



# Genome-Wide Analysis of the NADK Gene Family in Plants

Wen-Yan Li, Xiang Wang, Ri Li, Wen-Qiang Li, Kun-Ming Chen\*

State Key Laboratory of Crop Stress Biology in Arid Areas, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, China

## Abstract

**Background:** NAD(H) kinase (NADK) is the key enzyme that catalyzes *de novo* synthesis of NADP(H) from NAD(H) for NADP(H)-based metabolic pathways. In plants, NADKs form functional subfamilies. Studies of these families in *Arabidopsis thaliana* indicate that they have undergone considerable evolutionary selection; however, the detailed evolutionary history and functions of the various NADKs in plants are not clearly understood.

**Principal Findings:** We performed a comparative genomic analysis that identified 74 NADK gene homologs from 24 species representing the eight major plant lineages within the supergroup Plantae: glaucophytes, rhodophytes, chlorophytes, bryophytes, lycophytes, gymnosperms, monocots and eudicots. Phylogenetic and structural analysis classified these NADK genes into four well-conserved subfamilies with considerable variety in the domain organization and gene structure among subfamily members. In addition to the typical NAD\_kinase domain, additional domains, such as adenylate kinase, dual-specificity phosphatase, and protein tyrosine phosphatase catalytic domains, were found in subfamily II. Interestingly, NADKs in subfamily III exhibited low sequence similarity (~30%) in the kinase domain within the subfamily and with the other subfamilies. These observations suggest that gene fusion and exon shuffling may have occurred after gene duplication, leading to specific domain organization seen in subfamilies II and III, respectively. Further analysis of the exon/intron structures showed that single intron loss and gain had occurred, yielding the diversified gene structures, during the process of structural evolution of NADK family genes. Finally, both available global microarray data analysis and qRT-RCR experiments revealed that the NADK genes in *Arabidopsis* and *Oryza sativa* show different expression patterns in different developmental stages and under several different abiotic/biotic stresses and hormone treatments, underscoring the functional diversity and functional divergence of the NADK family in plants.

**Conclusions:** These findings will facilitate further studies of the NADK family and provide valuable information for functional validation of this family in plants.

**Citation:** Li W-Y, Wang X, Li R, Li W-Q, Chen K-M (2014) Genome-Wide Analysis of the NADK Gene Family in Plants. PLoS ONE 9(6): e101051. doi:10.1371/journal.pone.0101051

**Editor:** Manoj Prasad, National Institute of Plant Genome Research, India

**Received:** December 14, 2013; **Accepted:** June 2, 2014; **Published:** June 26, 2014

**Copyright:** © 2014 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by grants from the National Natural Science Foundation of China (grant number 31270299), the Basic Scientific Fund of The State Key Laboratory of Crop Stress Biology for Arid Areas, NWAUFU (CSBAAZD1301), the Basic Scientific Fund of NWAUFU (2014YB033), and the Program for New Century Excellent Talents in University of China (NCET-11-0440). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* Email: kunmingchen@nwsuaf.edu.cn

## Introduction

NAD(H) and NADP(H) are crucial coenzymes and play important and distinguishable roles in all living organisms [1]. NAD(H) is primarily involved in catabolic reactions, whereas NADP(H) participates in anabolic reactions, such as NADP(H)-dependent reductive anabolic pathways, signal transduction, and defense against oxidative stress [1,2]. Hence, regulation of the intracellular balance of NAD(H) and NADP(H) is critical. NAD(H) kinase (NADK) is the key enzyme in the *de novo* biosynthesis of NADP(H), catalyzing the transfer of a phosphoryl group from ATP to NAD(H), and thus plays an important role in the regulation of intracellular NAD(H)/NADP(H) balance for NADP(H)-based metabolic pathways.

NADK genes have been found in nearly all living organisms, including Archaea, eubacteria and eukaryotes, except for the intracellular parasite *Chlamydia trachomatis* [3]. NADK genes have

been cloned from a wide variety of species, including Archaea (*Methanococcus jannaschii* [4]), eubacteria (*Mycobacterium tuberculosis* [5]), *Escherichia coli* [6], yeast (*Saccharomyces cerevisiae* [7,8]), humans (*Homo sapiens* [9]) and plants (*Arabidopsis thaliana* [10,11,12]). Moreover, the number of NADKs in different organisms varies, with most prokaryotic organisms, such as Archaea and eubacteria, having only one NADK, whereas most eukaryotic organisms, such as yeast and plants, have several NADKs. For example, *Euglena gracilis* [13] has two NADKs and *S. cerevisiae* [14,15] and *Arabidopsis* have three [10,11,16].

The enzymatic properties of natural or recombinant NADKs from several organisms, including their substrate specificity and structural properties, especially of the active site, have been well characterized [2,9,17,18,19]. All characterized NADKs are homomultimers and exhibit species-dependent phosphoryl donor and acceptor specificity. Hence, these enzymes are classified into two groups: ATP-NAD kinases and inorganic polyphosphate

poly(P)/ATP-NAD kinases, based on the phosphoryl donor specificity [2]. For example, NADKs from gram-positive bacteria (*M. tuberculosis*) and Archaea (*M. jannaschii*) utilize both ATP and poly(P) as phosphoryl donors and they phosphorylate NAD<sup>+</sup>/NADH, whereas NADKs from gram-negative bacteria (*E. coli*) and eukaryotes (*S. cerevisiae*, *Arabidopsis* and human) utilize ATP but not poly(P), and phosphorylate NAD<sup>+</sup> and NAD<sup>+</sup>/NADH, respectively [2]. In addition, analysis of the sequences and crystal structures of NADKs identified three highly conserved and functionally important motifs within the NADK family (a GGDG motif, an NE/D motif and a Gly-rich motif) and a possible phosphate transfer mechanism [20,21,22,23,24].

The biological significance and physiological functions of NADKs have been characterized from studies of several organisms. For example, mutation of the single NADK gene in *M. tuberculosis* and *Salmonella enterica* is lethal [3,4]. Similarly, mutation of all three NADK genes in *S. cerevisiae* (*utr1/yef1/pos5*) or two of the three genes (*utr1/pos5*) also causes lethality, whereas the respective NADK gene mutations are not lethal in *Arabidopsis* [11,16,25,26]. In addition, studies of the physiological functions of NADKs in *S. cerevisiae* and humans revealed that NADKs play a major role in protecting living cells against oxidative stress [8,15,26,27] because NADPH is vital in the intracellular anti-oxidative defense system of most organisms [28].

In plants, NADKs are involved in regulating redox balance, biotic and abiotic stress responses and various developmental processes. Notably, NADK was the first protein in plants demonstrated to be regulated by the calcium-sensing protein calmodulin (CaM) [29], and plant NADKs are divided into CaM-independent and CaM-regulated isoforms. CaM-dependent NADK is essential for survival of plants under difficult conditions and for protecting plants against invading pathogens by helping to provide reductants for the NADPH-dependent oxidative burst [29,30,31]. For example, CaM-dependent NADK activity, but not CaM-independent NADK activity, increases under cold stress in green bean [32], and decreased in response to high salinity and drought in tomato [33] and wheat [34]. In *Arabidopsis*, three genes encoding NADK (*NADK1/NADK2/NADK3*) have been identified and their physiological functions have been characterized [2,35]. *AtNADK1* is located in the cytosol and expressed mainly in roots, *AtNADK2* is located in the chloroplasts and expressed mainly in leaves, whereas *AtNADK3* is located in the peroxisome and is strongly expressed in reproductive tissues, such as the stigma, pollen and carpel vasculature [10,11,36]. The *AtNADK1*-deficient mutant exhibits sensitivity to  $\gamma$ -irradiation and paraquat-induced oxidative stress [16]. The *AtNADK2*-deletion mutant displays hypersensitivity to environmental stresses that induce oxidative stress, such as UVB-irradiation, drought, heat shock and high salinity [11]. Similarly, the *AtNADK3*-null mutant shows hypersensitivity to oxidative stress, including methyl viologen, high salinity and osmotic shock [10]. Moreover, *AtNADK2* plays a vital role in chlorophyll synthesis and protects chloroplasts against oxidative damage [11]. In addition, plants growth and fertility are affected in the NADK-deficient mutants [2,11,37].

Although these studies in bacteria, yeast, human and *Arabidopsis* have led to an understanding of the biochemical properties and physiological functions of NADKs, there has been no systematic study of the evolution and functional divergence of the NADK gene family, especially in Plantae. Here, we performed a comprehensive analysis of the NADK gene family in 24 species, representing the eight major plant lineages within the supergroup Plantae. Phylogenetic analysis was performed to delineate the evolutionary history of the NADK family in Plantae, and exon/intron structure analysis was performed to gain insight into the

possible mechanisms of the structural diversity of NADK gene family. Finally, the tissue-specificity and inducibility of NADK gene expression in *Arabidopsis* and rice (*Oryza sativa*) were characterized by examining publicly available microarray datasets and by qRT-PCR experiments. The results obtained here will broaden our understanding of the roles of plant NADKs and provide a framework for further functional investigations of these genes in plants.

## Results

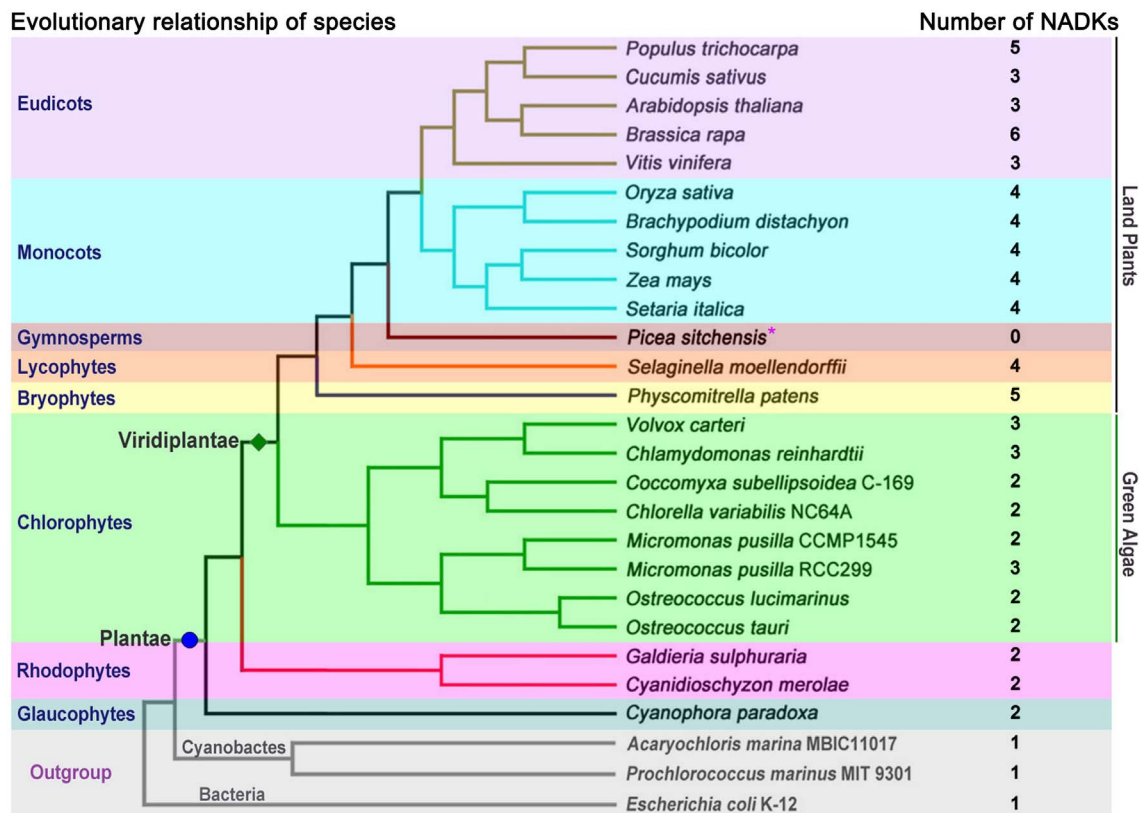
### Identification of NADK family members in plants

To comprehensively investigate and characterize the NADK gene family in plants, 24 species representing the eight major plant lineages within the supergroup Plantae, were selected for analysis (Figure 1). A hidden Markov model (HMM) search was performed with the obtained sequences and 74 NADK homologs were identified (Figure 1, Table S1). Except for the gymnosperm *Picea sitchensis*, for which the full genome sequence was not yet available, two or more NADK genes were identified in each genome of the selected species (Figure 1). Most of the aquatic algae, including glaucophytes, rhodophytes and chlorophytes, contained two NADK genes per genome and only three species in the chlorophytes, *Micromonas pusilla* RCC299, *Chlamydomonas reinhardtii* and *Volvox carteri*, carried three NADK genes per genome (Figure 1). In contrast, three or more NADK genes were found in all land plants, including bryophytes, lycophytes, monocots and eudicots (Figure 1). In addition, only one NADK gene was identified per genome in the cyanophytes (*Acaryochloris marina* MBIC11017 and *Prochlorococcus marinus* MIT 9301) and bacteria (*E. coli* K-12), outgroup (Figure 1).

The Pfam and SMART databases were used to analysis the functional domains of the identified NADK candidates. All of the putative NADKs possess a typical NAD\_kinase domain (Pfam accession number PF01513) and some also contain other functional domains such as an adenylate kinase domain (ADK; PF00406), a dual-specificity phosphatase, catalytic domain (DSPc; PF00782), and/or a protein tyrosine phosphatase, catalytic domain/DSPc domain (PTPc/DSPc; SMART accession number SM000012). Only one NADK candidate, CpNADK2, which did not contain a complete NAD\_kinase domain, was excluded from the following analysis.

### Phylogenetic analysis and classification of the NADK family

To explore the phylogenetic relationships among NADK family members in plants, we first generated a rooted maximum-likelihood phylogenetic tree with the 73 NADKs from the 24 species (Figure 2A), which was inferred from the amino acid sequences of their NAD\_kinase domains (Figure S1). Furthermore, phylogenetic trees reconstructed by the neighbor joining, minimum evolution and maximum parsimony methods showed very similar topologies to the maximum-likelihood tree (data not shown). Using *EcNADK1*, *AmNADK1* and *PmNADK1* from bacteria and cyanophytes as the outgroup, all NADK homologs in plants can be classified into four well-conserved subfamilies (I–IV; Figure 2A) with high statistical support, according to the topology and the deep duplication nodes of NADK paralogues in the maximum-likelihood tree. Interestingly, the topological relationship of members within subfamilies was highly consistent with the evolutionary relationships between species in Plantae (Figures 1 and 2A). The majority of aquatic algae contained two NADK genes per genome and grouped into subfamilies II and IV, whereas the majority of land plants, except for the gymnosperm



**Figure 1. Systematic evolutionary relationships of 24 species among eight lineages within the supergroup Plantae.** The numbers of NADK homologs in each species are listed next to the tree. \*, the genome sequencing of *Picea sitchensis* is not complete. doi:10.1371/journal.pone.0101051.g001

*Picea sitchensis*, carried  $\geq$  three NADK genes per genome and were clustered into subfamilies I, II and III (Figure 2A). Moreover, only one of the NADK genes from green algae (*CrNADK3*) fell into subfamily I and only two (*VcNADK3* and *MpNADK3*) fell into subfamily III (Figure 2A).

Examination of the chromosomal locations of the NADK family genes in the genomes of algae (*Ostreococcus lucimarinus*, *Ostreococcus tauri*, *C. reinhardtii* and *V. carteri*), bryophytes (*Physcomitrella patens*), lycophytes (*Selaginella moellendorffii*), monocots (*Brachypodium distachyon*, rice, *Sorghum bicolor* and *Zea mays*) and eudicots (*Arabidopsis*, *Brassica rapa*, *Populus trichocarpa* and *Vitis vinifera*) showed that NADK genes are randomly distributed (Figure S3). Moreover, a search for NADK paralogs using the Plant Genome Duplication Database (PGDD; <http://chibba.agtec.uga.edu/duplication/>) [38] revealed eight paralogous gene pairs in *P. patens*, rice, *Z. mays*, *B. rapa* and *P. trichocarpa* (Figure S4A), but none in the other species.

To further explore association of positive selection with duplication and divergence of NADK family genes, the rate of non-synonymous substitution ( $K_a$ ), synonymous substitution ( $K_s$ ) and the  $K_a/K_s$  ratios were calculated for the eight paralogous gene pairs and used to estimate duplication and divergence times. The  $K_a/K_s$  ratios varied from 0.18 to 0.39 among the five different species (Figure S4B). The fact that the  $K_a/K_s$  ratios were  $<1$  suggests that the NADK family genes have undergone strong negative selection pressure, and the duplication event was estimated to have occurred  $\sim$ 13.8–67.6 million years ago. Divergence of *P. patens*, rice, *Z. mays*, *P. trichocarpa* and *B. rapa* was estimated to have occurred 67.6, 61.5, 13.8, 20.8 and 23.1–33.1 million years ago, respectively (Figure S4B).

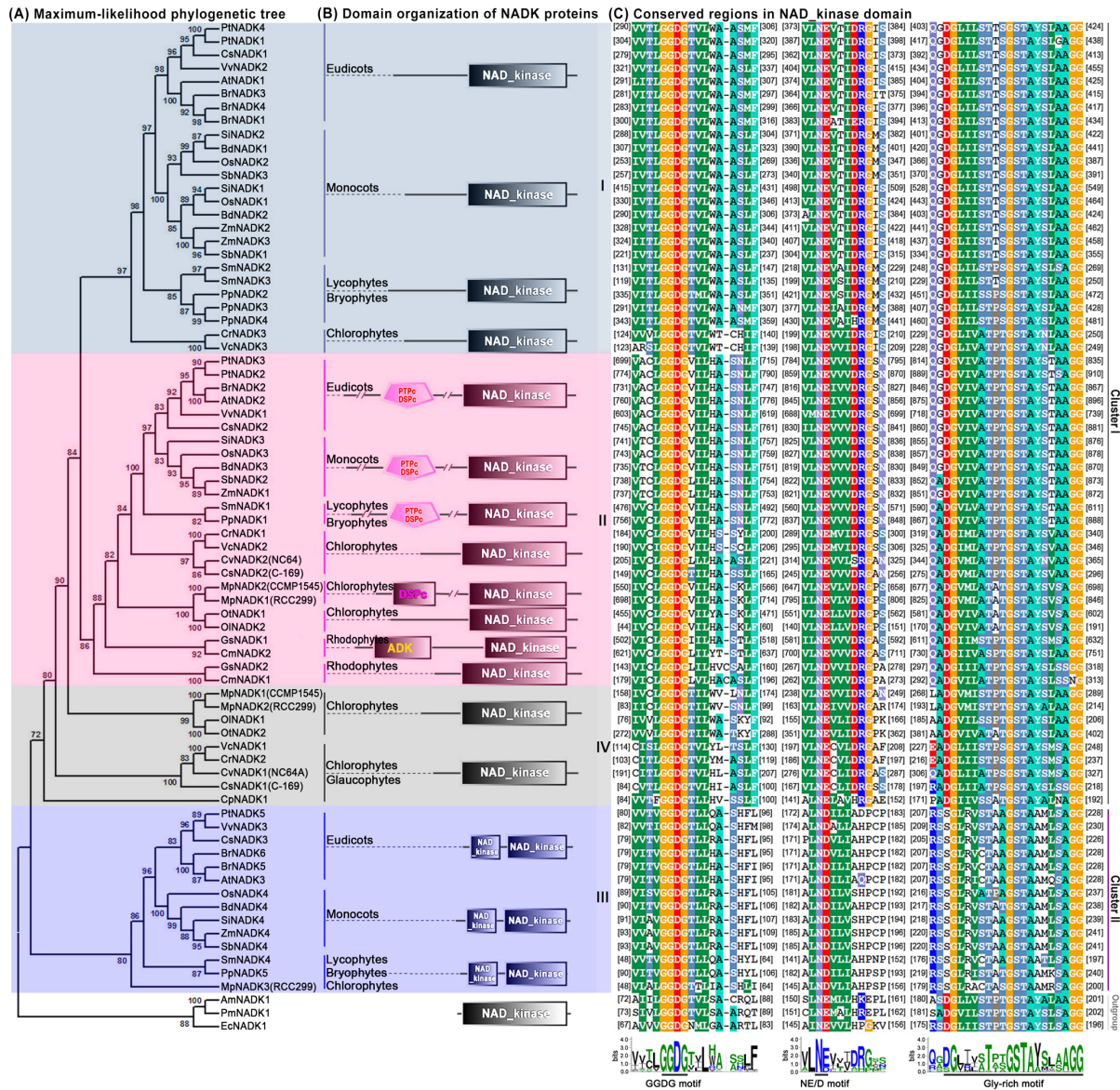
Further analysis of the functional domains showed that domain organization of NADKs in the different subfamilies varied considerably (Figure 2B). All of the identified NADKs contained a typical NAD\_kinase domain at the C terminus, but the proteins in subfamily II also carried the additional N-terminal catalytic domains noted earlier (DSPc, PTPc/DSPc or ADK; Figure 2B). Interestingly, the NAD\_kinase domain of proteins in subfamily III was not only divided into two parts, but also exhibited low sequence similarity ( $\sim$ 30%) between subfamily III members and the other subfamilies (Figures 2C and S1). In addition, based on the NAD\_kinase domain sequence similarity, the NADK gene subfamilies could be further divided into two clusters: cluster I (subfamilies I, II and IV) and cluster II (subfamily III).

More detailed analysis of the NAD\_kinase domains of the 73 NADKs revealed three highly conserved and functionally important motifs: a GGDG motif, an NE/D motif and a Gly-rich motif (Figures 2C and S1). The GGDG motif is involved in ATP-binding, whereas the NE/D and Gly-rich motifs are involved in NAD(H) binding [2,20,21,23,24].

### Structure analysis of NADK family genes

Intron position is generally very well-conserved in orthologous genes over long evolutionary time intervals, whereas exon/intron structure is slightly less, but sufficiently, conserved in paralogous genes to reveal evolutionary relationships between introns [39,40,41]. To investigate the gene structural diversity and possible mechanisms for the structural evolution of NADK homologs in green plants (Viridiplantae, excluding glaucophytes and rhodophytes), we analyzed the exon/intron organization in the coding sequence. Overall, there was considerable diversity in





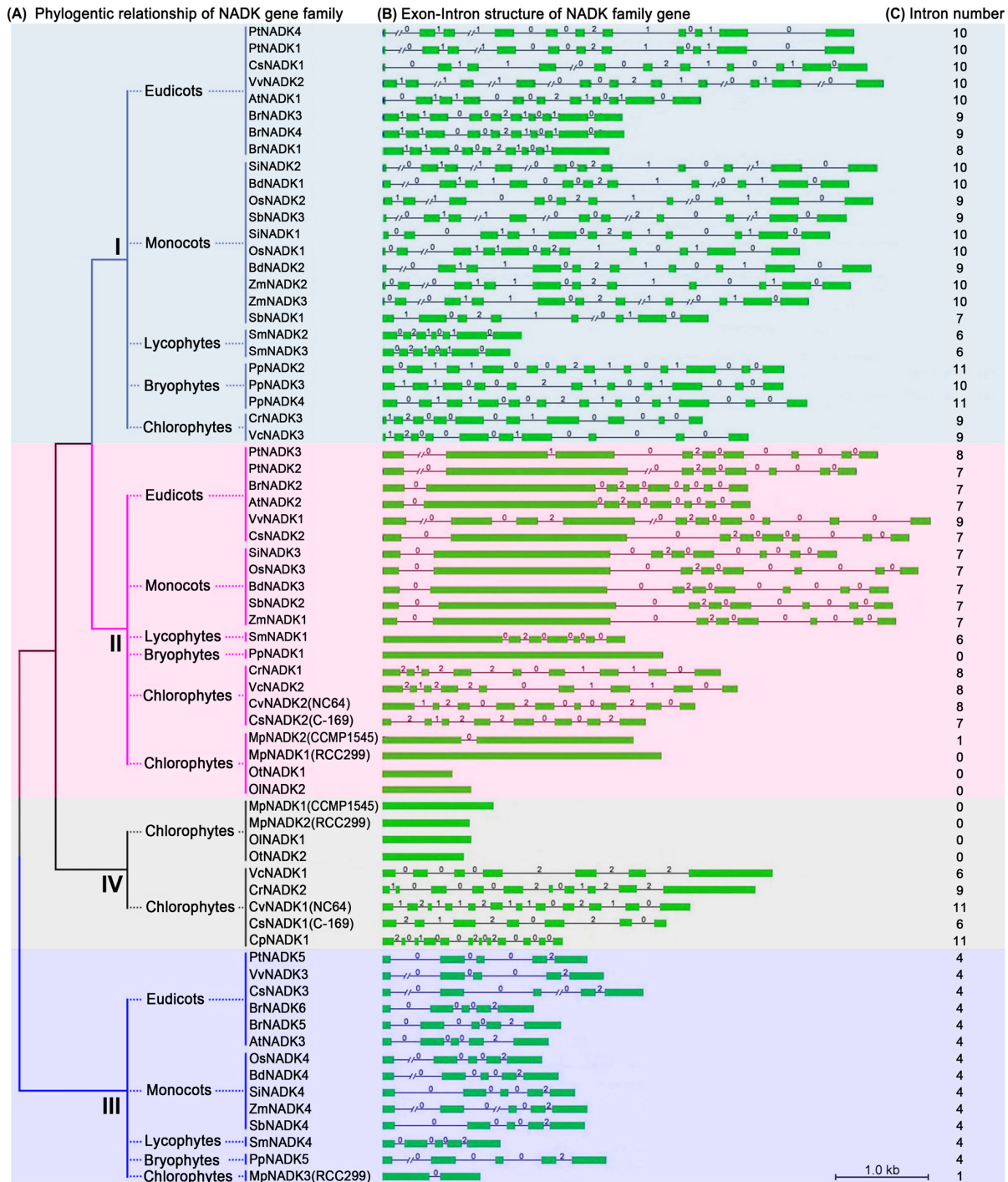
**Figure 2. Phylogenetic relationships and domain organization of NADK genes in plants.** (A) The rooted maximum-likelihood phylogenetic tree of NADK family members was inferred from the amino acid sequence alignment of the NAD\_kinase domain. Numbers above the nodes represent bootstrap values from 1000 replications. (B) Domain organization of the NADKs. (C) Amino acid sequence alignment of conserved motifs within the NAD\_kinase domain. doi:10.1371/journal.pone.0101051.g002

the number of introns (0–11) and the length of introns (50–7452 bp) in the NADK family genes (Figure 3). Interestingly, NADK family members within the same subfamily shared similar gene structure in terms of intron number, exon length, and/or intron phases, with the exception of subfamily IV (Figure 3). For instance, NADK genes in subfamily I had 6–11 introns, 48% (12/25) and 84% (21/25) of which contained 10 and 9–11 introns, respectively. We also investigated intron phases with respect to codons in the NADK genes. The intron phases were remarkably well conserved among subfamily members, whereas the intron arrangement and phases were strikingly distinct between subfamilies (Figure 3B).

To further explore intron loss or gain within the NADK family genes, we next examined the exon/intron organization of paralogs and orthologs in the land plants (Figure 4A). This analysis revealed

that single intron loss and gain likely occurred during the structural evolution of NADK family genes in land plants. For example, The paralogous genes *PtNADK1/4* and *SmNADK2/3* showed conserved exon/intron structure in terms of the number of introns and exon length, whereas a single intron appears to have been lost during the evolution of the *BrNADK1/4* and *PpNADK3/4* paralogs (Figures 3B and 4A). By contrast, a single intron gain occurred between the paralogs *PtNADK2* and *3* (Figures 3B and 4A). Among the orthologous NADK genes in land plants, the majority of subfamilies I, II and III members contained 10, 7 and 4 conserved common introns, respectively (Figure 4A), whereas some subfamily I and II NADK orthologs contained fewer introns. For example, *OsNADK1*, *SiNADK1*, *BdNADK1/2*, *zmNADK2/3* (subfamily I) have only nine conserved common introns. Similar intron losses were seen with *BrNADK1/3/4*, *SbNADK3* and



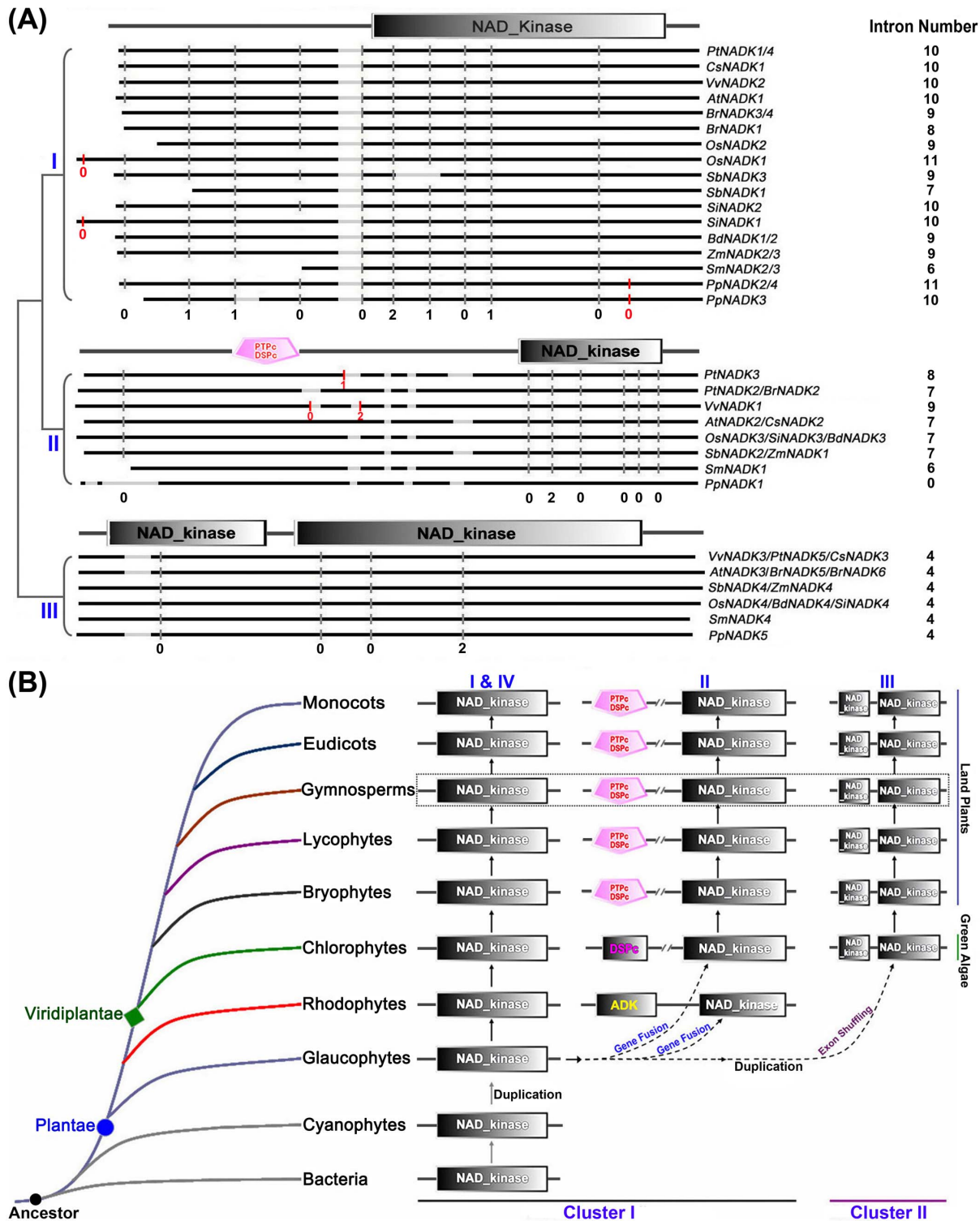


**Figure 3. Exon/intron structure of NADK family genes of green plants.** Green boxes represent exons; black lines represent introns; numbers 0, 1 and 2 are intron phases. The length of the boxes and lines are scaled relative to the length of the gene, and longer introns are denoted by a double slash. doi:10.1371/journal.pone.0101051.g003

*SmNADK2/3* (subfamily I), *PpNADK1* (subfamily II), and *SmNADK4* (subfamily III; Figure 4A). Single intron gain was also observed in the NADK genes of land plants. For instance, *OsNADK1*, *SiNADK1*, and *PpNADK2/3/4* in subfamily I and *PtNADK3* in subfamily II appear to have gained an intron, whereas *VuNADK1* in subfamily II appears to have gained two introns (Figure 4A).

### Tissue-specific expression patterns of NADK genes in *Arabidopsis* and rice

To investigate differences in expression of NADK genes in *Arabidopsis* and rice during plant development, we first analyzed the expression profiles of *AtNADK* and *OsNADK* genes available in expressed sequence tag (EST) databases from vegetative and reproductive development stages (Figures 5A and 5B and Tables



**Figure 4. The expansion and evolution of the NADK gene family in plants.** (A) Schematic comparison of intron distribution in NADK orthologs of land plants generated with the CIWOG software. Black horizontal lines are aligned sequences; gray horizontal lines are gaps in the alignment; gray vertical bars are conserved common introns; red vertical bars are gained introns. The numbers 0, 1 and 2 are intron phases. (B) A model for the expansion and evolution of the NADK gene family in Plantae. doi:10.1371/journal.pone.0101051.g004

S3 and S4, respectively). Varying levels of ESTs for each of the *AtNADK* and *OsNADK* genes were found the database, indicating that all of the NADK genes in *Arabidopsis* and rice are

expressed, and that they are likely differential expressed in different tissues or developmental stages. We then used *Arabidopsis* (ATH1, 22 k array) and rice (*Os* 51 k array) microarray data in

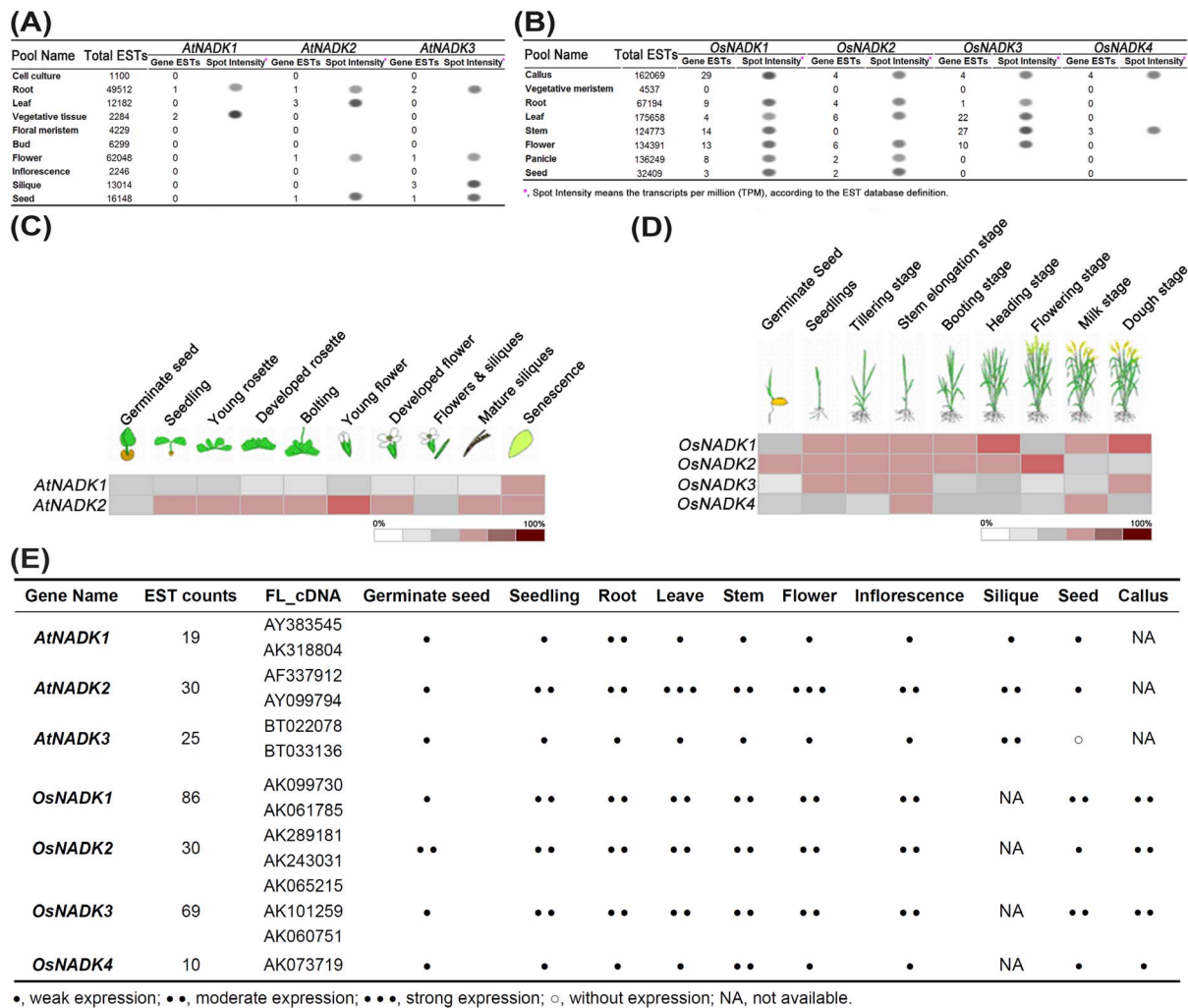
Genevestigator to analyze the expression patterns of the NADK genes in 10 and 9 developmental stages/tissues, respectively (Figure 5C, D and S5). Essentially identical developmental expression profiles for the *AtNADK* and *OsNADK* genes were also obtained with data from the *Arabidopsis* and rice eFP browsers in the Bio-Analytic Resource (<http://bar.utoronto.ca/welcome.htm>) database [42]. *AtNADK3* was not included in the comparison of tissue/developmental expression patterns because it was not included in the microarray data. The combined expression patterns of *AtNADKs* and *OsNADKs* from the EST and microarray data analyses are summarized in Figure 5E.

Overall, all of the *AtNADK* and *OsNADK* genes were expressed during the vegetative and reproductive development stages, and they displayed strong tissue specificity. In *Arabidopsis*, *AtNADK2* showed higher expression in most tissues and developmental stages than *AtNADK1* (Figures 5C and S5). The highest *AtNADK2* expression is seen in leaves and flowers (Figures 5A, C and S5), consistent with Waller et al. (2010) [36]; whereas *AtNADK1* expression is highest in the dry seed and senescence stages (Figures 5C and S5). *AtNADK3* appears to be mainly

expressed in the reproductive tissues such as flowers and siliques (Figure 5A), also consistent with the findings of Waller et al. (2010) [36]. In rice, the *OsNADK1*, *OsNADK2* and *OsNADK3* transcripts tend to accumulate to higher levels than the *OsNADK4* transcript in most tissues/developmental stages (Figures 5B, D and S5), but their individual expression patterns differ. For instance, during the germinate seed stage, *OsNADK1* transcript levels are low compared to *OsNADK2* (Figure 5D), whereas the opposite is observed in callus (Figure 5B). The highest levels of *OsNADK3* transcript are seen in leaf and stem (Figures 5D and S5); whereas relatively low levels of the *OsNADK4* transcript are seen in most tissues/developmental stages except for stem and inflorescence (Figures 5D and S5).

**Response profiles of NADK genes under abiotic/biotic stresses and hormone treatments in *Arabidopsis* and rice**

To understand the molecular mechanism of *AtNADKs* and *OsNADKs* transcriptional regulation under abiotic/biotic stresses and hormone treatments, we first identified potential *cis*-elements in the promoter regions of each NADK gene using the



**Figure 5. Developmental expression patterns of NADK family genes in *Arabidopsis* and rice.** Expression profiles of (A) *AtNADKs* and (B) *OsNADKs* inferred from public EST data. Expression profiles of (C) *AtNADKs* (except *AtNADK3*) and (D) *OsNADKs* in different developmental stages obtained from microarray data reported in Genevestigator. Results are shown as heat maps in white/gray/red (low to high) that reflect the percent of expression. (E) Summary of the relative expression patterns of *AtNADK* and *OsNADK* genes inferred from the combined EST (A, B; Tables S3 and S4) and microarray (C, D) profiles.

doi:10.1371/journal.pone.0101051.g005



PlantCARE program (Tables 1 and S2). We found that the promoter regions of AtNADK and OsNADK genes contain response elements for several abiotic/biotic stresses, such as low temperature, heat, drought (MYB binding sites), anaerobic conditions, and pathogens (TC-rich repeats, W box, GCC box, Box S, Box-W1 and EIRE) (Tables 1 and S2). In addition, the AtNADK and OsNADK gene promoters contain *cis*-elements for responding to several hormones, such as auxin, gibberellin, abscisic acid (ABA), ethylene, salicylic acid (SA) and methyl jasmonic acid (MeJA) (Tables 1 and S2). Further analysis showed that all of the *AtNADK* and *OsNADK* promoters contain anaerobic- and ABA-responsive elements; and the majority of them also carry heat stress, drought (MYB binding sites) and MeJA response elements (Tables 1 and S2).

To further demonstrate that the expression of AtNADK and OsNADK genes is induced by abiotic and biotic stress, we again examined *Arabidopsis* and rice microarray data in the Genevestigator database, as well as by qRT-PCR experiments. As shown in Figures 6, 7 and S6, the expression of AtNADK and OsNADK genes is induced to varying degrees by abiotic stress such as cold, heat, drought (PEG), salt (NaCl) and oxidative (methyl viologen, MV), as well as by biotic stress such as the pathogens *Botrytis cinerea*, *Blumeria graminis*, *Pseudomonas syringae*, *Agrobacterium tumefaciens*, *Mycosphaerella graminicola*, *Magnaporthe oryzae*, and *Magnaporthe grisea*. In *Arabidopsis*, *AtNADK1* is up-regulated in whole plants under heat and MV stresses, whereas it is down-regulated and does not change under PEG and cold treatment, respectively (Figure 6). By contrast, *AtNADK2* is down-regulated or does not change under these conditions (Figure 6). *AtNADK3* is up-regulated under cold stress, while it is down-regulated under other abiotic stresses including heat, PEG, NaCl, and MV (Figure 6). Additionally, both *AtNADK1* and *AtNADK2* can be up-regulated by some biotic stresses, although with different expression profiles under the same stress (Figure S6). In rice, *OsNADK1/2* are up-regulated in shoot and root under cold stress, in root under heat stress and in shoot under MV stress, respectively; while they are slightly down-regulated or does not change under these conditions (Figure 7). *OsNADK3* shows up-regulated under PEG in shoot and down-regulated or does not change under other abiotic stresses such as cold, heat, NaCl and MV (Figure 7). Similarly, *OsNADK4* is up-regulated in root under cold and heat and down-regulated or does not change under other abiotic stresses including PEG, NaCl, and MV (Figure 7). Additionally, *OsNADK1–4* can be induced by biotic stresses. For instance, *OsNADK1* is up-regulated in callus after infection by *A. tumefaciens* and root after infection by *M. oryzae*, whereas they are down-regulated in root and leaf after infection by *M. graminicola* and *M. grisea*, respectively (Figure S6).

As several hormone response elements were identified in the promoter regions of the AtNADK and OsNADK genes, we also examined their expression profiles under phytohormone treatments in the public *Arabidopsis* and rice microarray data (Figure S7), as well as by qRT-PCR experiments (Figures 6 and 7). *AtNADK1* and *AtNADK2* are differentially expressed in seedlings and cells treated with phytohormones, such as the auxins indole-3-acetic acid (IAA) and naphthaleneacetic acid (NAA), zeatin, SA, ABA, and ABA + SA (Figure S7). Interestingly, *AtNADK1* is up-regulated by ABA, ABA + SA, and ABA + MeJA treatment; whereas *AtNADK2* is down-regulation by ABA and ABA + SA treatment (Figure S7). The expressions of AtNADK genes by qRT-PCR showed *AtNADK1–2* are up-regulated by MeJA treatment and down-regulated or do not change by ABA treatment; *AtNADK3* is down-regulated by ABA treatment and does not change by MeJA treatment (Figure 6). In rice, *OsNADK1–4* are differentially expressed in seedlings and leaves treated with

IAA, NAA, zeatin, gibberellin, kinetin, ABA, SA and jasmonic acid (JA) (Figure S7). For instance, *OsNADK3* is up-regulated in roots under *trans*-zeatin treatment, whereas *OsNADK4* is down-regulated in leaves under the same treatment (Figure S7). It is also notable that *OsNADK1–3* can be induced in shoot by ABA treatment and do not change in root, but *OsNADK4* is not obviously changed with the same treatment. All *OsNADK1–4* were up-regulated or not obviously changed by MeJA treatment.

## Discussion

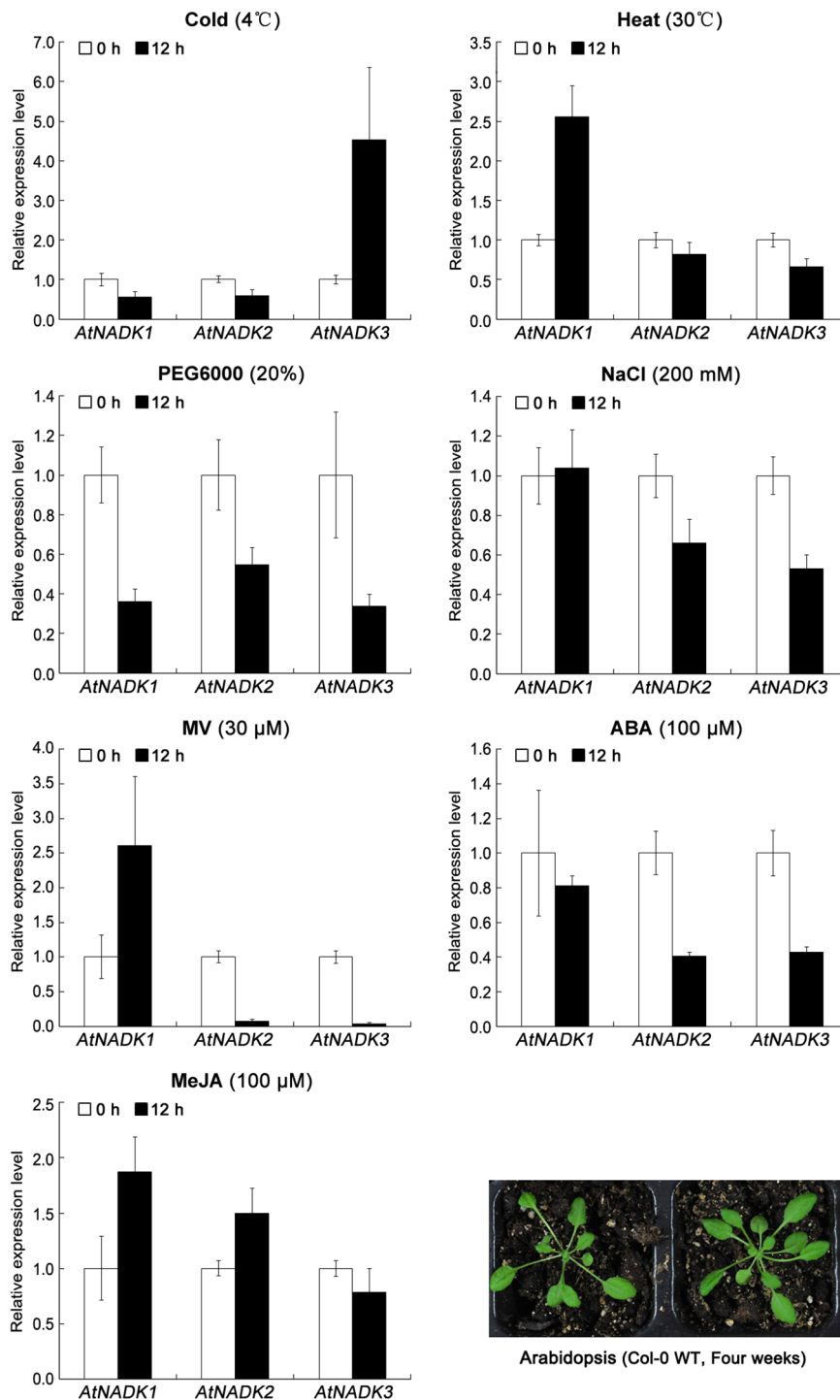
### Gene fusion and exon shuffling after gene duplication contributed to the expansion and evolution of the NADK family in plants

Gene duplication is a common phenomenon in eukaryotes and contributes to biological diversity during evolution [43,44,45]. The NADKs are represented by at least one gene in nearly every living organism. In this study, we found that the 24 representative plant species from the supergroup Plantae examined contain 2–6 NADK homologs, whereas cyanophytes, considered to be the ancestor of Plantae, had only one NADK homolog per genome, suggesting that a single gene duplication leading to the expansion of the NADK gene family, occurred during the divergence of ancestral cyanophytes from Plantae. The scattered distribution of NADK family genes on chromosomes (Figure S3) and the eight paralogous NADK gene pairs found in five land plants (Figure S4) suggest that segmental duplications may have been involved in the expansion of the NADK gene family and caused differences in the number of NADK genes within subfamilies and species of land plants after divergence from aquatic algae.

Gene fusion and exon shuffling after gene duplication are mechanisms that can enhance the functional divergence of duplicated genes by creating additional domains or rearranging the original functional domains [46,47,48,49]. In this study, phylogenetic analysis, together with the domain organizations and gene structures of NADK family genes, showed distinct evolutionary differences among subfamilies (Figure 2). Therefore, a model was constructed to account for the expansion and evolution of the NADK gene family in plants (Figure 4B). In this model, all NADK family members originated from a common ancestor, which contained only the typical NAD\_kinase domain and existed in all living organisms from prokaryotic bacteria to eukaryotic angiosperms; cyanophytes were considered to be the ancestor of Plantae, and at least one gene duplication occurred to yield two NADK genes in eukaryotic glaucophytes (Figures 1 and 4B). Moreover, during the evolutionary diversification from glaucophytes to rhodophytes and Viridiplantae, gene fusion events occurred and additional catalytic domains (ADK, DSPc or PTPc/DSPc) were acquired in the N-terminus, leading to the diversified domain organization seen in subfamily II (Figures 2B and 4B). In addition, exon shuffling after gene duplication contributed to the formation of the bifurcated NAD\_kinase domain in subfamily III (Figures 2B and 4B).

### Single intron loss and gain lead to the diversified gene structures

Gene structural diversity within gene families is another evolutionary mechanism that promotes variability, and intron loss or gain is important in generating structural diversity and complexity [41,50]. Analyzing the exon/intron structures of NADK genes, we found that the number of introns and intron phases among subfamily members are remarkably conserved, whereas the intron arrangement and intron phases are strikingly distinct between subfamilies (Figures 3 and 4). Further analysis of



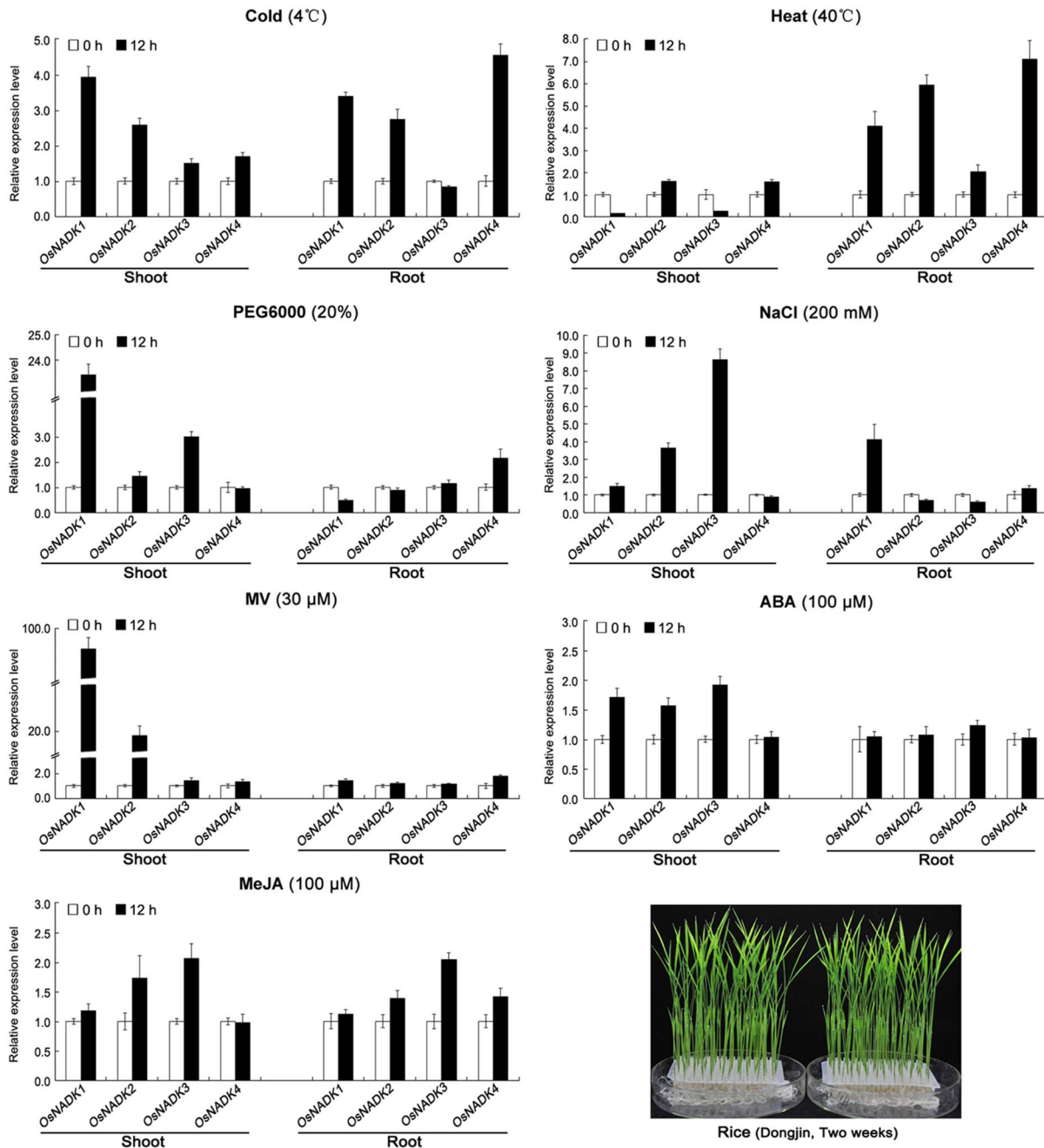
**Figure 6. Expression patterns of NADK family genes in *Arabidopsis* under abiotic stress and hormone treatments.** Expression levels of *AtNADK1–3* assayed by qRT-PCR under cold (4°C), heat (30°C), drought (20% PEG6000), salt (200 mM NaCl), oxidative (30 µM MV) stresses and MeJA (100 µM), ABA (100 µM) hormone treatments. Data are means  $\pm$  SD (n=3) and are representative of similar results from three independent experiments.

doi:10.1371/journal.pone.0101051.g006

the orthologous and paralogous genes in land plants (Figure 4) suggests that single intron loss as well as intron gain likely occurred and contributed to the diversification of gene structure, and consequent functional diversity and divergence, during the evolution of the NADK family in plants.

NADKs are involved in plant responses to several abiotic and biotic stresses

The three NADK genes of *Arabidopsis* belong to three different subfamilies according to our phylogenetic analysis (Figures 1 and 2). The three *AtNADKs* are known to have anti-oxidative functions [10,11,12,16,36], but they have distinct mechanisms by which



**Figure 7. Expression patterns of NADK family genes in rice under abiotic stress and hormone treatments.** Expression levels of *OsNADK1–4* assayed by qRT-PCR under cold (4°C), heat (30°C), drought (20% PEG6000), salt (200 mM NaCl), oxidative (30 µM MV) stresses and MeJA (100 µM), ABA (100 µM) hormone treatments. Data are means  $\pm$  SD (n=3) and are representative of similar results from three independent experiments.

doi:10.1371/journal.pone.0101051.g007

they facilitate plant resistance to oxidative stress. AtNADK1 (subfamily I) is a cytosolic enzyme that indirectly provides cytosolic NADP for plasma membrane NADPH oxidases, which are the key producers of reactive oxygen species under both normal and stress conditions in plants [36,51,52,53]. AtNADK2 (subfamily II) is a chloroplastic enzyme that plays a vital role in chlorophyll synthesis and chloroplast protection against oxidative damage by regulating plastidic NADP-biosynthesis [11,36]. AtNADK3 (subfamily III) is a peroxisomal enzyme that plays a prominent anti-oxidation role by providing the peroxisomal reductant NADPH [10,36]. NADK

genes in subfamily IV were only present in aquatic algae and grouped with *S. cerevisiae* *NADK1* (Pos5) in our broader phylogenetic analysis (Figure S2). Pos5 is located in the mitochondria and its deletion causes slow growth, sensitivity to oxidative stress, such as paraquat, hyperoxia, H<sub>2</sub>O<sub>2</sub>, and Cu<sup>2+</sup>, and biosynthesis deficiency on iron-sulfur clusters [25,54,55]. These observations suggest that plant NADKs in subfamily IV may also localize to mitochondria and protect against oxidative damage by regulating mitochondrial NADP-biosynthesis.



We found that plant NADKs are also involved in responding to several abiotic/biotic stresses, including cold, heat, drought, salt, oxidative and pathogens. *Arabidopsis NADK1* expression is up-regulated by H<sub>2</sub>O<sub>2</sub>, irradiation and the bacterial pathogen *P. syringae* pv. *tomato*; and the *AtNADK1* deficient mutant is sensitive to  $\gamma$ -irradiation and paraquat-induced oxidative stress [16]. The *AtNADK2* deletion mutant is also hypersensitive to environmental stresses that trigger oxidative stress, such as UVB-irradiation, drought, heat, and high salinity [11]; similarly *AtNADK3* transcription can be induced by MV, high salinity and osmotic shock [10,36]. In this study, we found that *AtNADK1* and *OsNADK1/2* belongs to subfamily I and can be induced to varying degrees by cold, drought (PEG), salt (NaCl), oxidative (MV) and pathogens (Figures 6, 7, S6 and S7). We also found that *AtNADK2*, which belongs to subfamily II, is up-regulated by pathogens such as *P. syringae* pv. *phaseolicola* and *Hyaloperonospora arabidopsidis* (Figure S6); *OsNADK3*, which also belongs to subfamily II, is up-regulated by *Nilaparvata lugens* and down-regulated under anaerobic conditions (Figure S6). *OsNADK4* belongs to subfamily III, and is up-regulated under anaerobic conditions, *N. lugens* and heat in root, and slightly up-regulated by drought or PEG in root (Figures 7 and S6). These observations, together with our analysis of *cis*-elements (Table 1), suggest that NADK genes in subfamilies I, II and III play an important role in plant responses to invading pathogens and abiotic stresses. The functions of NADKs in subfamily IV in abiotic/biotic-stress responses remains to be examined.

### NADKs may participate in regulation of hormone signaling

*AtNADK3* transcription can be slightly induced by ABA and the mutant is hypersensitive to ABA [10]. In this study, we found that all *AtNADK1–3* showed down-regulated or slightly down-regulated in response to ABA (Figure 6), while *OsNADK1–3* were up-regulated by ABA treatment (Figure 7). It's important to note that response profiles of NADK genes by qRT-PCR are not completely consisted with microarray data or previous studies, because of the different experimental material, processing and analyzing methods used in this study. Moreover, seven *AtNADK* and *OsNADK* genes (*AtNADK1–3* and *OsNADK1–4*) contain *cis*-acting ABA responsive elements (Table 1), suggesting that NADKs probably participate in regulation of ABA signaling in plants.

JA and its derivative MeJA are important signaling molecules in plant responses to many abiotic and biotic stresses such as wounding and pathogens [56,57]. Moreover, JA biosynthesis occurs in peroxisomes [57,58]. *AtNADK3* is also a peroxisomal enzyme and may phosphorylate the NADH derived from  $\beta$ -oxidation to yield NADPH needed for anti-oxidant defense [36]. *AtNADK1/2* and *OsNADK1–4* were up-regulated or slightly up-regulated under MeJA treatment (Figures 6 and 7). Moreover, several of these genes (*AtNADK1/2/3* and *OsNADK1/4*) contain CGTCA-MeJA-responsiveness motifs in their promoter regions (Table 1). These observations suggest that NADKs may also participate in regulation of JA or MeJA signaling in plants.

It should be pointed out that although transcription factors and *cis*-regulatory elements play important roles in regulating gene expression, the relationship is not always direct and the inducibility of expression is often affected by many factors and on multiple levels [59,60]. For example, ABA-responsive *cis*-regulatory elements were identified in the promoter regions of all seven NADK genes in *Arabidopsis* and rice, but only *OsNADK1–3* were up-regulated or slightly up-regulated by ABA treatment (Figures 7 and S7). Moreover, considering the limitations of the PlantCARE program in predicting *cis*-regulatory elements, further experiments are necessary to verify the relationship between the inducible

expression profiles of NADK genes and the *cis*-elements within their promoters.

## Materials and Methods

### Data retrieval and identification of NADK genes

The protein sequences of 24 completely or partially sequenced plant genomes representing the eight major plant lineages were retrieved from public databases. All of the protein sequences were the most current non-redundant sequences from the following sources (Table S1): the glaucophyte *Cyanophora paradoxa* from the *Cyanophora* Genome Project (<http://cyanophora.rutgers.edu/cyanophora/home.php>) [61]; the rhodophytes *Cyanidioschyzon merolae* and *Galdieria sulphuraria* from the *Cyanidioschyzon* Genome Project (<http://merolae.biol.s.u-tokyo.ac.jp/>) [62] and the *Galdieria* Genome Project (<http://genomics.msu.edu/galdieria/>) [63], respectively; the chlorophytes *O. tauri* (version 2.0) [64], *O. lucimarinus* (version 2.0) [65], *M. pusilla* strain RCC299 (version 3.0) and strain CCMP1545 (version 3.0) [66], *Chlorella variabilis* NC64A (version 1.0) [67], *Coccomyxa subellipsoidea* C-169 (version 2.0) [68], *C. reinhardtii* (version 4.0) [69] and *V. carteri* (version 2.0) and the bryophyte *P. patens* (version 3.0) [70] and lycophyte *S. moellendorffii* (version 1.1) [71] from the Joint Genome Institute (JGI, <http://genome.jgi-psf.org/>); for the gymnosperm *P. sitchensis* [72] from NCBI (<http://www.ncbi.nlm.nih.gov/>, partial sequences only because of its genome is only partially sequenced); the monocots *B. distachyon* (version 1.2) [73], *Setaria italica* (version 2.1) [74] and *S. bicolor* (version 2.1) [75] from JGI, rice (version 7.0) and *Z. mays* (version 5.6) from the Institute for Genomic Research Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/index.shtml>) [76] and MaizeSequence (<http://www.maizesequence.org/index.html>) [77], respectively; the eudicots *P. trichocarpa* (version 2.2) [78] and *Cucumis sativus* (version 1.0) [79] from JGI, *Arabidopsis* (version 10.0) from the Arabidopsis Information Resource (<http://www.arabidopsis.org/>) [80], *B. rapa* (version 1.5) from the Brassica Database (<http://brassicadb.org/brad/>) [81] and *V. vinifera* (version 1.0) from GenoScope (<http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/>) [82]. All of the above protein sequences were integrated into a local protein database for the subsequent identification of NADK homologs.

To identify the NADK genes and their homologs in the Plantae supergroup, HMMER v3.0 [83,84] was used to perform an HMM search against the local protein database, using the family-specific NAD\_kinase domain (PF01513) HMM profile obtained from the Pfam database [85,86]. The HMM search was performed with the default parameters and an E-value cutoff of  $1e^{-5}$ . If a candidate gene had multiple alternative splice variants, the longest variant was used to represent the candidate protein. The Pfam and SMART [87,88] databases were employed to detect conserved domains in the candidate proteins, and the search results were refined manually to eliminate partial NADK domains and other potential false positives.

### Sequence alignment and phylogenetic analysis

The NAD\_kinase domain sequences of the candidate proteins were aligned using the MUSCLE v3.8 program with the default parameters [89,90] and the alignments were manually edited using the BioEdit v7.0 program [91]. The rooted maximum-likelihood tree was inferred from the resulting alignments using the Phylip v3.68 package [92] under the  $\gamma$ -corrected Jones–Taylor–Thornton model [93] with default parameters, and the reliability of interior branches was assessed with 1000 bootstrap resamplings. In addition, the MEGA v5.0 program [94] was also used to reconstruct the phylogenetic trees by the neighbor joining,

**Table 1.** Abiotic/biotic stress and hormone response elements in *AtNADK* and *OsNADK* promoters\*.

| Cis-elements                     | Core sequences | Functions of the cis-elements          | <i>OsNADK1</i> | <i>OsNADK2</i> | <i>OsNADK3</i> | <i>OsNADK4</i> | <i>AtNADK1</i> | <i>AtNADK2</i> | <i>AtNADK3</i> |
|----------------------------------|----------------|--|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| <b>Abiotic/biotic stress</b>     |                |  |                |                |                |                |                |                |                |
| LTR                              | CCGAAA         | low-temperature responsiveness         |                | ✓              |                |                |                |                |                |
| HSE                              | AAAAAATTC      | heat-stress responsiveness             | ✓              | ✓              | ✓              | ✓              | ✓              |                | ✓              |
| MBS                              | TAAC TG        | MYB binding site, drought inducibility | ✓              | ✓              | ✓              | ✓              |                |                | ✓              |
| ARE                              | TGGTTT         | anaerobic induction                    | ✓              | ✓              | ✓              | ✓              | ✓              |                | ✓              |
| TC-rich repeats                  | ATTCTTAAC      | defense and stress responsiveness      | ✓              | ✓              | ✓              | ✓              | ✓              | ✓              | ✓              |
| W box                            | TTGACC         | wound and pathogen responsiveness      |                | ✓              |                |                | ✓              |                |                |
| Box S                            | AGCCACC        | wound and pathogen responsiveness      | ✓              |                |                |                |                |                |                |
| GCC box                          | AGCCGCC        | wound and pathogen responsiveness      |                |                | ✓              |                |                |                |                |
| EIRE                             | TTCGACC        | Elicitor responsiveness                |                |                |                |                | ✓              |                |                |
| Box-W1                           | TTGACC         | fungal elicitor responsiveness         |                |                | ✓              |                | ✓              |                | ✓              |
| <b>Hormone response elements</b> |                |  |                |                |                |                |                |                |                |
| AuxRR-core                       | GGTCCAT        | auxin responsiveness                   |                | ✓              |                |                |                |                |                |
| TGA-element                      | AACGAC         | Auxin responsiveness                   |                |                |                |                | ✓              | ✓              |                |
| GARE-motif                       | AAACAGA        | Gibberellin responsiveness             |                | ✓              |                |                | ✓              |                |                |
| ABRE                             | CAGGTG         | abscisic acid responsiveness           | ✓              | ✓              | ✓              | ✓              | ✓              | ✓              | ✓              |
| ERE                              | ATTCAAA        | Ethylene responsiveness                | ✓              | ✓              | ✓              | ✓              | ✓              |                |                |
| TCA-element                      | GAGAAGAATA     | salicylic acid responsiveness          |                |                |                |                | ✓              |                | ✓              |
| CGTCA-motif                      | CGTCA          | MeJA responsiveness                    | ✓              |                |                | ✓              | ✓              | ✓              | ✓              |

\*the cis-elements were identified with the PlantCARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) using the sequences 1500 bp upstream from the transcription start site of each NADK gene. The “ ✓ ” means the NADK gene contains this cis-element in the promoter region.  
doi:10.1371/journal.pone.0101051.t001

minimal evolution and maximum parsimony methods, and to display the phylogenetic trees.

The rates of non-synonymous substitution (Ka) and synonymous substitution (Ks) were estimated for the orthologous and paralogous gene pairs of NADK family genes using the Codeml program in PAML v4.3 [95] interface tool of PAL2NAL [96], based on the aligned amino acid sequences and the corresponding nucleotide sequences. The duplication and divergence times of each gene pairs were estimated from the Ks of  $\lambda$  substitutions per synonymous site per year as  $T = Ks/2\lambda$  ( $\lambda = 6.5 \times 10^{-9}$ ) [97,98].

### Exon/intron structure and conserved motif analysis

The exon/intron structures of individual NADK genes were obtained through the Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn>) [99] by aligning the coding or cDNA sequences with their corresponding genomic DNA sequences from Phytozome v9.1 (<http://www.phytozome.net/>) [100]. To illustrate the evolution of introns, gene models were inspected for annotation of introns, and exon/intron boundaries were manually checked. For a subset of genes, predictions pertaining to the types of introns were independently checked using Common Introns Within Orthologous Genes software (<http://ciwog.gdcb.iastate.edu/>) [101].

The conserved functional motifs within NAD\_kinase domains were identified using the MEME v4.9 program (<http://meme.sdsc.edu>) [102] with the default parameters, and the sequence logos (graphical representations) of these motifs or domains were generated with the WebLogo v3.3 server (<http://weblogo.threeplusone.com/create.cgi>) [103] based on the results of protein sequence alignments.

### Cis-regulatory elements and expression profile analysis

The 1500 bp upstream of the transcription start site of all NADK genes in *Arabidopsis* and rice were obtained from Phytozome v9.1 (<http://www.phytozome.net/>), and the cis-regulatory elements were identified using the PlantCARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [104,105].

ATH1 22 k and Os 51 k microarray data in the GeneInvestigator V3 database were used to analyze the tissue-specific and inducible expression profiles of NADK genes in *Arabidopsis* and rice [106], respectively. In addition, EST profiles of each NADK gene in *Arabidopsis* and rice were also analyzed by BLASTN search against the corresponding NCBI (<http://www.ncbi.nlm.nih.gov/>) EST database, with the coding sequences of the individual NADK gene as query. The BLASTN searches were performed with the following criteria: E-value <  $1e^{-10}$  and nucleotide identities > 95% over 150 bp.

### Plant materials, treatments and quantitative real-time PCR (qRT-PCR) analysis

The *Arabidopsis* (Col-0) seeds were pretreated at 4°C for 3 days and directly sown in Sunshine MVP potting soil, and then grown in growth chambers under 16 h light/8 h dark at 22±1°C; while rice (*Oryza sativa* ssp. *japonica* cv. Dongjin) seeds were pretreated germination at 26±1°C for 4 days and transplanted to 1/2 Hogland nutrient solution, and then grown in growth chambers under 16 h light/8 h dark at 26±1°C. Following two (rice) or four (*Arabidopsis*) weeks of growth, the seedlings were grown under 4°C or 30°C (*Arabidopsis*)/40°C (rice) for 12 h for cold and heat treatments, respectively; submerged in 200 mM NaCl or 20% PEG6000 solutions for 12 h for drought and salt treatments, respectively. For oxidative stress and hormone treatments,

solutions of 30 µM MV, 100 µM MeJA and ABA were separately sprayed on seedlings for 12 h. All the MV, MeJA and ABA used for treatments were purchased from Sigma-Aldrich. Samples were collected following treatment and immediately frozen at -80°C.

Total RNA was extracted by using Trizol reagent (Takara, Japan) according to the manufacturer's instructions and treated with RNase-free DNase I (Invitrogen, USA) for 15 min to remove any DNA contamination. RNA concentration and quality were verified by using the NanoDrop 1000 Spectrophotometers (Thermo, USA), and cDNAs were synthesized by using oligo d(T)<sub>18</sub> reverse primer from 5 µg of total RNA in a total volume of 20 µL by using EasyScript First-Strand cDNA Synthesis SuperMix (TransGen Biotech, China). qRT-PCR reactions were carried out in 96-well (20 µL) format by using the FastStart Essential DNA Green Master (Roche, Switzerland), and were performed in an CFX96 Touch Real-Time PCR Detection System (BIO-RAD, USA). The reactions were repeated three times and the quantitative analysis used the  $2^{-\Delta\Delta CT}$  method. All gene-specific primers were designed to avoid the conserved region and span introns or cross an exon-exon junction. The detailed primer sequences are shown in Table S5. The *AtTub6* (AT5G12250) and *OsActin1* (accession ID KC140126) were chosen as the internal control in *Arabidopsis* and rice, respectively.

### Supporting Information

**Figure S1 Amino acid sequence alignment of NAD\_kinase domains in NADKs.**

(PDF)

**Figure S2 Phylogenetic relationship and domain organization of NADK genes in plants, yeast and humans.**

(PDF)

**Figure S3 Chromosomal locations of NADK family genes in plants.**

(PDF)

**Figure S4 Segmental duplications and duplication and divergence times of NADK family genes in land plants.**

(PDF)

**Figure S5 Developmental expression patterns of NADK family genes in *Arabidopsis* and rice.**

(PDF)

**Figure S6 Expression patterns of NADK family genes in *Arabidopsis* and rice under abiotic and biotic stresses.**

(PDF)

**Figure S7 Expression patterns of NADK family genes in *Arabidopsis* and rice with various hormone treatments.**

(PDF)

**Table S1 The NADK family genes in plants.**

(PDF)

**Table S2 Cis-element analysis of the *AtNADK* and *OsNADK* promoters.**

(PDF)

**Table S3 EST profiles of NADK genes in *Arabidopsis*.**

(PDF)

**Table S4 EST profiles of NADK genes in rice.**

(PDF)

**Table S5 The gene-specific primers used for qRT-PCR in this study.**

(PDF)



## Author Contributions

Conceived and designed the experiments: WYL KMC. Performed the experiments: XW RL. Analyzed the data: WYL WQL. Contributed

reagents/materials/analysis tools: WYL XW RL. Wrote the paper: WYL WQL. Provided funding and critically revised the manuscript: KMC.

## References

1. Ying W (2008) NAD<sup>+</sup>/NADH and NADP<sup>+</sup>/NADPH in cellular functions and cell death: regulation and biological consequences. *Antioxid Redox Signal* 10: 179–206.
2. Kawai S, Murata K (2008) Structure and function of NAD kinase and NADP phosphatase: key enzymes that regulate the intracellular balance of NAD(H) and NADP(H). *Biosci Biotechnol Biochem* 72: 919–930.
3. Grose JH, Joss L, Velick SF, Roth JR (2006) Evidence that feedback inhibition of NAD kinase controls responses to oxidative stress. *Proc Natl Acad Sci U S A* 103: 7601–7606.
4. Kawai S, Fukuda C, Mukai T, Murata K (2005) MJ0917 in archaeon *Methanococcus jannaschii* is a novel NADP phosphatase/NAD kinase. *J Biol Chem* 280: 39200–39207.
5. Kawai S, Mori S, Mukai T, Suzuki S, Yamada T, et al. (2000) Inorganic Polyphosphate/ATP-NAD kinase of *Micrococcus flavus* and *Mycobacterium tuberculosis* H37Rv. *Biochem Biophys Res Commun* 276: 57–63.
6. Kawai S, Mori S, Mukai T, Hashimoto W, Murata K (2001) Molecular characterization of *Escherichia coli* NAD kinase. *Eur J Biochem* 268: 4359–4365.
7. Kawai S, Suzuki S, Mori S, Murata K (2001) Molecular cloning and identification of UTR1 of a yeast *Saccharomyces cerevisiae* as a gene encoding an NAD kinase. *FEMS Microbiol Lett* 200: 181–184.
8. Outten CE, Culotta VC (2003) A novel NADH kinase is the mitochondrial source of NADPH in *Saccharomyces cerevisiae*. *EMBO J* 22: 2015–2024.
9. Lerner F, Niere M, Ludwig A, Ziegler M (2001) Structural and functional characterization of human NAD kinase. *Biochem Biophys Res Commun* 288: 69–74.
10. Chai MF, Wei PC, Chen QJ, An R, Chen J, et al. (2006) NADK3, a novel cytoplasmic source of NADPH, is required under conditions of oxidative stress and modulates abscisic acid responses in *Arabidopsis*. *Plant J* 47: 665–674.
11. Chai MF, Chen QJ, An R, Chen YM, Chen J, et al. (2005) NADK2, an *Arabidopsis* chloroplastic NAD kinase, plays a vital role in both chlorophyll synthesis and chloroplast protection. *Plant Mol Biol* 59: 553–564.
12. Turner WL, Waller JC, Vanderheld B, Snedden WA (2004) Cloning and characterization of two NAD kinases from *Arabidopsis*. Identification of a calmodulin binding isoform. *Plant Physiol* 135: 1243–1255.
13. Stephan C, Renard M, Montrichard F (2000) Evidence for the existence of two soluble NAD(+) kinase isoenzymes in *Euglena gracilis* Z. *Int J Biochem Cell Biol* 32: 855–863.
14. Li YF, Shi F (2006) Partial rescue of pos5 mutants by YEF1 and UTR1 genes in *Saccharomyces cerevisiae*. *Acta Biochim Biophys Sin (Shanghai)* 38: 293–298.
15. Shi F, Kawai S, Mori S, Kono E, Murata K (2005) Identification of ATP-NADH kinase isozymes and their contribution to supply of NADP(H) in *Saccharomyces cerevisiae*. *FEBS J* 272: 3337–3349.
16. Berrin JG, Pierrugues O, Brutusco C, Alonso B, Montillet JL, et al. (2005) Stress induces the expression of AtNADK-1, a gene encoding a NAD(H) kinase in *Arabidopsis thaliana*. *Mol Genet Genomics* 273: 10–19.
17. Oganessian V, Huang C, Adams PD, Jancarik J, Yokota HA, et al. (2005) Structure of a NAD kinase from *Thermotoga maritima* at 2.3 Å resolution. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 61: 640–646.
18. Mori S, Kawai S, Mikami B, Murata K (2001) Crystallization and preliminary X-ray analysis of NAD kinase from *Mycobacterium tuberculosis* H37Rv. *Acta Crystallogr D Biol Crystallogr* 57: 1319–1320.
19. Ando T, Ohashi K, Ochiai A, Mikami B, Kawai S, et al. (2011) Structural Determinants of Discrimination of NAD<sup>+</sup> from NADH in Yeast Mitochondrial NADH Kinase Pos5. *J Biol Chem* 286: 29984–29992.
20. Raffaelli N, Finaurini L, Mazzola F, Pucci L, Sorci L, et al. (2004) Characterization of *Mycobacterium tuberculosis* NAD kinase: functional analysis of the full-length enzyme by site-directed mutagenesis. *Biochemistry* 43: 7610–7617.
21. Garavaglia S, Raffaelli N, Finaurini L, Magni G, Rizzi M (2004) A novel fold revealed by *Mycobacterium tuberculosis* NAD kinase, a key allosteric enzyme in NADP biosynthesis. *J Biol Chem* 279: 40980–40986.
22. Labesse G, Douguet D, Assairi L, Gilles AM (2002) Diacylglyceride kinases, sphingosine kinases and NAD kinases: distant relatives of 6-phosphofructokinases. *Trends Biochem Sci* 27: 273–275.
23. Mori S, Yamasaki M, Maruyama Y, Momma K, Kawai S, et al. (2005) NAD-binding mode and the significance of intersubunit contact revealed by the crystal structure of *Mycobacterium tuberculosis* NAD kinase-NAD complex. *Biochem Biophys Res Commun* 327: 500–508.
24. Liu J, Lou Y, Yokota H, Adams PD, Kim R, et al. (2005) Crystal structures of an NAD kinase from *Archaeoglobus fulgidus* in complex with ATP, NAD, or NADP. *J Mol Biol* 354: 289–303.
25. Shianna KV, Marchuk DA, Strand MK (2006) Genomic characterization of POS5, the *Saccharomyces cerevisiae* mitochondrial NADH kinase. *Mitochondrion* 6: 94–101.
26. Bieganski P, Seidle HF, Wojcik M, Brenner C (2006) Synthetic lethal and biochemical analyses of NAD and NADH kinases in *Saccharomyces cerevisiae* establish separation of cellular functions. *J Biol Chem* 281: 22439–22445.
27. Pollak N, Niere M, Ziegler M (2007) NAD kinase levels control the NADPH concentration in human cells. *J Biol Chem* 282: 33562–33571.
28. Singh R, Mailloux RJ, Puisieux-Dao S, Appanna VD (2007) Oxidative stress evokes a metabolic adaptation that favors increased NADPH synthesis and decreased NADH production in *Pseudomonas fluorescens*. *J Bacteriol* 189: 6665–6675.
29. Anderson JM, Charbonneau H, Jones HP, McCann RO, Cormier MJ (1980) Characterization of the plant nicotinamide adenine dinucleotide kinase activator protein and its identification as calmodulin. *Biochemistry* 19: 3113–3120.
30. Harding SA, Oh SH, Roberts DM (1997) Transgenic tobacco expressing a foreign calmodulin gene shows an enhanced production of active oxygen species. *EMBO J* 16: 1137–1144.
31. Zhou CY, Wu GL, Duan ZQ, Wu LL, Gao YS, et al. (2010) H<sub>2</sub>O<sub>2</sub>-NOX system: an important mechanism for developmental regulation and stress response in plants. *Chinese Bulletin of Botany* 45: 615–631.
32. Ruiz JM, Sanchez E, Garcia PC, Lopez-Lefebvre LR, Rivero RM, et al. (2002) Proline metabolism and NAD kinase activity in green bean plants subjected to cold-shock. *Phytochemistry* 59: 473–478.
33. Delumeau O, Paven M-CM-L, Montrichard F, Laval-Martin DL (2000) Effects of short-term NaCl stress on calmodulin transcript levels and calmodulin-dependent NAD kinase activity in two species of tomato. *Plant, Cell & Environment* 23: 329–336.
34. Zagdanska B (1990) NAD kinase activity in wheat leaves under water deficit. *Acta Biochim Pol* 1990;37(3): 385–9.
35. Shi F, Li Y, Wang X (2009) Molecular properties, functions, and potential applications of NAD kinases. *Acta Biochim Biophys Sin (Shanghai)* 41: 352–361.
36. Waller JC, Dhanoa PK, Schumann U, Mullen RT, Snedden WA (2010) Subcellular and tissue localization of NAD kinases from *Arabidopsis*: compartmentalization of de novo NADP biosynthesis. *Planta* 231: 305–317.
37. Wu LL, Zhou CY, Gao YS, Cong YX, Chen KM, et al. (2011) Cloning and genetic transformation of *OxNADK3* gene in rice. *Journal of Nuclear Agricultural Sciences* 25: 0863–0870.
38. Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res* 41: D1152–1158.
39. Hardison RC (1996) A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc Natl Acad Sci U S A* 93: 5675–5679.
40. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr Biol* 13: 1512–1517.
41. Li W, Liu B, Yu L, Feng D, Wang H, et al. (2009) Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. *BMC Evol Biol* 9: 90.
42. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, et al. (2007) An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2: e718.
43. Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433–438.
44. Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: a tale of duplications. *Trends Plant Sci* 14: 680–688.
45. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R (2013) Gene duplication as a major force in evolution. *J Genet* 92: 155–161.
46. Kolkman JA, Stemmer WP (2001) Directed evolution of proteins by exon shuffling. *Nat Biotechnol* 19: 423–428.
47. Jones CD, Begun DJ (2005) Parallel evolution of chimeric fusion genes. *Proc Natl Acad Sci U S A* 102: 11373–11378.
48. Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res* 20: 1313–1326.
49. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, et al. (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecific diversity in maize. *Nat Genet* 37: 997–1002.
50. Zhang Z, Kishino H (2004) Genomic background predicts the fate of duplicated genes: evidence from the yeast genome. *Genetics* 166: 1995–1999.
51. Wang GF, Li WQ, Li WY, Wu GL, Zhou CY, et al. (2013) Characterization of Rice NADPH Oxidase Genes and Their Expression under Various Environmental Conditions. *Int J Mol Sci* 14: 9440–9458.
52. Sagi M, Fluhr R (2006) Production of reactive oxygen species by plant NADPH oxidases. *Plant Physiol* 141: 336–340.
53. Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, et al. (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* 422: 442–446.

54. Strand MK, Stuart GR, Longley MJ, Graziewicz MA, Dominick OC, et al. (2003) POS5 gene of *Saccharomyces cerevisiae* encodes a mitochondrial NADH kinase required for stability of mitochondrial DNA. *Eukaryot Cell* 2: 809–820.
55. Pain J, Balamurali MM, Dancis A, Pain D (2010) Mitochondrial NADH kinase, Pos5p, is required for efficient iron-sulfur cluster biogenesis in *Saccharomyces cerevisiae*. *J Biol Chem* 285: 39409–39424.
56. Chehab EW, Kaspi R, Savchenko T, Rowe H, Negre-Zakharov F, et al. (2008) Distinct roles of jasmonates and aldehydes in plant-defense responses. *PLoS One* 3: e1904.
57. Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. *Ann Bot* 111: 1021–1058.
58. Schaller A, Stüntzi A (2009) Enzymes in jasmonate biosynthesis - structure, function, regulation. *Phytochemistry* 70: 1532–1538.
59. Ma Q, Chim GW, Szustakowski JD, Bakhtiarova A, Kosinski PA, et al. (2008) Uncovering mechanisms of transcriptional regulations by systematic mining of *cis* regulatory elements with gene expression profiles. *BioData Min* 1: 4.
60. Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11: 113–127.
61. Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, et al. (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335: 843–847.
62. Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428: 653–657.
63. Barbier G, Oesterheld C, Larson MD, Halgren RG, Wilkerson C, et al. (2005) Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol* 137: 460–474.
64. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, et al. (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci U S A* 103: 11647–11652.
65. Palenik B, Grimwood J, Aerts A, Salamov A, et al. (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A* 104: 7705–7710.
66. Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324: 268–272.
67. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, et al. (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* 22: 2943–2955.
68. Blanc G, Agarkova I, Grimwood J, Kuo A, Bruggeman A, et al. (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* 13: R39.
69. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318: 245–250.
70. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.
71. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, et al. (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332: 960–963.
72. Ralph SG, Chun HJ, Kolosova N, Cooper D, Oddy C, et al. (2008) A conifer genomics resource of 200,000 spruce (*Picea spp.*) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* 9: 484.
73. International-Brachypodium-Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763–768.
74. Zhang G, Liu X, Quan Z, Cheng S, Xu X, et al. (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30: 549–554.
75. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551–556.
76. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.
77. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.
78. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.
79. Huang S, Li R, Zhang Z, Li L, Gu X, et al. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41: 1275–1281.
80. Arabidopsis-Genome-Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
81. Wang X, Wang H, Wang J, Sun R, Wu J, et al. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43: 1035–1039.
82. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
83. Zhang Z, Wood WI (2003) A profile hidden Markov model for signal peptides generated by HMMER. *Bioinformatics* 19: 307–308.
84. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37.
85. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247–251.
86. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40: D290–301.
87. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857–5864.
88. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40: D302–305.
89. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
90. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
91. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
92. Krawetz S, Retief J (1999) *Phylogenetic Analysis Using PHYLIP*. *Bioinformatics Methods and Protocols: Humana Press*. 243–258.
93. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275–282.
94. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
95. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
96. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34: W609–612.
97. Puranik S, Sahu PP, Mandal SN, B VS, Parida SK, et al. (2013) Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *PLoS One* 8: e64594.
98. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
99. Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. *Yi Chuan* 29(8): 1023–6.
100. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40: D1178–1186.
101. Wilkerson MD, Ru Y, Brendel VP (2009) Common introns within orthologous genes: software and application to plants. *Briefings in Bioinformatics* 10: 631–644.
102. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37: W202–208.
103. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–1190.
104. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, et al. (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30: 325–327.
105. Rombauts S, Dehais P, Van Montagu M, Rouze P (1999) PlantCARE, a plant *cis*-acting regulatory element database. *Nucleic Acids Res* 27: 295–296.
106. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, et al. (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008: 420747.