb, 2013; [14]). The protein sequences of six archaeobacteria, three bacteria, three cyanobacteria, three protozoa, five fungi and seven species from animalia were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/protein>, last accessed: Feb, 2013). The Hidden Markov Model (HMM) profile of NACHT (PF05729.7) retrieved from Pfam [15] were used to screen the whole protein dataset for the identification of NACHT containing protein candidates in each organism by means of HMMER v3.1 (hmmsearch program) [16] with an e-value cutoff of 1e-04. A similar approach was carried out to find NB-ARC protein candidates in plants using HMM profile of NB-ARC domain (PF00931.17). The identified candidates were further confirmed for the presence of both NTPases (NACHT and NB-ARC) using PfamScan program with options `-e_seq:1e-04`; `-e_dom:1e-04` and `-clan_overlap` where “-clan\_overlap” option shows the information regarding overlapping domains belong to the same clan or family.

### Multiple Sequence Alignment (MSA) and Phylogenetic Analysis

The NACHT and NB-ARC NTPases specific amino acid sequences were extracted using in-house scripts and EMBOSS 6.6 (extractseq program) [17]. The extracted NACHT specific amino acid sequences were aligned using ClustalW v2.1 with default parameters [18]. The alignments were further employed for construction of the phylogenetic tree using RAxML version 7.2.8-ALPHA [19]. For finding the best empirical substitution model, we used ProtTest 3.4 [20] software package. The JTT +G+F was observed to be the best substitution model with respect to all four used statistical criteria—Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (AICc) and Decision Theory (DT) Criterion [21]—with confidence interval of 100.0. The phylogeny was then inferred through Maximum likelihood (ML) statistical method with gamma model of rate heterogeneity and JTT matrix. GAMMA model parameters were estimated up to an accuracy of 0.1000000000 Log Likelihood units. The bootstrap analysis of 100 replicates was applied to evaluate the topological robustness of phylogenetic tree. MEGA v5.2.2 were used to visualize the phylogenetic tree [22].

In the case study of *S. moellendorffii*, MEGA v5.2.2 was used to evaluate the evolutionary patterns of NACHT and NB-ARC through neighbor joining method. The evolutionary distances were computed using the Poisson correction method. The topological robustness for each branch of phylogenetic tree generated was assessed by bootstrap analysis with 1000 replicates. Subsequently, a final consensus tree was constructed with a 50% cutoff. The multiple sequence alignment of NACHT and NB-ARC specific protein sequences were visualized using ClustalX v2.1 [23] and AliView v1.14 [24].

### Analysis of the conserved motifs

The protein sequences of NACHT and NB-ARC in *S. moellendorffii* were analyzed for the prediction of conserved potential patterns by employing the MEME V4.9.0 (Multiple Expectation

Maximization for Motif Elicitation) [25]. The MEME analysis was done with a set of parameters i.e., minimum width: 8; maximum width: 20; maximum number of motifs: 20. The expected values and iterative cycles were set as default by MEME.

### *In silico* Structural comparison

The protein sequence of NACHT candidates of *S. moellendorffii* were used to build a model by means of I-TASSER web server which are based on threading approach [26]. The best model was selected on the basis of c-score (confidence score) calculated by I-TASSER to estimate the quality of predicted models. It is typically in the range of (-5,2), where a high value correlates with high confidence towards quality of model and vice-versa. The TM-score predicted by I-TASSER was also used for qualitative analysis of topology between two structure where a value of more than 0.5 infers the correct topology of predicted structure. TM-align program was implicated to compare protein structure [27] and PyMOL was employed for visualization of aligned structure [28].

## Results and Discussion

### Unusual presence of putative NACHT NTPases in early green plants

Recent genome-wide studies reported the presence of both types of STAND P-loop NTPases (NACHT and NB-ARC) in the diverse species. In this analysis, we have identified the presence of NACHT and NB-ARC NTPase in the genome of 67 organisms from diverse taxa (Fig 1; S1 Table). Fascinatingly, we found the presence of NACHT NTPase in the genome of green plants in addition to other eukaryotic species that was not reported previously and caught our attention for revisiting phylogenetic history. Therefore, to investigate the evolutionary relatedness of these NACHT NTPases, we performed multiple sequence alignment and reconstructed their phylogeny. From the phylogenetic analysis, it was observed that NACHT in plants grouped discretely (Fig 2); on the other hand, NACHT of archaeobacteria, cyanobacteria, bacteria, protozoans, fungi and animalia were clustered separately where NACHT of few plant species were also making a clan which was beyond our expectation (Fig 2).

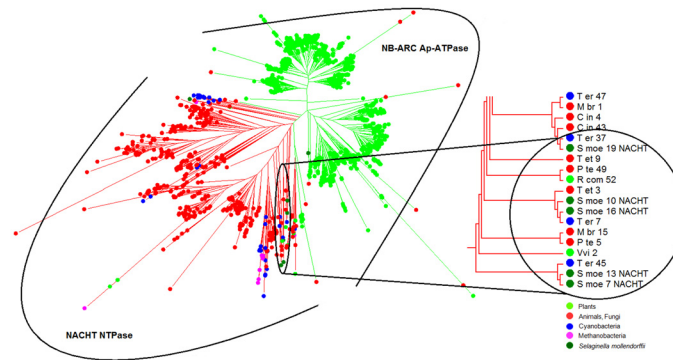
Later on analyzing the pfam results of NACHT specific sequences, we found that NACHT in most of the plants were overlapped with NB-ARC. However, the NACHT of *S. moellendorffii*, *C. reinhardtii*, and *C. subellipsoidea* were not observed to be overlapped with NB-ARC. Here, we used the term "NACHT NTPase" specific for NACHT to distinguish them from "NACHT overlap with NB-ARC". Further sequence analysis of "NACHT overlap with NB-ARC" and few other STAND P-loop NTPase from other plant species form a clan within the NACHT NTPase group revealed the presence of NB-ARC specific sequences, were therefore considered as AP-ATPases (NB-ARC).

In addition, one NACHT NTPase of short length but with well-conserved motif specific to this NTPase and significant e-value ( $4.3e-05$ ) was found in *Physcomitrella patens*. The NACHT protein candidates were also identified in several green algal species (*C. reinhardtii* (4) and *C. subellipsoidea* (2)) (Table 1). To examine the NACHT NTPase at conserved amino acid residue level, we performed the MSAs with NACHTs from *S. moellendorffii*, *C. reinhardtii*, *C. subellipsoidea*, *P. patens*; NLRC4 NACHT (NP\_001028539) were used as a NACHT representative. This analysis revealed the conservation of Walker-A and Walker-B motifs specific to NACHT NTPase (Fig 3). Whether so found NACHT is the sequence artifact, we also carried out blast search of all NACHT NTPases of the early green plants. The NACHT of *S. moellendorffii* were found mostly sharing a range of 15–20% sequence identity with well-known NACHT of human and mouse; green algae were found to share 21–35% similarity with NACHT of *Nostoc sp* PCC 7120 (S2 Table).

	Species	NB-ARC	NACHT
Archaeobacteria	<i>Methanosarcina acetivorans</i>	-	+
	<i>Methanosarcina barkeri</i>	+	+
	<i>Methanospirillum hungatei</i>	-	-
	<i>Methanocaldococcus jannaschii</i>	-	-
	<i>Natronomonas pharaonis</i>	-	-
	<i>Sulfolobus tokodaii</i>	+	-
Bacteria	<i>Escherichia coli</i>	-	+
	<i>Rhodococcus opacus</i>	+	+
	<i>Streptomyces coelicolor</i>	+	+
Cyanobacteria	<i>Nostoc punctiforme</i>	+	+
	<i>Synechococcus sp</i>	+	+
	<i>Trichodesmium erythraeum</i>	+	+
Protozoa	<i>Monosiga brevicollis</i>	-	+
	<i>Paramecium tetraurelia</i>	+	+
	<i>Tetrahymena thermophila</i>	-	+
Fungi	<i>Aspergillus terreus</i>	+	+
	<i>Chaetomium globosum</i>	+	+
	<i>Coprinopsis cinerea</i>	+	+
	<i>Magnaporthe grisea</i>	+	+
	<i>Saccharomyces cerevisiae</i>	-	-
Animalia	<i>Branchiostoma floridae</i>	+	+
	<i>Ciona intestinalis</i>	+	+
	<i>Homo sapiens</i>	+	+
	<i>Mus musculus</i>	+	+
	<i>Strongylocentrotus purpuratus</i>	+	+
	<i>Caenorhabditis elegans</i>	+	+
	<i>Nematostella vectensis</i>	+	+
Chlorophyte	<i>Chlamydomonas reinhardtii</i>	-	+
	<i>Coccomyxa subellipsoidea</i>	-	+
	<i>Micromonas pusilla CCMP1545</i>	-	-
	<i>Micromonas pusilla RCC299</i>	-	-
	<i>Ostrecoccus lucimarinus</i>	-	-
	<i>Volvox carteri</i>	-	-
Bryophyte	<i>Physcomitrella patens</i>	+	+
Lycophyte	<i>Selaginella moellendorffii</i>	+	+
Pteridophyte <sup>#</sup>			
Gymnosperm	<i>Pinus monti</i> <sup>*</sup>	+	?
Angiosperm	<i>Aquilegia coerulea</i>	+	-
	<i>Arabidopsis lyrata</i>	+	-
	<i>Arabidopsis thaliana</i>	+	-
	<i>Brachypodium distachyon</i>	+	-
	<i>Citrus clementia</i>	+	-
	<i>Carica papaya</i>	+	-
	<i>Capsella rubella</i>	+	-
	<i>Cucumis sativus</i>	+	-
	<i>Citrus sinensis</i>	+	-
	<i>Eucalyptus grandis</i>	+	-
	<i>Fragaria vesca</i>	+	-
	<i>Glycine max</i>	+	-
	<i>Gossypium raimondii</i>	+	-
	<i>Linum usitatissimum</i>	+	-
	<i>Malus domestica</i>	+	-
	<i>Manihot esculenta</i>	+	-
	<i>Mimulus guttatus v1.1</i>	+	-
	<i>Medicago truncatula</i>	+	-
	<i>Prunus persica</i>	+	-
	<i>Populus trichocarpa</i>	+	-
	<i>Phaseolus vulgaris</i>	+	-
	<i>Ricinus communis</i>	+	-
	<i>Solanum lycopersicum</i>	+	-
	<i>Solanum tuberosum</i>	+	-
	<i>Theobroma cacao</i>	+	-
	<i>Thellungiella halophila</i>	+	-
	<i>Vitis vinifera</i>	+	-
<i>Oryza sativa</i>	+	-	
<i>Panicum virgatum</i>	+	-	
<i>Sorghum bicolor</i>	+	-	
<i>Setaria italica</i>	+	-	
<i>Zea mays</i>	+	-	

**Fig 1. Genome-wide survey of 67 organisms for NB-ARC and NACHT NTPases.** The genome-wide survey of 67 diverse organisms shows the presence (+) or absence (-) of NB-ARC and NACHT NTPases in respective organism.

doi:10.1371/journal.pone.0150634.g001



**Fig 2. Phylogenetic tree of NB-ARC and NACHT NTPases of three major kingdoms of life.** The NB-ARC and NACHT NTPases present in the diverse organisms were aligned by ClustalW v2.1 and phylogenetic tree was constructed by employing RAxML version 7.2.8-ALPHA. The phylogenetic tree was visualized using MEGA v5.2.2. Green and red branches denote the NB-ARC AP-ATPase and NACHT NTPase group respectively. The phylogenetic analysis inferred that few members of STAND P-loop NTPase of plants were making a clan in NACHT NTPase group, are putative NACHT NTPase of plants.

doi:10.1371/journal.pone.0150634.g002

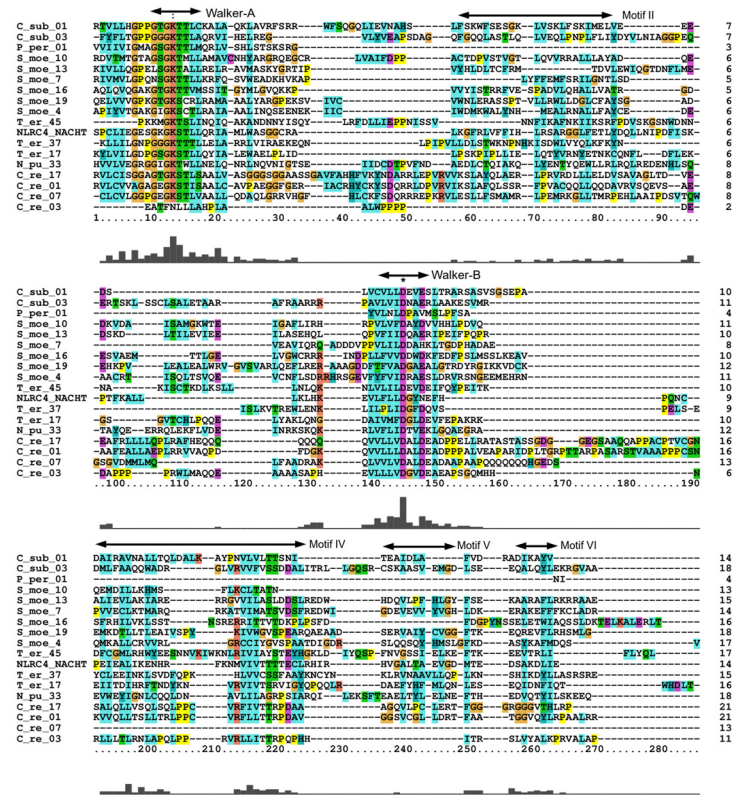
To scrutinize the evolutionary fate of both STAND P-loop NTPases, identified NB-ARC and NACHT sequences were analyzed with molecular phylogenetic methods. While investigating this phylogeny, we have discarded some of the NB-ARC from our analysis if either of Walker-A or Walker-B motif were found to be totally absent. We used NLRC4 as a representative of NACHT (accession no: NP\_001028539) and APAF1 (accession no: ABQ59028) as a NB-ARC representative for carrying out the phylogenetic analysis. The NACHT of cyanobacteria (that were grouped with *S. moellendorffii* NACHTs of NACHT NTPase group; Fig 2) were also included in this phylogenetic comparison. Interestingly, this evolutionary analysis suggests

**Table 1. List of NACHT domain containing protein sequences in early green plants.**

Organism Name/ Accession Number	Name used in our study	Division	E-value	Length (Amino acids)
<b><i>Selaginella moellendorffii</i></b>				
<b>Lycophyte</b>				
407161 PACid:15416689	S_moe_4		6e-05	172
417146 PACid:15401511	S_moe_7		1.2e-05	142
422201 PACid:15417115	S_moe_10		9.1e-07	129
424082 PACid:15421485	S_moe_13		1.1e-05	158
425148 PACid:15402815	S_moe_16		6.9e-05	168
444094 PACid:15413793	S_moe_19		4.1e-05	184
<b><i>Chlamydomonas reinhardtii</i></b>				
<b>Green algae</b>				
Cre01.g039500.t1.2 PACid:27578790	C_re_01		5e-05	217
Cre03.g152550.t1.3 PACid:27576858*	C_re_03		1e-06	111
Cre07.g350451.t1.2 PACid:27564305	C_re_07		3.7e-05	133
Cre17.g723450.t1.3 PACid:27572133	C_re_17		2.7e-07	213
<b><i>Coccomyxa subellipsoidea</i></b>				
<b>Green algae</b>				
14805 PACid:27390995	C_sub_01		6.6e-06	147
44528 PACid:27394369	C_sub_03		1.4e-06	181
<b><i>Physcomitrella patens</i></b>				
<b>Bryophyte</b>				
Pp1s114_147V6.1 PACid:18037264*	P_per_01		4.3e-05	48

\*Highly diverged /truncated putative NACHT NTPase

doi:10.1371/journal.pone.0150634.t001

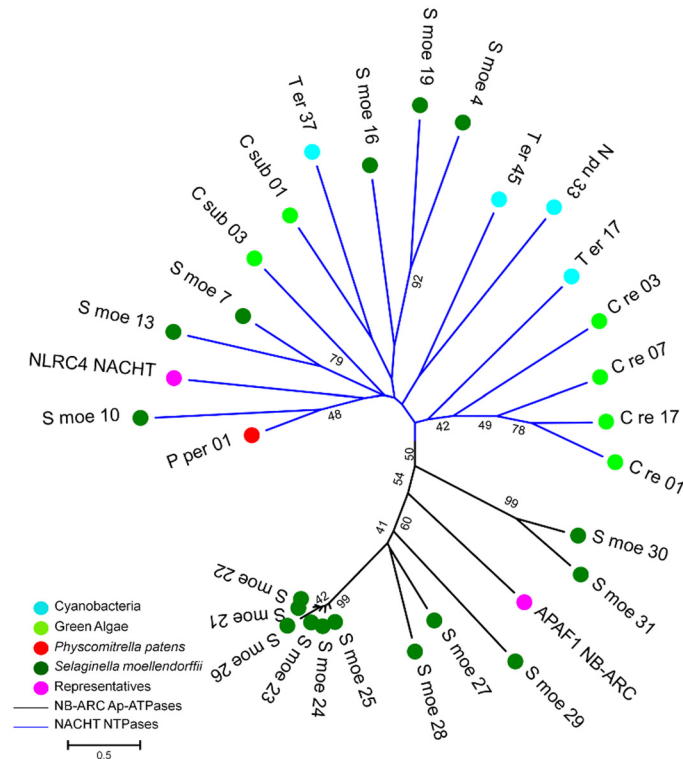


**Fig 3. Multiple Sequence Alignment of NACHT NTPases.** Multiple sequence alignment (MSA) for NACHT of *Selaginella moellendorffii*, *Physcomitrella patens*, green algae, cyanobacteria and NLRC4 showing the conserved region at the site of Walker-A and Walker-B and other motifs [10]. (The S\_moe\_\* denotes NACHT NTPases in *S. moellendorffii*, C\_re\_\* for *C. reinhardtii* and C\_sub\_\* for *C. subellipsoidea* where \* represents any integer number). More than 80% consensus background coloring of amino acid residues reflects as follows: hydrophobic residues (ACFILMVVWY), aromatic residues (FHVWY), and aliphatic residues (VIL) are shaded cyan/blue; polar residues (STQN) are colored green; acidic residues (DE) are colored magenta; basic residues (KR) are colored red; Glycine (G) is shaded with mustard color; and Proline (P) is colored yellow.

doi:10.1371/journal.pone.0150634.g003

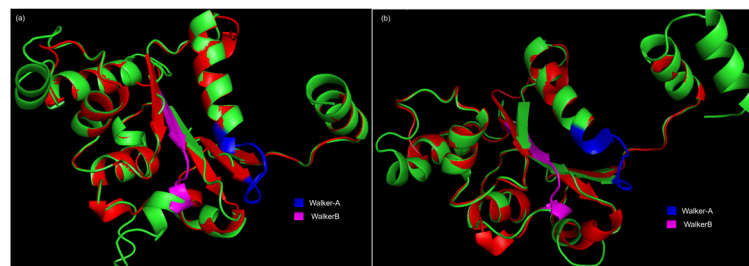
that both NACHT and NB-ARC arise as an independent entity during the evolution of STAND P-loop NTPases (Fig 4), further supporting their independent evolution [29].

To further evaluate NACHT at structural level, *in silico* fold recognition method was used to predict the structure for identified putative NACHT NTPase in *S. moellendorffii*. As expected, *in silico* predicted structure of *S. moellendorffii* NACHTs exhibit the same topological arrangements as that of experimentally derived structure of (PDB ID: 4KXF, chain K). The structure of NACHT comprises alternating alpha helix and beta sheets in the core region of Walker-A and Walker-B motifs. Interestingly, the structure of NLRC4 NACHT and *S. moellendorffii* NACHT superimpose well with a RMSD value of 0.79 Å (from NLRC4 NACHT and 407161|PACid:15416689) and 0.74 Å (from NLRC4 NACHT and 425148|PACid:15402815) (Fig 5) for Cα traces of the entire chain, despite relatively low sequence identity (< 30% sequence identity). To investigate the conserved residues responsible for functional activity, NACHT sequences of *S. moellendorffii* were fed to MEME. The consensus motifs were found to be G [P/A]KG[S/T]GK[S/T] and hhhhD[RQG]AE respectively, which was strikingly similar to well-known NACHT of NLRC4. And as expected, in well-known NB-ARC architecture, Walker-A and Walker-B motifs of *S. moellendorffii*, similarly showed GA[GS]GAGKT and hhhhDDVW.



**Fig 4. Unrooted phylogenetic tree of NB-ARC and NACHT NTPases.** The members of NACHT and NB-ARC NTPases of *S. moellendorffii* were aligned with NACHT NTPases of green algae, cyanobacteria and *Physcomitrella patens* using ClustalW v2.1 and a phylogenetic tree was constructed by employing MEGA v5.2.2. The NLRC4 NACHT was taken as representative of NACHT and APAF1 NB-ARC as NB-ARC representative. (The S\_moe\_\* denotes NACHT NTPases and NB-ARC AP-ATPases in *S. moellendorffii*, C\_re\_\* for *C. reinhardtii* and C\_sub\_\* for *C. subellipsoidea* where \* represents any integer number.)

doi:10.1371/journal.pone.0150634.g004



**Fig 5. Structural comparison of NACHT of *Selaginella moellendorffii* with NLRC4 NACHT.** The superposition of NACHT of *S. moellendorffii* with NLRC4 NACHT using TM-align. (a) NLRC4 NACHT and 407161|PACid:15416689 (b) NLRC4 NACHT and 425148|PACid:15402815. Red colour cartoon shows NLRC4 NACHT and green colour cartoons show *S. moellendorffii* NACHT domains. The residue range for Walker-A and Walker-B are 169–176 and 244–252 (NLRC4 NACHT); 82–89 and 169–177 (407161|PACid:15416689); 161–168 and 238–245 (425148|PACid:15402815) respectively. The quality of model was assessed on basis of c-score and TM-score. The predicted value of c-score and TM-score by I-TASSER are -1.36 and 0.55+0.15 for 407161|PACid:15416689 and are -2.74 and 0.40+0.13 for 425148|PACid:15402815. The root mean square deviation (RMSD) calculated by TM-align are 0.79 and 0.74 for 407161|PACid:15416689 and 425148|PACid:15402815 respectively.

doi:10.1371/journal.pone.0150634.g005

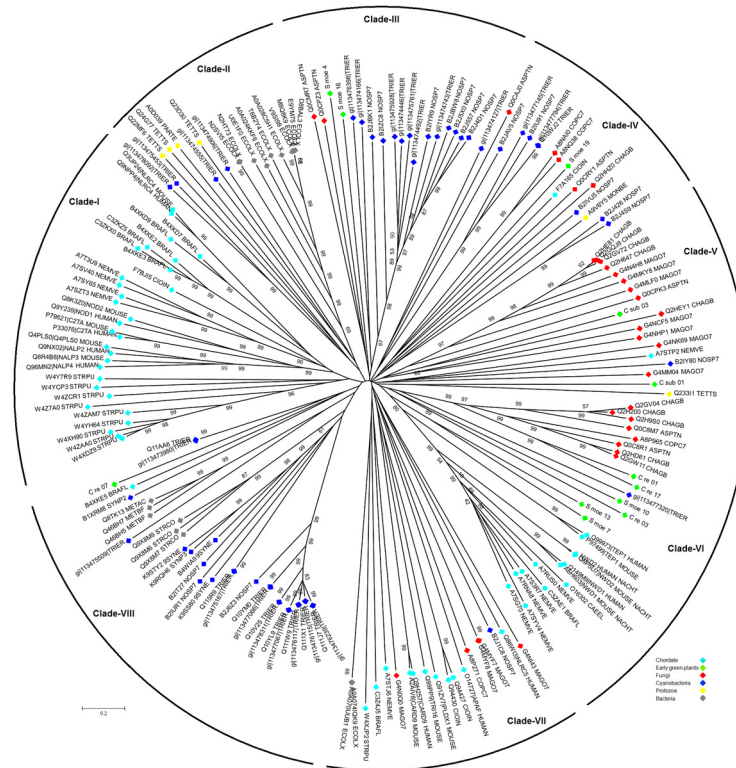


Hence, one may presume the presence of another STAND P-loop NTPase as NACHT NTPase in this lycophyte plant. This inspection also inferred and supported that the divergence at Walker-B motif might have happened between these two NTPases which were responsible for the variation of biochemical activity: NTP binding and hydrolysis [13, 30].

The NACHT NTPase perform diverse function in addition to signal transduction in apoptosis because of different type of domain combinations (CARD-NACHT-WD40, PYD--NACHT-LRR, NACHT-WD40 or NACHT-LRR). However, we did not found any type of domain association in recently observed *S. moellendorffii* NACHT NTPases. Interestingly, we reported for the first time, the WD40 association at the C-terminus in the NACHT NTPases of chlorophyte, *C. reinhardtii*. The WD40 are the short motif of tryptophan and aspartic acid with 4–16 repeating units and involved in signaling pathways of programmed cell death (PCD) in eukaryotes [31, 32]. Furthermore, the delta-blast searches of NACHT NTPase of *C. reinhardtii* also showed the top similarity hits with a significant e-value for predicted NACHT NTPase (NACHT-WD40 protein) of *Nostoc* sp. PCC 7120 (sp|Q8YRI1.1|YY46\_NOSS1) (S1 Table). The strikingly significant similarities of NACHT of *C. reinhardtii* with NACHT of *Nostoc* sp PCC 7120 may suggest their close evolutionary relationships among these STAND P-loop NTPases. The resulting observation is an indicative of the presence of NACHT NTPases in the early green plants which may be unusual regarding their unknown function in the plant genomes.

### Evolutionary scheme for horizontal gene transfer of NACHT NTPase genes in early green plants

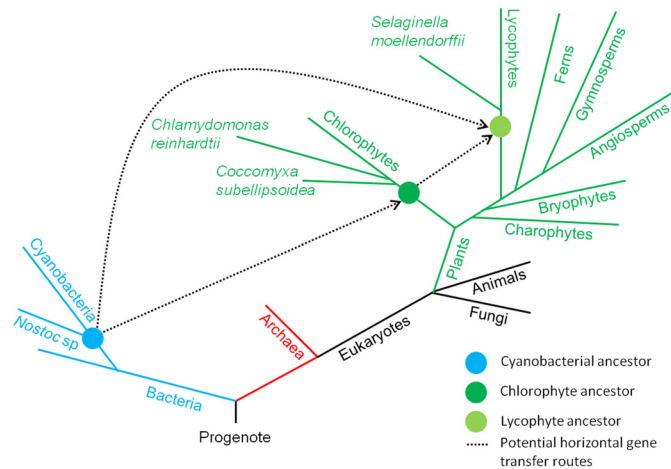
The Ordovician period was heralded with biggest mass extinction lead to the loss of life. At that time period, the green algae was likely to be dominated flora and then first terrestrial plant life have been evolved. The NACHT NTPases were involved in signaling cascades of apoptosis/programmed cell death (PCD) which compliment the cell development and differentiation in the diverse organisms of three major kingdoms of life. These NACHT NTPases has been later lost in plant lineages however still present in the genome of early green plants (*S. moellendorffii*, *C. reinhardtii* and *C. subellipsoidea*) raises a quandary for their distribution. To inspect the phenomenon behind the acquisition of NACHT NTPases in early green plants, we performed the phylogenetic comparison of identified NACHT NTPases in the present study and some well-known NACHT NTPases from different organisms. The NACHT of early green plants were grouped either with NACHT of cyanobacteria or fungi in clades III, IV, V, VI, and VIII (Fig 6; S3 Table). However, the delta-blast results revealed the top hits with NACHT NTPase of *Nostoc* sp for NACHT in *C. reinhardtii*. The clan of NACHT of green algae, *C. reinhardtii* grouped with cyanobacterial NACHT and further the blast search result showing the highest similarity with *Nostoc* sp PCC 7120 infers that the DNA for NACHT in green algae may have been transferred from the ancestors of extant cyanobacteria. The transmission of genetic information in progeny from parents usually occurs by vertical gene transfer (VGT) by means of sexual and asexual reproduction in the complex multicellular eukaryotes. However, recent reports assessed the importance of horizontal gene transfer (HGT) in the evolution of eukaryotic genomes which was considered to be predominant in prokaryotes [33, 34]. HGT, often known as lateral gene transfer (LGT), is the process of introducing the novel genes between evolutionary unrelated species [35], thus facilitating phenotypic variation and adaptation to the changing environments. Incongruent relationships observed in the molecular phylogenetic trees is the general method used to detect the HGT events in unrelated species [36]. During the evolution, plants gradually developed complexity in their biological systems to survive themselves to the shifting environments where HGT also played a significant role in this adaption [37]. Previous



**Fig 6. Unrooted phylogenetic tree of NACHT NTPases from diverse organisms of three major kingdoms of life.** The multiple sequence alignment was performed for putative NACHT NTPases in early green plants (green algae, bryophyte, and lycophyte) and well-known NACHT NTPases in other organisms including bacteria, archaeobacteria, cyanobacteria, protozoa, fungi, and chordates and phylogenetic tree was constructed using MEGA v5.2.2. The phylogenetic analysis shows that NACHT NTPases of early green plant were grouped in a clan of NACHT NTPases of cyanobacteria or fungi (in clades III, IV, V, VI, and VII). Moreover, results of delta-blast search of NACHT NTPase in *C. reinhardtii* revealed the the top hits with *Nostoc* sp PCC 7120. From phyletic patterns and delta-blast search results from the present study, we infer that the DNA for NACHT in green algae may have been transferred from the ancestors of extant cyanobacteria.

doi:10.1371/journal.pone.0150634.g006

reports have also documented that cyanobacterial invasion and subsequent HGT is the likely resource for the evolution of diverse types of proteins in the modern flowering plants [38–41]. Moreover, HGT served as a frequent scenario for integration of microorganismal genes into the plant nuclear genomes, speculated in a few phylogenetic studies of thousands protein-encoding genes [42–46]. In addition, previous studies also divulged the cyanobacteria as a well-known symbionts and cyanobionts for wide spectrum of fungi and plants (bryophytes, petri-dophytes, and spermatophytes [47]. Based on the current insights on phylogeny of NACHT NTPases, close homology relationship of NACHT NTPases of *C. reinhardtii* with NACHT NTPases of *Nostoc* sp. (S2 Table) and absence in other plant species suggested that NACHT NTPase might have been acquired in early green plant lineages through HGT events (Fig 7). The plausible reason for the absence of NACHT NTPases in other plant species is that their function might have been overtaken by a large repertoire of NB-ARC ATPases. The observation of putative NACHT NTPases distribution in early green plants adds a probable item to the list of functionalities acquired through HGT in green plants from the ancestors of extant cyanobacteria. However, further investigation will be needed regarding current lack of data to assess the HGT scenario of NACHT NTPases in early green plants with greater clarity.



**Fig 7. Hypothetical pictorial representation of horizontal gene transfer of NACHT NTPase genes in early green plants.** Probable horizontal gene transfer (HGT) of NACHT NTPase from cyanobacterial ancestor to early green plant lineages. The NACHT NTPase occurs in all three major kingdoms of life that arose from single progenote and may have acquired in single-celled green alga (chlorophytes) through HGT from an ancestor of cyanobacteria (*Nostoc* sp.).

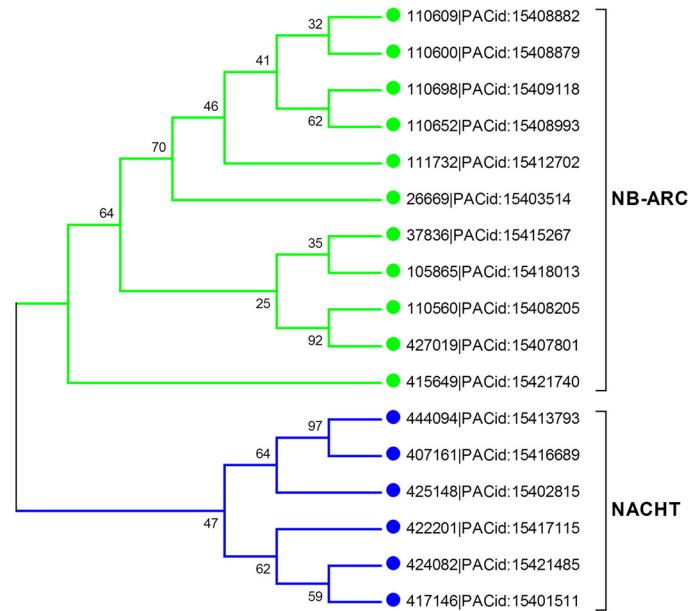
doi:10.1371/journal.pone.0150634.g007

### *Selaginella moellendorffii* as an important evolutionary node

*S. moellendorffii* is the member of the oldest living vascular plant lineage which lack true leaves and roots. It first appeared in fossil record 400 million years ago and considered as a model organism for comparative genomic studies [6]. The phylogenetic analysis of the identified AP-ATPases (NB-ARC) and NACHT-NTPases in this lycophyte divulges the two discrete clades specific for NB-ARC and NACHT respectively (Figs 4 and 8), suggesting the independent evolution supported by previous literature [29]. The sequence, motif analysis and structural considerations confirm the presence of NACHT in addition to NB-ARC NTPases in this ancient vascular plant. Moreover, the complement of NB-ARC and NACHT was found to be in commensurable number in comparison to the large repertoire of NB-ARC in other land plant species that make *S. moellendorffii* more interesting for studying the evolutionary tree of life.

Few reports documented more similarity for the organelle genome composition in *S. moellendorffii* with chlorophyceae in comparison with other land plants [5] whereas other studies also demonstrated the close relationship for the identified secondary metabolites in *S. moellendorffii*, a non-seed plant, with a highly advanced flowering plants [48]. In the case of *S. moellendorffii*, both NB-ARC and NACHT were found to be in comparable number (11 NB-ARC and 6 NACHT) whereas other land plants show a large repository for NB-ARC; for example, the bryophyte, *P. patens* shows a higher number of NB-ARC (88). This observation further led us to exemplify the significance of this extant lycophyte in the evolution of STAND P-loop NTPases and suggests *S. moellendorffii* as an important piece of the puzzle in understanding the evolution of land plants [6].

Our report of NACHT NTPase acquisition in early green plants raises a question: why it had never been pervaded in other green plants? We hypothesize that the absence of associated domains responsible for NACHT function is the most likely reason for the loss of NACHT NTPase in plants. Moreover, the large complement of NB-ARC in plants may overtake the function of NACHT, which is the another possible explanation behind the absence of NACHT NTPase in green plant species. The evolutionary analysis revealed that cyanobacteria may have acted as donors of genetic materials for more than thousand proteins in the modern flowering



**Fig 8. Phylogenetic tree of NB-ARC and NACHT NTPases in *Selaginella moellendorffii*.** The NACHT and NB-ARC of *S. moellendorffii* were aligned using ClustalW and phylogenetic tree was constructed using MEGA v5.2.2. The green lines represents the NB-ARC AP-ATPases and blue lines denotes NACHT NTPases proteins forming two separate clades. The members of both NACHT and NB-ARC NTPase grouped discretely supporting the independent evolution of these STAND P-loop NTPases.

doi:10.1371/journal.pone.0150634.g008

plants as result of HGT or endosymbiotic gene transfer [38–41]. Our finding of the close homology between NACHT NTPase in *C. reinhardtii* and *Nostoc sp* PCC 7120 proteins demonstrates that HGT may have played a decisive role in the distribution of the NACHT domain in the early green plants from the free-living cyanobacterial ancestor species. Most interestingly, the comparable number of both STAND P-loop NTPases, in the extant lycophyte, *S. moellendorffii*, supports their independent origin. This study states the independent origin of both STAND P-loop NTPases in eukaryotes and invasion of NACHT NTPases in early green plants with apparent HGT.

## Supporting Information

**S1 Table. List of protein sequence accession number used for phylogenetic analyses in the report from different species.**

(XLS)

**S2 Table. Blast hits of corresponding protein sequences of early green plant lineages.**

(XLS)

**S3 Table. List of NACHT NTPases used for phylogenetic analysis in Fig 6.**

(XLS)

## Acknowledgments

We are grateful to the Director, CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT), Palampur, India for providing necessary facilities. PA gratefully acknowledge University Grant Commission (UGC), Government of India for providing fellowship. This manuscript represents CSIR-IHBT communication number: 2396.

## Author Contributions

Conceived and designed the experiments: PA VA. Performed the experiments: PA. Analyzed the data: PA VA. Contributed reagents/materials/analysis tools: PA VA. Wrote the paper: PA VA.

## References

1. Sanderson MJ, Thorne JL, Wikström N, Bremer K. Molecular evidence on plant divergence times. *Am J Bot.* 2004; 91: 1656–65. doi: [10.3732/ajb.91.10.1656](https://doi.org/10.3732/ajb.91.10.1656) PMID: [21652315](https://pubmed.ncbi.nlm.nih.gov/21652315/)
2. Rubinstein C V, Gerrienne P, de la Puente GS, Astini RA, Steemans P. Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). *New Phytol.* 2010; 188: 365–9. doi: [10.1111/j.1469-8137.2010.03433.x](https://doi.org/10.1111/j.1469-8137.2010.03433.x) PMID: [20731783](https://pubmed.ncbi.nlm.nih.gov/20731783/)
3. Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LGG, et al. Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol Evol.* 2010; 2: 488–503. doi: [10.1093/gbe/evq032](https://doi.org/10.1093/gbe/evq032) PMID: [20644220](https://pubmed.ncbi.nlm.nih.gov/20644220/)
4. Wodniok S, Brinkmann H, Glöckner G, Heidel AJ, Philippe H, Melkonian M, et al. Origin of land plants: do conjugating green algae hold the key? *BMC Evol Biol.* 2011; 11: 104. doi: [10.1186/1471-2148-11-104](https://doi.org/10.1186/1471-2148-11-104) PMID: [21501468](https://pubmed.ncbi.nlm.nih.gov/21501468/)
5. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and Molecular Evolution of the Green Algae. *CRC Crit Rev Plant Sci.* Taylor & Francis Group; 2012; 31: 1–46. doi: [10.1080/07352689.2011.615705](https://doi.org/10.1080/07352689.2011.615705)
6. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, et al. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science.* 2011; 332: 960–3. doi: [10.1126/science.1203810](https://doi.org/10.1126/science.1203810) PMID: [21551031](https://pubmed.ncbi.nlm.nih.gov/21551031/)
7. Aravind L, Dixit VM, Koonin E V. The domains of death: evolution of the apoptosis machinery. *Trends Biochem Sci.* 1999; 24: 47–53. doi: [10.1016/S0968-0004\(98\)01341-3](https://doi.org/10.1016/S0968-0004(98)01341-3) PMID: [10098397](https://pubmed.ncbi.nlm.nih.gov/10098397/)
8. Leippe DD, Koonin EV, Aravind L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol.* 2004; 343: 1–28. doi: [10.1016/j.jmb.2004.08.023](https://doi.org/10.1016/j.jmb.2004.08.023) PMID: [15381417](https://pubmed.ncbi.nlm.nih.gov/15381417/)
9. van der Biezen EA, Jones JDG. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr Biol.* 1998; 8: R226–R228. doi: [10.1016/S0960-9822\(98\)70145-9](https://doi.org/10.1016/S0960-9822(98)70145-9) PMID: [9545207](https://pubmed.ncbi.nlm.nih.gov/9545207/)
10. Koonin EV., Aravind L. The NACHT family—a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends Biochem Sci.* 2000; 25: 223–224. doi: [10.1016/S0968-0004\(00\)01577-2](https://doi.org/10.1016/S0968-0004(00)01577-2) PMID: [10782090](https://pubmed.ncbi.nlm.nih.gov/10782090/)
11. Hanson PI, Whiteheart SW. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol.* Nature Publishing Group; 2005; 6: 519–29. doi: [10.1038/nrm1684](https://doi.org/10.1038/nrm1684)
12. Snider J, Houry WA. AAA+ proteins: diversity in function, similarity in structure. *Biochem Soc Trans.* 2008; 36: 72–77. doi: [10.1042/BST0360072](https://doi.org/10.1042/BST0360072) PMID: [18208389](https://pubmed.ncbi.nlm.nih.gov/18208389/)
13. Rairdan G, Moffett P. Brothers in arms? Common and contrasting themes in pathogen perception by plant NB-LRR and animal NACHT-LRR proteins. *Microbes Infect.* 2007; 9: 677–86. doi: [10.1016/j.micinf.2007.01.019](https://doi.org/10.1016/j.micinf.2007.01.019) PMID: [17379561](https://pubmed.ncbi.nlm.nih.gov/17379561/)
14. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012; 40: D1178–86. doi: [10.1093/nar/gkr944](https://doi.org/10.1093/nar/gkr944) PMID: [22110026](https://pubmed.ncbi.nlm.nih.gov/22110026/)
15. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz H-R, et al. The Pfam protein families database. *Nucleic Acids Res.* 2008; 36: D281–8. doi: [10.1093/nar/gkm960](https://doi.org/10.1093/nar/gkm960) PMID: [18039703](https://pubmed.ncbi.nlm.nih.gov/18039703/)
16. Eddy SR. HMMer user's guide (version 3.1). Profile hidden Markov models for biological sequence analysis. 2003. Available: <http://hmmer.janelia.org/>
17. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 2000; 16: 276–277. doi: [10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2) PMID: [10827456](https://pubmed.ncbi.nlm.nih.gov/10827456/)
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22: 4673–80. doi: [10.1093/nar/22.22.4673](https://doi.org/10.1093/nar/22.22.4673) PMID: [7984417](https://pubmed.ncbi.nlm.nih.gov/7984417/)

19. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22: 2688–90. doi: [10.1093/bioinformatics/btl446](https://doi.org/10.1093/bioinformatics/btl446) PMID: [16928733](https://pubmed.ncbi.nlm.nih.gov/16928733/)
20. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27: 1164–5. doi: [10.1093/bioinformatics/btr088](https://doi.org/10.1093/bioinformatics/btr088) PMID: [21335321](https://pubmed.ncbi.nlm.nih.gov/21335321/)
21. Minin V, Abdo Z, Joyce P, Sullivan J. Performance-Based Selection of Likelihood Models for Phylogeny Estimation. *Syst Biol*. 2003; 52: 674–683. doi: [10.1080/10635150390235494](https://doi.org/10.1080/10635150390235494) PMID: [14530134](https://pubmed.ncbi.nlm.nih.gov/14530134/)
22. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28: 2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
23. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23: 2947–8. doi: [10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404) PMID: [17846036](https://pubmed.ncbi.nlm.nih.gov/17846036/)
24. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014; 30: 3276–8. doi: [10.1093/bioinformatics/btu531](https://doi.org/10.1093/bioinformatics/btu531) PMID: [25095880](https://pubmed.ncbi.nlm.nih.gov/25095880/)
25. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*. 2006; 34: W369–73. doi: [10.1093/nar/gkl198](https://doi.org/10.1093/nar/gkl198) PMID: [16845028](https://pubmed.ncbi.nlm.nih.gov/16845028/)
26. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010; 5: 725–38. doi: [10.1038/nprot.2010.5](https://doi.org/10.1038/nprot.2010.5) PMID: [20360767](https://pubmed.ncbi.nlm.nih.gov/20360767/)
27. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005; 33: 2302–9. doi: [10.1093/nar/gki524](https://doi.org/10.1093/nar/gki524) PMID: [15849316](https://pubmed.ncbi.nlm.nih.gov/15849316/)
28. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010. Available: <https://www.pymol.org>
29. Yue J-X, Meyers BC, Chen J-Q, Tian D, Yang S. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol*. 2012; 193: 1049–63. doi: [10.1111/j.1469-8137.2011.04006.x](https://doi.org/10.1111/j.1469-8137.2011.04006.x) PMID: [22212278](https://pubmed.ncbi.nlm.nih.gov/22212278/)
30. MacDonald JA, Wijekoon CP, Liao K-C, Muruve DA. Biochemical and structural aspects of the ATP-binding domain in inflammasome-forming human NLRP proteins. *IUBMB Life*. 2013; 65: 851–62. doi: [10.1002/iub.1210](https://doi.org/10.1002/iub.1210) PMID: [24078393](https://pubmed.ncbi.nlm.nih.gov/24078393/)
31. Jiang Q, Qin S, Wu Q-Y. Genome-wide comparative analysis of metacaspases in unicellular and filamentous cyanobacteria. *BMC Genomics*. 2010; 11: 198–11. doi: [10.1186/1471-2164-11-198](https://doi.org/10.1186/1471-2164-11-198) PMID: [20334693](https://pubmed.ncbi.nlm.nih.gov/20334693/)
32. Hamada M, Shoguchi E, Shinzato C, Kawashima T, Miller DJ, Satoh N. The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Mol Biol Evol*. 2013; 30: 167–76. doi: [10.1093/molbev/mss213](https://doi.org/10.1093/molbev/mss213) PMID: [22936719](https://pubmed.ncbi.nlm.nih.gov/22936719/)
33. Lawrence JG. Gene transfer in bacteria: speciation without species? *Theor Popul Biol*. 2002; 61: 449–60. doi: [10.1006/tpbi.2002.1587](https://doi.org/10.1006/tpbi.2002.1587) PMID: [12167364](https://pubmed.ncbi.nlm.nih.gov/12167364/)
34. Fortune PM, Roulin A, Panaud O. Horizontal transfer of transposable elements in plants. *Commun Integr Biol*. 2008; 1: 74–7. doi: [10.4161/cib.1.1.6328](https://doi.org/10.4161/cib.1.1.6328) PMID: [19513203](https://pubmed.ncbi.nlm.nih.gov/19513203/)
35. Yue J, Hu X, Sun H, Yang Y, Huang J. Widespread impact of horizontal gene transfer on plant colonization of land. *Nat Commun*. Nature Publishing Group; 2012; 3: 1152. doi: [10.1038/ncomms2148](https://doi.org/10.1038/ncomms2148)
36. Bock R. The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci*. Elsevier; 2010; 15: 11–22. doi: [10.1016/j.tplants.2009.10.001](https://doi.org/10.1016/j.tplants.2009.10.001)
37. Graham LE, Cook ME, Busse JS. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci U S A*. 2000; 97: 4535–40. doi: [10.1073/pnas.97.9.4535](https://doi.org/10.1073/pnas.97.9.4535) PMID: [10781058](https://pubmed.ncbi.nlm.nih.gov/10781058/)
38. Rujan T, Martin W. How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends Genet*. 2001; 17: 113–20. doi: [10.1016/S0168-9525\(00\)02209-5](https://doi.org/10.1016/S0168-9525(00)02209-5) PMID: [11226586](https://pubmed.ncbi.nlm.nih.gov/11226586/)
39. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A*. 2002; 99: 12246–51. doi: [10.1073/pnas.182432999](https://doi.org/10.1073/pnas.182432999) PMID: [12218172](https://pubmed.ncbi.nlm.nih.gov/12218172/)
40. Raven JA, Allen JF. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol*. 2003; 4: 209. doi: [10.1186/gb-2003-4-3-209](https://doi.org/10.1186/gb-2003-4-3-209) PMID: [12620099](https://pubmed.ncbi.nlm.nih.gov/12620099/)
41. Xi Z, Bradley RK, Wurdack KJ, Wong K, Sugumaran M, Bombliés K, et al. Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics*. 2012; 13: 227. doi: [10.1186/1471-2164-13-227](https://doi.org/10.1186/1471-2164-13-227) PMID: [22681756](https://pubmed.ncbi.nlm.nih.gov/22681756/)

42. Syvanen M. Horizontal gene transfer: evidence and possible consequences. *Annu Rev Genet. Annual Reviews* 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303–0139, USA; 1994; 28: 237–61. doi: [10.1146/annurev.ge.28.120194.001321](https://doi.org/10.1146/annurev.ge.28.120194.001321)
43. Copley SD, Dhillon JK. Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes. *Genome Biol.* 2002; 3: research0025. PMID: [12049666](https://pubmed.ncbi.nlm.nih.gov/12049666/)
44. Emiliani G, Fondi M, Fani R, Gribaldo S. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol Direct.* 2009; 4: 7. PMID: [19220881](https://pubmed.ncbi.nlm.nih.gov/19220881/)
45. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell.* 2009; 21: 1897–911. doi: [10.1105/tpc.109.065805](https://doi.org/10.1105/tpc.109.065805) PMID: [19584142](https://pubmed.ncbi.nlm.nih.gov/19584142/)
46. Knie N, Polsakiewicz M, Knoop V. Horizontal gene transfer of chlamydial-like tRNA genes into early vascular plant mitochondria. *Mol Biol Evol.* 2015; 32: 629–34. doi: [10.1093/molbev/msu324](https://doi.org/10.1093/molbev/msu324) PMID: [25415968](https://pubmed.ncbi.nlm.nih.gov/25415968/)
47. Papaefthimiou D, Hrouzek P, Mugnai MA, Lukesova A, Turicchia S, Rasmussen U, et al. Differential patterns of evolution and distribution of the symbiotic behaviour in nostocacean cyanobacteria. *Int J Syst Evol Microbiol.* 2008; 58: 553–64. doi: [10.1099/ijs.0.65312-0](https://doi.org/10.1099/ijs.0.65312-0) PMID: [18319454](https://pubmed.ncbi.nlm.nih.gov/18319454/)
48. Weng J-K, Noel JP. Chemodiversity in Selaginella: a reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci.* 2013; 4: 119. doi: [10.3389/fpls.2013.00119](https://doi.org/10.3389/fpls.2013.00119) PMID: [23717312](https://pubmed.ncbi.nlm.nih.gov/23717312/)