Genome-wide analysis of haloacid dehalogenase genes reveals their function in phosphate starvation responses in rice

Zezhen Du1,2, Suren Deng1,2, Zixuan Wu1,2, Chuang Wang1,2*

1 Microelement Research Center, College of Resources & Environment, Huazhong Agricultural University, Wuhan, P. R. China, 2 Key Laboratory of Arable Land Conservation (Middle and Lower Reaches of Yangtze River), MOA, Huazhong Agricultural University, Wuhan, P. R. China

* chuang.wang@mail.hzau.edu.cn

Abstract

The HAD superfamily is named after the halogenated acid dehalogenase found in bacteria, which hydrolyses a diverse range of organic phosphate substrates. Although certain studies have shown the involvement of HAD genes in Pi starvation responses, systematic classification and bioinformatics analysis of the HAD superfamily in plants is lacking. In this study, 41 and 40 HAD genes were identified by genomic searching in rice and Arabidopsis, respectively. According to sequence similarity, these proteins are divided into three major groups and seven subgroups. Conserved motif analysis indicates that the majority of the identified HAD proteins contain phosphatase domains. A further structural analysis showed that HAD proteins have four conserved motifs and specified cap domains. Fewer HAD genes show collinearity relationships in both rice and Arabidopsis, which is consistent with the large variations in the HAD genes. Among the 41 HAD genes of rice, the promoters of 28 genes contain Pi-responsive cis-elements. Mining of transcriptome data and qRT-PCR results showed that at least the expression of 17 HAD genes was induced by Pi starvation in shoots or roots. These HAD proteins are predicted to be involved in intracellular or extracellular Po recycling under Pi stress conditions in plants.

Introduction

Phosphorus (P) is one of the essential macronutrients for plant growth and development. There are two different forms of P in soil: inorganic P (Pi) and organic P (Po), of which plants can only absorb and utilize water-soluble Pi. Although P is abundant in the Earth’s crust, it is usually present in soil in the form of Po or fixed with other metals, making it unavailable to plants and insufficient to support the optimal growth of plants [1, 2]. To cope with Pi deficiency stress, plants have evolved a series of physiological and biochemical strategies to enhance the uptake and utilization of P [3]. Inducing the expression and synthesis of phosphatases is one of the important responses to Pi stress in plants [4].

Phosphatase (EC 3.1.3) is a type of hydrolase that acts to free attached phosphate groups from other molecules [5]. It was reported that Pi starvation induced the synthesis of both
intracellular and extracellular acid phosphatase isozymes in plants [6]. More specifically, the activity of intracellular acid phosphatases (IAPs) increases after two days of Pi starvation, which is considered to be involved in recycling Pi from intracellular Po compounds [6, 7]. In contrast, the activity of secreted acid phosphatases (SAPs) is enhanced under prolonged Pi starvation conditions and participates in the degradation of extracellular Po [6, 8, 9]. Many Pi starvation-induced IAPs and SAPs have been purified from different plant species, and the majority of them are encoded by purple acid phosphatase (PAP) genes [8, 10–16]. Compared with PAP genes, other types of Pi starvation-induced phosphatases are less studied in plants.

The HAD superfamily is named after the halogenated acid dehalogenase in bacteria, which catalyses carbon or phosphoryl group transfer reactions on a diverse range of substrates [17]. Enzymes exist in all three superkingdoms of life, showing very disparate biological functions [18]. The majority of HAD proteins are involved in phosphoryl transfer, including phosphatases, P-type ATPases, phosphonatases, and phosphotransferases [19]. All proteins of the HAD superfamily contain a similar core catalytic domain, which is a unified Rossmannoid fold [18]. Moreover, members of the HAD superfamily insert three kinds of cap structures in the core Rossmannoid fold [18]. Although the HAD superfamily proteins show little overall sequence similarity (<15%), these proteins contain four highly conserved sequence motifs [19, 20].

Several HAD genes have been reported to regulate Pi deficiency stress responses in different plant species. In Arabidopsis, the expression of AtPECP1 and AtPS2 was significantly induced by Pi starvation. Subsequently, these enzymes were proven to recycle Pi from phosphocholine and phosphoethanolamine under Pi starvation conditions [21–25]. Pi starvation also induces the secretion of HAD phosphatases in Arabidopsis, which may participate in the utilization of extracellular Po [26]. However, mutation or overexpression of these HAD genes does not show any growth phenotypes under either Pi-replete or Pi-depleted conditions in Arabidopsis [23, 26]. In contrast, overexpression of OsHAD1 significantly enhanced extracellular Po utilization and stimulated the growth of rice under Pi-deficient conditions [27]. Overexpression of LePS2 resulted in increased acid phosphatase activity and anthocyanin accumulation and delayed flowering in tomatoes [28]. Similarly, Pi starvation-induced PvPS2.1 and PvPS2.2 increase the Pi adaptation ability in plants [29]. Moreover, soybean GmHAD1 was identified to be associated with tolerance to low Pi stress by map-based cloning, and GmHAD1 overexpression led to increased acid phosphatase activity and Pi utilization efficiency in both soybean and Arabidopsis plants [30].

Although a number of HAD genes are involved in Pi starvation responses in plants, systematic classification of the HAD superfamily in plants and functional prediction of Pi stress adaptation are lacking. Because the protein similarities of HAD family members are relatively few, it is difficult to search out all HAD family genes by direct blastP as PAPs [4]. Therefore, the hidden Markov model (HMM) of the HAD family was built and used to search against the rice and Arabidopsis protein sequences. Forty-one and 40 typical HAD proteins have been identified in rice and Arabidopsis, respectively. A further combination of bioinformatics and expression analysis showed that many HAD genes may be involved in intracellular and extracellular Po degradation under Pi stress conditions in plants.

Materials and methods

Identification of HAD proteins and motifs in rice and Arabidopsis

The whole protein sequences of Arabidopsis thaliana and rice were downloaded from the Ensembl database (http://plants.ensembl.org). Hidden Markov models (HMMs) of the HAD family were obtained from the Pfam protein database (http://pfam.xfam.org/) and constructed from reported plant HAD proteins [31]. The two hidden Markov models of the HAD domain
were used to search against the *Arabidopsis* and rice protein sequence data by HMMER software using default parameters (E-value = 0.5) [32]. The identified proteins were further examined by hand to verify whether the protein contained conserved motif I of HAD proteins. The final sequences were defined as HAD proteins and submitted to ExPASy (https://web.expasy.org) to calculate the isoelectric point and molecular weight [33]. Conserved motifs of the identified proteins were predicted by MEME using the following parameters: the search window of motifs was between 2–30 amino acids, and the motif number was set as 20 [34].

**Chromosomal location, gene structure, and synteny analysis**

The corresponding gene loci of the identified proteins were downloaded from the Ensembl database (http://plants.ensembl.org). All the gene loci were mapped on the *Arabidopsis* and rice chromosomes with Mapchart. Gene collinearity relationships were analysed by MCScanX [35, 36]. The collinear relationship of HAD genes from *Arabidopsis* and rice was finally drawn by Circos [35]. Introns and exons of all the gene loci were extracted from the rice comment files of the Ensembl database and drawn with TBtools [37].

**Phylogenetic analysis**

The full-length protein sequences of rice and *Arabidopsis* were aligned by MEGA-X using default parameters. The unrooted maximum likelihood method phylogenetic tree was constructed by MEGA-X using the Jones-Taylor-Thornton (JTT) model [38].

**Analysis of cis-elements in the promoters of HAD genes**

The promoter sequences (1.5 Kb) of rice HAD genes were downloaded from the Ensembl database (http://plants.ensembl.org) and were then submitted to the PlantCARE database for cis-element analysis [39]. The positions of cis-element P1BS and the W-box are marked on the promoters.

**Protein three-dimensional structure simulation**

Three-dimensional structures of the HAD proteins were built by SWISS-MODEL (https://swissmodel.expasy.org/) [40]. Targeted proteins were imported, and the template was searched by the software. Homologous modelling was used to simulate the 3D structure based on the template. Each model with a sequence similarity of greater than 30% was employed for the next step.

**Plant materials and growth condition**

Rice seeds (*Oryza sativa*, cv Nipponbare) were sterilized with 1% nitric acid and then germinated at 24 °C in the dark for two days. Uniform germinated seeds were grown in normal nutrient solution for 10 days and then transferred to nutrient solutions without Pi for 10 days, followed by two days of recovery in normal solution. The nutrient solution contained 1.425 mM NH₄NO₃, 0.998 mM CaCl₂·2H₂O, 0.323 mM NaH₂PO₄·2H₂O, 0.513 mM K₂SO₄, 1.643 mM MgSO₄·7H₂O, 9.5 μM MnCl₂·4H₂O, 0.075 μM (NH₄)₆Mo₇O₂₄·4H₂O, 19 μM H₂BO₃, 0.152 μM ZnSO₄·7H₂O, 0.155 μM CuSO₄·5H₂O, 0.125 mM FeSO₄·7H₂O, 0.125 mM Na₂EDTA·2H₂O, and 0.25 mM Na₂SiO₃·9H₂O, pH 5.5. The hydroponic experiments were carried out in a greenhouse with a 12/12-h photoperiod (200 μmol photons m⁻² s⁻¹) at 30/24 °C and 60% relative humidity. The solution was refreshed every three d.
**Quantitative real-time PCR analysis**

Total RNA was extracted from the shoots or roots of rice plants using TRIzol reagent (CoWin). Two micrograms of RNA were used to synthesize cDNA using a reverse transcription kit (CoWin) according to the manufacturer’s instructions. qRT-PCR experiments were performed on SYBR Green Real-Time PCR Master Mix reagent (YEASEN) dye on an Applied Biosystems Quantstudio™ Real-Time PCR System machine. The relative expression levels were calculated using the $2^{-\Delta\Delta C_T}$ method [41]. The primer sequences of the qRT-PCR are listed in S3 Table.

**Statistical analysis**

SPSS statistics base software (version 22) was used for statistical analysis. Significant differences were evaluated using one-way ANOVA and Tukey’s test.

**Results**

**Identification of HAD genes in rice and Arabidopsis**

To obtain the HAD proteins in rice and Arabidopsis, hidden Markov models of HAD were obtained from the Pfam protein database (PF12710) and constructed from reported plant HAD proteins (S1 Fig). The two different models of the HAD domain were used to search the whole protein sequence data of rice and Arabidopsis by HMMER [32]. Seventy-nine and 85 putative HAD proteins, which corresponded to 63 and 60 gene loci, were identified in rice and Arabidopsis, respectively (S2 Fig). Although the organization of HAD proteins is variable, the reactions catalysed by HAD require two core Asp residues in a DxD signature of motif I [42]. Then, only the sequences that contain the corresponding motif I are defined as typical plant HAD proteins (S2 Fig). In this way, 41 and 40 HAD genes were annotated as HAD coding genes in rice and Arabidopsis, respectively (S1 Table). The isoelectric points (pI) and molecular weight (MW) of the HAD proteins ranged from 4.65 to 10.78 and 19.7 to 117.7 kDa, respectively (S1 Table and S3 Fig).

**Phylogenetic and conserved motif analysis of the HAD proteins**

To study the evolutionary relationship between HAD family genes, the identified rice and Arabidopsis HAD protein sequences were aligned, and a phylogenetic tree was constructed by the maximum likelihood method. The 81 proteins were divided into three major groups (families I, II, and III) and seven subgroups (subfamilies Ia, Ib, Ic, IIa, IIb, IIIa, IIIb, and IIIC), each with more than 95% bootstrap support (Fig 1A). The rice and Arabidopsis HAD genes are named according to the phylogenetic tree from Ia to IIIc, which are OsHAD1 to OsHAD41 and AtHAD1 to AtHAD40, respectively (Fig 1A and S1 Table).

The motif variation of HAD proteins was further analysed by MEME and annotated with the Pfam database. Twenty motifs with the lowest E-values are listed (Table 1). Interestingly, five motifs (motifs 4, 7, 8, 12, and 19) were annotated as haloacid dehalogenase-like hydrolases or HAD superfamilies, while six other motifs (motifs 1, 5, 6, 16, 18 and 20) were annotated as phosphatases (Table 1). The functions of motifs 2, 3, 9, 10, 11, 13, and 17 are unknown. These 20 motifs were then mapped on each HAD protein. HAD proteins from the same subgroups have a similar motif arrangement, which supported the results of phylogenetic analysis (Fig 1B). Moreover, the number of conserved motifs ranges from one to 11 in different HAD members, which is consistent with the large divergence of HAD proteins.

Although the overall homology between HAD groups is very low, sequence comparisons have shown that all members of the HAD proteins contain four conserved motifs (Fig 2), which could be used to define HAD proteins in plants. To determine whether the HAD proteins are
Fig 1. Phylogenetic tree (A), conserved protein motifs (B), and gene structures (C) of HAD genes from rice and Arabidopsis. The phylogenetic tree of the HAD protein family in rice and Arabidopsis was constructed using MEGA-X software based on the maximum likelihood method and a JTT matrix-based model. Different colours represent different subgroups. The conserved motifs of the protein were analysed with MEME software and marked with different colours. For the gene structures, green represents untranslated 5’- and 3’-regions, and yellow represents exons.

https://doi.org/10.1371/journal.pone.0245600.g001
secreted, the signal peptides and subcellular localization were predicted by using TargetP 2.0. In total, only 11 HAD proteins contained a signal peptide in rice and Arabidopsis (S2 Table).

### Cap modules of HAD family in plants

In addition to the different conserved motifs, the HAD proteins were classified by their structure. Three different cap modules (C0, C1, and C2), which determine the substrate specificity,
have been found in the HAD superfamily. There were 2, 34, and 5 HAD proteins in rice with C0, C1, and C2 caps, respectively (S1 Table). Similarly, 2 C0, 35 C1, and 3 C2 caps were found among the HAD proteins in Arabidopsis (S1 Table).

Three-dimensional structures of OsHAD14, OsHAD24, and OsHAD25, which represent proteins of the C0, C1, and C2 caps, were simulated by SWISS-MODEL. As expected, all three types of proteins have a squiggle and a flap structure in their core catalytic domain, which are two key structural signatures of the HAD domain (Fig 3). The C0 cap of OsHAD25 has a small insertion between the two strands of the flap, while the C1 cap of OsHAD14 contains four alpha-helices in the middle of the flap motif (Fig 3A and 3B). In contrast, the C2 cap has a more complex insertion that contains four α-helices and five β-sheets (Fig 3C).

**Gene structure, chromosomal localization and collinearity analysis of HAD genes in rice and Arabidopsis**

Genomic sequences of the HAD genes were downloaded from the database and used to analyse the gene structures. There were great differences in the gene structures of the HAD genes, and the number of introns ranged from 0 to 33 (Fig 1C). Moreover, the HAD genes were unequally distributed on the chromosomes of rice and Arabidopsis (S4 Fig).

To explore the collinearity relationship of genes between species in the evolutionary process, BLASTP and MCScanX software were used to analyse the HAD genes of Arabidopsis and rice, and the results were visualized with Circos. Only two and four pairs of collinearity relationships between HAD genes in rice and Arabidopsis, respectively, were observed (Fig 4). There were another six pairs of collinearity between HAD genes of rice and Arabidopsis (Fig 4). The low collinearity of HAD genes indicates that large variations exist among the HAD genes.

**Pi-responsive cis-acting elements and the expression of HAD genes under different Pi supply conditions**

To explore whether the expression of HAD genes is regulated by Pi nutrients, a 1.5-kb promoter sequence was extracted to analyse cis-acting regulatory elements in rice. The promoters of 10 HAD genes contain the PiBS site (Fig 5), which is the binding site of PHR and PHR-like transcription factors [43]. Moreover, 23 promoters of HAD genes contain W-box sites (Fig 5), which are the binding sites of WRKY transcription factors [44]. Both the PHR and WRKY transcription factors are the key regulators of Pi starvation signalling in rice, indicating the putative relationship between HAD genes and Pi nutrients.

To further confirm whether the expression of HAD genes was regulated by Pi starvation, previously reported transcriptome data were reanalysed in rice [45]. The response of HAD genes to Pi nutrients can be divided into three different clusters (Fig 6). The first cluster contained eight genes, the expression of which showed a huge response to prolonged Pi starvation in roots. The re-supply of Pi for one day quickly recovered the Pi starvation-induced expression of these genes in roots. The expression of OsHAD2, OsHAD6 and OsHAD16 was also induced by Pi starvation by more than 10-fold in shoots (Fig 6). The expression of the second cluster of HAD genes did not respond to, or were suppressed by, Pi starvation conditions according to the transcript profiling. The third cluster of HAD genes showed a medium response to Pi starvation, while the expression of seven genes was induced by more than 2-fold under certain Pi starvation conditions in roots or shoots (Fig 6).
Fig 3. Simulated three-dimensional structures of typical HAD proteins in rice by SWISS-MODEL. (A) The three-dimensional structure of the OsHAD25 protein, which represents the C0 cap domain. The template accession No. is 3pg1.1 for OsHAD25. (B) The three-dimensional structure of the OsHAD14 protein, which represents the C1 cap domain. The template accession No. is 3l5k.1 for OsHAD14. (C) The three-dimensional structure of the OsHAD24 protein, which represents the C2 cap domain. The template accession No. is 1rko.1 for OsHAD24. The β strands are
qRT-PCR analysis of the transcripts of HAD genes under different Pi treatment conditions

Eight HAD genes were further selected, and their expression was measured by qRT-PCR under different Pi conditions. Among the eight selected HAD genes, OsHAD6, OsHAD16, OsHAD29, and OsHAD34 belong to cluster I genes, OsHAD12 and OsHAD28 belong to cluster II genes, and OsHAD20 and OsHAD24 belong to cluster III genes (Fig 6). Consistent with the transcription data, the expression of OsHAD6, OsHAD16, OsHAD29, and OsHAD34 was significantly

![Fig 4. Chromosomal distribution and interchromosomal relationship between HAD family genes in rice and Arabidopsis. The transparent lines indicate all the collinear blocks in the rice and Arabidopsis genomes. The red lines indicate duplicated HAD gene pairs. The chromosome names are marked on the outermost edge of the circle.](https://doi.org/10.1371/journal.pone.0245600.g004)
Fig 5. Distribution of the cis-element P1BS and W-box in the 1.5-kb regions upstream of ATG in the HAD genes of rice.

[Link](https://doi.org/10.1371/journal.pone.0245600.g005)
induced by prolonged Pi starvation in roots, while the expression of OsHAD6 and OsHAD16 was also induced by Pi starvation in shoots (Fig 7). The expression of OsHAD20 and OsHAD24 was induced by Pi starvation in roots and suppressed in the shoots by resupply Pi to Pi-starved plants (Fig 7). Although the transcription data showed that the transcripts of OsHAD12 and OsHAD28 did not respond to Pi starvation, a previous study detected the induced expression of OsHAD12 under Pi starvation conditions [27]. qRT-PCR proved that the expression of OsHAD12 and OsHAD28 was induced by Pi starvation in both roots and shoots (Fig 7).
Discussion

Identification and structural analysis of HAD proteins in plants

HAD proteins are composed of ATPase (~20%) and acid phosphatase (~79%), which exist in almost all organisms [46]. There are 28 and 183 HAD proteins from *E. coli* and humans, respectively [46]. Although a number of HAD domain proteins have been reported, no research has systematically analysed HAD proteins at the genomic level in plants. According to the hydrolysing substrates of HAD proteins, these enzymes can be divided into protein phosphatases, small molecule phosphatases, dehalogenases, phosphonatases and β-phosphoglucose mutases [47]. Although sequence comparisons found that the HAD domain contains four conserved sequence motifs, the similarities of whole sequences of HAD superfamily members are relatively low (<15%) [48]. Therefore, it is impossible to search out all the HAD family genes by the BLAST tool. In this study, the proteins containing the HAD domain were searched using hidden Markov models of HAD (S1 and S2 Figs). Among the conserved motifs of HAD proteins, a mutation of aspartic acid in motif I would lead to a significant reduction or even loss of enzyme activity [42]. Therefore, the identified proteins by HMMER software were further screened with motif I, which finally found 41 and 40 HAD genes from rice and *Arabidopsis*, respectively (S1 Table).

Using the identified HAD proteins from rice and *Arabidopsis*, HMM of the four conserved motifs was extracted with typical signatures. There are two conserved aspartic acids (D) in both motifs I and IV, which are involved in coordinating Mg$^{2+}$ in the active site (Fig 2). Motif II is characterized by a highly conserved serine or threonine (S/T), while motif III contains a conserved lysine (K) (Fig 2). Motifs II and III contribute to the stability of the reaction intermediates of the hydrolysis reaction [18]. In addition to the identified HAD proteins in this study, many other proteins were also annotated as HAD domain-containing proteins in the UniProt database. However, most of them lack critical motif I and were not analysed in this study.

Except for the four conserved motifs, the HAD superfamily shares common three-dimensional structures [48]. All HAD domains have a squiggle and flap in the Rossmannoid fold,
which is distinguished from the Rossmannoid fold in other types of phosphatases [18]. Moreover, different kinds of cap structures insert in the middle of the flap structure or adjacent to motif III. The cap structures surround the catalytic centre and function to adjust substrate specificity [49, 50]. Therefore, understanding the cap structure is very important for studying the substrate specificity of the HAD protein. Among the 41 HAD genes in rice, two are C0 genes, five are C2 genes, and the rest are C1 cap structural genes (S1 Table). The cap structure has a great influence on the substrate specificity of the enzyme. Mutation of the amino acids (ΔF53, ΔN54, ΔN55) in the cap structure will reduce the kinetics of AtHAD15 [51]. The substrate specificity of HAD proteins with C1 and C2 cap structures is lower than that of C0, which may be related to the specific enzyme residues of caps that can bind to different substrates [50].

A number of HAD phosphatases are involved in intracellular and extracellular Pi recycling

Eighty-one HAD proteins were found in this research and divided into seven subgroups (Ia, Ib, Ic, II, IIIa, IIIb, and IIIc). Among the 20 most conserved motifs in HAD proteins of plants, six are annotated as putative phosphatases, which is in accordance with the phosphoryl transfer activity of the HAD domain (Table 1). Interestingly, the promoters of 28 HAD genes contain P1BS or W-box in rice, indicating that these HAD genes may be regulated by Pi starvation (Fig 5). Combining transcriptome data and qRT-PCR analysis showed that the expression of at least 17 HAD genes was induced by Pi starvation in shoots or roots (Figs 6 and 7). These Pi-responsive HAD genes may participate in the regulation of Pi stress adaptation responses in rice.

It has been reported that plants synthesize intercellular and extracellular phosphatases to recycle different Po under Pi starvation conditions [16]. The majority of the Pi starvation-induced secreted acid phosphatases in plants are PAPs, which belong to a unique acid phosphatase subfamily. Interestingly, AtVPS3 (AtHAD37) was recently identified as a novel secreted acid phosphatase isomorph under Pi starvation conditions in Arabidopsis [26]. Moreover, AtVSP1 (AtHAD34) and AtVSP2 (AtHAD33) also encode Pi starvation-induced secreted acid phosphatases in Arabidopsis [52]. These VSP proteins form subgroup IIIc HADs, which have eight and three HAD members in Arabidopsis and rice, respectively (Fig 1). Subcellular localization prediction showed that all the subgroup IIIc HADs contained a secreted signal peptide (S2 Table). Similar to Arabidopsis, the expression of OsHAD39 and OsHAD41 is induced by Pi starvation in roots, indicating that they may be involved in recycling extracellular Po under Pi starvation conditions. In contrast to the subgroup IIIc HAD proteins, other groups of HAD phosphatases do not contain a secreted signal peptide and may be involved in utilizing intercellular Po in plants.

Notably, seven of the 10 subgroup IIIa HAD genes were induced by Pi starvation in rice (Figs 6 and 7). These HAD genes belong to the glycerol-3-phosphate acyltransferase (GPAT) family, which participates in the acylation reaction at the sn-2 position of glycerol-3-phosphate (G3P) to form sn-2 lysophosphatidic acid (sn-2 LPA) and further dephosphorylation of sn-2 LPA [53, 54]. These GPATs may be involved in recycling Pi from G3P under Pi stress conditions. In Arabidopsis, AtGPAT1–8 (AtHAD23–28) are required for the synthesis of cutin and suberin [54]. Interestingly, the expression of OsHAD27, OsHAD29, OsHAD33, and OsHAD34 was induced more than fivefold after 21 days of Pi starvation in roots (Fig 6). These genes may participate in the synthesis of suberin in roots under Pi-deficient conditions [54].

The cell membrane systems are composed of phospholipids, which are one of the most important Po pools in plants. In Arabidopsis, Pi starvation induced the expression of the AtNPC4 and AtNPC5 genes to degrade phosphatidyl-choline and phosphatidyl-ethanolamine (PC/PE) to phosphocholine and phosphoethanolamine. AtPS2 (or AtHAD1) and AtPECP1 (or
AtHAD2), which belong to the HAD Ia subgroups, further dephosphorylate phosphocholine and phosphoethanolamine to release Pi [22–25]. The gene expression of OsHAD1 and OsHAD2, which are close homologs of AtHAD1 and AtHAD2, was significantly induced under Pi stress conditions, indicating that a similar function of OsHAD1 and OsHAD2 may exist in rice. However, mutations in AtHAD1 and AtHAD2 do not influence Pi homeostasis or Pi starvation responses in Arabidopsis. The functions of OsHAD1 and OsHAD2 in Pi starvation adaptations need further analysis.

Supporting information
S1 Fig. Hidden Markov models (HMMs) of HAD proteins from the Pfam database (A) and constructed by reported plant HAD proteins (B).
(PDF)
S2 Fig. Sequence search flowchart of the HAD proteins and genes in rice and Arabidopsis.
(PDF)
S3 Fig. Molecular Weight (MW) of the HAD proteins in rice and Arabidopsis.
(PDF)
S4 Fig. Chromosomal location of the HAD genes in rice and Arabidopsis.
(PDF)
S1 Table. Gene information of the HAD superfamily.
(XLSX)
S2 Table. Secreted signal peptide prediction of HAD family proteins in Arabidopsis and rice.
(XLSX)
S3 Table. Primers used in this study.
(XLSX)

Acknowledgments
We thank Pro. Fangseng Xu, Pro. Lei Shi and Pro. Guangda Ding for the suggestions and corrections of the manuscript.

Author Contributions
Conceptualization: Chuang Wang.
Data curation: Zezhen Du.
Formal analysis: Zezhen Du, Suren Deng, Zixuan Wu.
Funding acquisition: Chuang Wang.
Software: Zezhen Du.
Writing – original draft: Zezhen Du.
Writing – review & editing: Suren Deng, Zixuan Wu, Chuang Wang.

References
1. Chiou TJ, Lin SI. Signaling network in sensing phosphate availability in plants. Annu Rev Plant Biol. 2011; 62:185–206. https://doi.org/10.1146/annurev-arplant-042110-103649 PMID: 21370979
Deng S, Lu L, Li J, Du Z, Liu T, Li W, et al. Purple acid phosphatase 10c (OsPAP10c) encodes a major

Liu P, Cai Z, Chen Z, Mo X, Ding X, Liang C, et al. A root-associated purple acid phosphatase,

Bozzo GG, Raghothama KG, Plaxton WC. Structural and kinetic properties of a novel purple acid phos-

tase Secretion into the Rhizosphere of Proteoid Roots from Phosphorus-Stressed White Lupin. Plant

Tran HT, Hurley BA, Plaxton WC. Feeding hungry plants: The role of purple acid phosphatases in phos-

Bozzo GG, Dunn EL, Plaxton WC. Differential synthesis of phosphate-starvation inducible purple acid phos-

tase, increased extracellular ATP utilization in rice. J Exp Bot. 2020; 71(14), 4321–4332. https://doi.org/10.1093/jxb/eraa179 PMID: 32270183

Puga MI, Rojas-Triana M, de Lorenzo L, Leyva A, Rubio V, Paz-Ares J. Novel signals in the regulation

Morais M, Zhang WH, Baker AS, Zhang GF, Mariano DD, Allen KN. The Crystal Structure of Bacillus cereus Phosphonoacetaldehyde Hydrolase Insight into Catalysis of Phosphorus Bond Cleavage and Catalytic Diversification within the HAD Enzyme Superfamily. Biochemistry. 2000; 39:10385–96. https://doi.org/10.1021/bi001171j PMID: 10956028

PMID: 9584613

A root-associated purple acid phosphatase, increased extracellular organic phosphorus utilization in rice. J Integr Plant Biol. 2020; 71(14), 4321–4332. https://doi.org/10.1093/jxb/eraa179 PMID: 32270183

Shen J, Yuan L, Zhang J, Li H, Bai Z, Chen X, et al. Phosphorus dynamics: from soil to plant. Plant Physiol. 2011; 156(3):997–1005. https://doi.org/10.1104/pp.111.175232 PMID: 21571668

Puga M, Rojas-Triana M, de Lorenzo L, Leyva A, Rubio V, Paz-Ares J. Novel signals in the regulation of Pi starvation responses in plants: facts and promises. Curr Opin Plant Biol. 2017; 39:40–9. https://doi.org/10.1016/j.pbi.2017.05.007 PMID: 28587933

Zhang Q, Wang C, Tian J, Li K, Shou H. Identification of rice purple acid phosphatases related to phos-

Bozzo GG, Dunn EL, Plaxton WC. Differentiation synthesis of phospha te-starvation inducible purple acid phos-

Veljanovski V, Vanderbeld B, Knowles VL, Snedden WA, Plaxton WC. Biochemical and molecular char-

Miller SS, Liu J, Allan DL, Menzhuber CJ, Fedorova M, Vance CP. Molecular Control of Acid Phospha-

Koonin EV, Tatusov RL. Computer Analysis of Bacteri al Haloacid Dehalogenases Defines a Large

Shen J, Yuan L, Zhang J, Li H, Bai Z, Chen X, et al. Phosphorus dynamics: from soil to plant. Plant Physiol. 2011; 156(3):997–1005. https://doi.org/10.1104/pp.111.175232 PMID: 21571668

Puga MI, Rojas-Triana M, de Lorenzo L, Leyva A, Rubio V, Paz-Ares J. Novel signals in the regulation of Pi starvation responses in plants: facts and promises. Curr Opin Plant Biol. 2017; 39:40–9. https://doi.org/10.1016/j.pbi.2017.05.007 PMID: 28587933

Zhang Q, Wang C, Tian J, Li K, Shou H. Identification of rice purple acid phosphatases related to phos-

Bozzo GG, Raghothama KG, Plaxton WC. Structural and kinetic properties of a novel purple acid phos-

Tran HT, Hurley BA, Plaxton WC. Feeding hungry plants: The role of purple acid phosphatases in phos-

Bozzo GG, Raghothama KG, Plaxton WC. Structural and kinetic properties of a novel purple acid phos-

Veljanovski V, Vanderbeld B, Knowles VL, Snedden WA, Plaxton WC. Biochemical and molecular char-

Miller SS, Liu J, Allan DL, Menzhuber CJ, Fedorova M, Vance CP. Molecular Control of Acid Phospha-

Koonin EV, Tatusov RL. Computer Analysis of Bacteri al Haloacid Dehalogenases Defines a Large
21. May A, Spinak M, Kock M. Arabidopsis thaliana PECP1: enzymatic characterization and structural organization of the first plant phosphoethanolamine/phosphocholine phosphatase. Biochim Biophys Acta. 2012; 1824(2):319–25. https://doi.org/10.1016/j.bbapap.2011.10.003 PMID: 2204570

22. Angkawijaya AE, Nakamura Y. Arabidopsis PECP1 and PS2 are phosphate starvation-inducible phosphoethanolamine phosphatases. Biochem Biophys Res Commun. 2017; 494(1–2):397–401. https://doi.org/10.1016/j.bbrc.2017.09.094 PMID: 28942147

23. Hanchi M, Thibaud MC, Legeret B, Kuwata K, Pochon N, Beisson F, et al. The Phosphate Fast-Responsive Genes PECP1 and PPsPase1 Affect Phosphocholine and Phosphoethanolamine Content. Plant Physiol. 2018; 176(4):2943–62. https://doi.org/10.1104/pp.17.01246 PMID: 29475899

24. Angkawijaya AE, Ngo AH, Nguyen VC, Gunawan F, Nakamura Y. Expression Profiles of 2 Phosphate Starvation-Inducible Phosphoethanolamine/Phosphocholine Phosphatases, PECP1 and PS2, in Arabidopsis. Front Plant Sci. 2019; 10:662. https://doi.org/10.3389/fpls.2019.00662 PMID: 31191788

25. Tannert M, May A, Diffe D, Berger S, Balcke GU, Tissier A, et al. Pi starvation-dependent regulation of ethanolamine metabolism by phosphoethanolamine phosphatase PECP1 in Arabidopsis roots. J Exp Bot. 2018; 69(3):467–81. https://doi.org/10.1093/jxb/exy408 PMID: 29294054

26. Sun L, Wang L, Zheng Z, Liu D. Identification and characterization of an Arabidopsis phosphate starvation-induced secreted acid phosphatase as a vegetative storage protein. Plant Sci. 2018; 277:278–84. https://doi.org/10.1016/j.plantsci.2018.09.016 PMID: 30466593

27. Pandey BK, Mehra P, Verma L, Bhadouria J, Giri J. OsHAD1, a Halaocid Dehalogenase-Like APase, Enhances Phosphatase Accumulation. Plant Physiol. 2017; 174(4):2316–32. https://doi.org/10.1104/pp.17.00571 PMID: 28637831

28. Baldwin JC, Karthikeyan AS, Cao A, Raghothama KG. Biochemical and molecular analysis of LePS2;1: a phosphate starvation induced protein phosphatase gene from tomato. Planta. 2008; 228(2):273–80. https://doi.org/10.1007/s00425-007-0736-y PMID: 18458947

29. Liang CY, Chen ZJ, Yao ZF, Tian J, Liao H. Characterization of two putative protein phosphatase genes and their involvement in phosphorus efficiency in Phaseolus vulgaris. J Integr Plant Biol. 2012; 54(6):400–11. https://doi.org/10.1111/j.1744-7909.2012.01126.x PMID: 22571280

30. Cai Z, Cheng Y, Xian P, Ma Q, Wen K, Xia Q, et al. Acid phosphatase gene GmHAD1 linked to low phosphorus tolerance in soybean, through fine mapping. Theor Appl Genet. 2018; 131(8):1715–28. https://doi.org/10.1007/s00122-017-2837-x PMID: 29754326

31. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019; 47(D1):D427–D32. https://doi.org/10.1093/nar/gky995 PMID: 30357350

32. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. Bmc Bioinformatics. 2010; 11:431. https://doi.org/10.1186/1471-2105-11-431 PMID: 20718988

33. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999; 112:531–52. https://doi.org/10.1385/1-59275-584-7:531 PMID: 10027275

34. Baldwin JC, Karthikeyan AS, Cao A, Raghothama KG. Biochemical and molecular analysis of LePS2;1: a phosphate starvation induced protein phosphatase gene from tomato. Planta. 2008; 228(2):273–80. https://doi.org/10.1007/s00425-007-0736-y PMID: 18458947

35. Cai Z, Cheng Y, Xian P, Ma Q, Wen K, Xia Q, et al. Acid phosphatase gene GmHAD1 linked to low phosphorus tolerance in soybean, through fine mapping. Theor Appl Genet. 2018; 131(8):1715–28. https://doi.org/10.1007/s00122-017-2837-x PMID: 29754326

36. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012; 40(7):e49. https://doi.org/10.1093/nar/gks1293 PMID: 2217600

37. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools—an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020. https://doi.org/10.1016/j.molp.2020.06.009 PMID: 32585190

38. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018; 35(6):1547–9. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

39. Lescom P, Déhaes P, Thijs G, Marchal K, Moreau Y, Peer YY, et al. PlantCARE: a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002; 30:325–7. https://doi.org/10.1093/nar/gqf339 PMID: 11752327

40. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018; 46(W1):W296–W303. https://doi.org/10.1093/nar/gky427 PMID: 29788355
41. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25(4):402–8. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609

42. Collet JF, Stroobant V, Schaftingen EV. Mechanistic Studies of Phosphoserine Phosphatase, an Enzyme Related to P-type ATPases. J biol chem. 1999; 274:33985–90. https://doi.org/10.1074/jbc.274.48.33985 PMID: 10567362

43. Rubio V, Linhares F, Solano R, Martin AC, Iglesias J, Leyva A, et al. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. Genes Dev. 2001; 15(16):2122–33. https://doi.org/10.1101/gad.204401 PMID: 11511543

44. Wang H, Xu Q, Kong YH, Chen Y, Duan JY, Wu WH, et al. Arabidopsis WRKY45 transcription factor activates PHOSPHATE TRANSPORTER1;1 expression in response to phosphate starvation. Plant Physiol. 2014; 164(4):2020–9. https://doi.org/10.1104/pp.113.235077 PMID: 24586044

45. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, et al. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife. 2015; 4. https://doi.org/10.7554/eLife.09343 PMID: 26196146

46. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamly. Curr Opin Struct Biol. 2009; 19(6):658–65. https://doi.org/10.1016/j.sbi.2009.09.008 PMID: 19889535

47. Kuznetsova E, Proudfoot M, Gonzalez CF, Brown G, Omelchenko MV, Borozan I, et al. Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. J Biol Chem. 2006; 281(47):36149–61. https://doi.org/10.1074/jbc.M605449200 PMID: 16990279

48. Seifried A, Schultz J, Gohlke A. Human HAD phosphatases: structure, mechanism, and roles in health and disease. FEBS J. 2013; 280(2):549–71. https://doi.org/10.1111/j.1742-4658.2012.08633.x PMID: 22607316

49. Lahiri SD, Zhang GF, Dai JY, Dunaway-Mariano D, Allen KN. Analysis of the substrate specificity Loop of the HAD Superfamily Cap Domain. Biochemistry. 2004; 43:2812–20. https://doi.org/10.1021/bi0356810 PMID: 15005616

50. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, et al. Panoramic view of a superfamily of phosphatases through substrate profiling. Proc Natl Acad Sci USA. 2015; 112(16):E1974–83. https://doi.org/10.1073/pnas.1423570112 PMID: 25848029

51. Caparros-Martín JA, McCarthy-Suárez I, Culianez-Macia FA. Sequence Determinants of Substrate Ambiguity in a HAD Phosphosugar Phosphatase of Arabidopsis Thaliana. Biology (Basel). 2019; 8(4). https://doi.org/10.3390/biology8040077 PMID: 31601035

52. Liu Y, Ahn JE, Datta S, Salzman RA, Moon J, Huyghues-Despointes B, et al. Arabidopsis vegetative storage protein is an anti-insect acid phosphatase. Plant Physiol. 2005; 139(3):1545–56. https://doi.org/10.1104/pp.105.066837 PMID: 16258019

53. Yang W, Pollard M, Li-Beisson Y, Beisson F, Feig M, Othrogge J. A distinct type of glycerol-3-phosphate acyltransferase with sn-2 preference and phosphatase activity producing 2-monoaoylglycerol. Proc Natl Acad Sci USA. 2010; 107(26):12040–5. https://doi.org/10.1073/pnas.0914149107 PMID: 20551224

54. Yang W, Simpson JP, Li-Beisson Y, Beisson F, Pollard M, Othrogge JB. A land-plant specific glycerol-3-phosphate acyltransferase family in Arabidopsis: substrate specificity, sn-2 preference, and evolution. Plant Physiol. 2012; 160(2):638–52. https://doi.org/10.1104/pp.112.201996 PMID: 22864585