*Pak. J. Bot.*, 54(4): 1365-1373, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-4(32)

# CLONING AND BIOINFORMATICS ANALYSIS OF PURPLE ACID PHOSPHATASE GENE, *PAP1*, FROM *AEGILOPS* SPECIES

**LIJIANG HOU[1,2], XIAOYU FENG[2], YUNZE LU[3], QIUFANG WU[1], XINQIANG GAO[1], JIAN DONG[2*] AND JUNWEI WANG[2*]**

[1]*School of Biological and Food Engineering, Anyang Institute of Technology, Anyang, Henan, 455000, China*
[2]*College of Agronomy, Northwest A&F University, Yangling, Shaanxi, 712100, China*
[3]*School of Landscape and Ecological Engineering, Hebei University of Engineering, Handan, Hebei, 056038, China*
**Corresponding author's email: wjw@nwsuaf.edu.cn; dj4322@163.com*

**Abstract**

Phosphorus (P)-starvation is one of the abiotic stresses that can adversely affect plants, causing reductions in their potential yield. In order to discover more P-starvation tolerant germplasm resources and associated candidate genes, we studied both the isolation and characterization of the key phosphorus catalytic enzyme gene *PAP1* in *Aegilops*, an inherently abiotic stress tolerant plant genus. Comprehensive phenotypic observations made under a P-starvation treatment evinced that the stress tolerance was stronger in *Aegilops crassa* than *Aegilops ventricosa*. Further, the CDS (coding DNA sequence) *PAP1* from the above *Aegilops* genotypes were cloned; this revealed that both *AePAP1s* encode low-molecular weight PAPs, much like mammalian PAPs (35 kD). They were highly expressed in roots and constitutively expressed under phosphorus (P) deprivation. The DNA sequence alignment uncovered the presence of 42 single nucleotide polymorphisms (SNPs) between *AeVePAP1* and *AeCrPAP1*, including 20 synonymous SNPs and 22 non-synonymous SNPs. According to the amino acid sequence alignment, the two AePAP1 proteins possessed unique, seven metal-binding residues as well as five conserved motifs of the PAPs family. The constructed phylogenic tree showed that the two AePAP1s from *Aegilops* shared a tight evolutionary relationship to other mammal-like PAPs from proximal species of *Triticum aestivum*. Finally, the signal peptide and subcellular localization analysis predicted that both *AeVePAP1*and *AeCrPAP1*could play a role in extracellular P scavenging in addition to P mobilization from the rhizosphere.

**Key words:** *Aegilops*, Purple acid phosphotase, P-starvation tolerance.

## Introduction

Phosphorus (P) is an essential nutrient element, participating in various physiological and biochemical processes, whose limited availability and uptake can influence the normal growth and development of plants (Hou *et al*., 2018; Tran *et al*., 2010; Vance *et al*., 2003). Although the P content in soil is high, the amount of available P that can be absorbed and utilized by plants is actually quite low and insufficient (Rehman, 2019). Research has shown that plants' utilization rate of phosphate fertilizer in the growing season is very low, at only 10%–25% (Fang *et al*., 2009). The annual application of P fertilizer not only pollutes groundwater but also accelerates the depletion of P resources (Richardson, 2009; Vance *et al*., 2003). Currently, more and more experts are suggesting that improving both the P utilization rate and the P-starvation tolerance of crops is an effective way for ensuring food security.

Acid phosphatases constitute a class of key enzymes, capable of catalyzing the hydrolysis of P from a broad and overlapping range of P monoesters with an acidic pH optimum, and which function in the production, transport, and recycling of P in plants (Raghothama, 1999). Purple acid phosphatases (PAPs) are a kind of special acid phosphatase, which belong to a metallophosphoesterase superfamily that includes phosphoprotein phosphatases and exonucleases. All PAPs possess seven metal-binding residues (D, D, Y, N, H, H, H) and five motifs (**D**xG-G**D**x$_2$**Y**-G**N**H(E,D)-Vx$_2$**H**-G**H**x**H**; (Strater *et al*., 1995), which form dimetallic active sites, being highly conserved amongst bacterial, mammalian, and plant purple acid phosphatases (Li *et al*., 2002; Olczak *et al*., 2003). In addition, PAPs also contain a binuclear metal center generally composed of $Fe^{3+}$-$Zn^{2+}$ or $Fe^{3+}$-$Mn^{2+}$ in plants (Klabunde *et al*., 1995; Olczak *et al*., 2003). According to the molecular weight of the corresponding protein, PAPs may be divided into two sub-families, namely low molecular weight PAPs (LMW PAPs, also termed mammal-like PAPs; about 35 kD) and high molecular weight PAPs (HMW PAPs; about 55 kD). Most HMW PAPs exist as homodimers in plants (Duff *et al*., 2010; Lebansky *et al*., 1992; Olczak *et al*., 2003; Olczak & Wtorek, 2010), whereas LMW PAPs occur as monomeric proteins lacking the equivalent of the NH2-domain of KBPAP, thus making them structurally similar to mammalian PAPs (Li *et al*., 2002). Many PAPs are glycoproteins that target the secretary pathway (Schenk *et al*., 2008) and play a key role in the P's scavenging and recycling from P esters and anhydrides under conditions of P starvation (Li *et al*., 2002; Olczak *et al*., 2003; Ravichandran *et al*., 2013; Robinson *et al*., 2012).

Many PAPs have been biochemically characterized from several plant species. The first PAP gene described (*GmPhy*) was isolated from soybean and found expressed in the cotyledons of germinating seedlings (Hegeman & Grabau, 2001). Later, a PAP cDNA was from *Medicago truncatula*, whose expression occurred in leaves and in roots as secreted enzymes that contributed to the acquisition of organic phosphorus (Xiao *et al*., 2005). The Arabidopsis purple acid phosphatase (AtPAP) family is encoded by 29 genes, but only 7 of its members (AtPAP7–AtPAP13) were responsive to phosphate deprivation (Li *et al*., 2002), of which only AtPAP15 and AtPAP23 possess phytase activities (Kuang *et al*., 2009; Zhu *et al*., 2005).

Furthermore, two phytases have been identified in mature grains of wheat (PHYI, approximately 66 kD; PHYII, approximately 68 kD), barley (P1 and P2, both 66 kD), and rice (F1, 66 kD; F2, 68 kD), all of which were synthesized and stored during grain filling (Dionisio *et al*., 2011; Greiner *et al*., 2000; Hayakawa *et al*., 1989; Nakano *et al*., 1999). Moreover, when *AtPAP15* was transferred into soybeans, it was able to enhance the efficiency utilization of P and increased the yield of soybeans (Wang *et al.*, 2009; Wu *et al.*, 2013). After *MtPAP1* was isolated and cloned from *Medicago*, it prompted the high-efficiency utilization of P once transformed into Arabidopsis (Xiao *et al.*, 2006; Xiao *et al.*, 2005). Yet, a P-starvation tolerance evaluation of *Aegilops* is lacking, and functional studies of *Aegilops PAP1* are also limited.

In this study, the P-starvation tolerance of two *Aegilops* species was systematically evaluated. Building on this, we cloned and compared the differences in sequence, physicochemical properties, protein structure, function, and evolution of the two *AePAP1* genes from *A. ventricosa* and *A. crassa* via bioinformatics methods. Our comprehensive study provides fundamental ecophysiological data for further explaining the P-starvation tolerance mechanism of *Aegilops*.

## Material and Methods

**Plant materials, growth conditions, and P-starvation treatment:** *Aegilops ventricosa* and *Aegilops crassa* were used in this study. Their seeds were sterilized and germinated in water-soaked filter papers at 4°C for 3 days, then transferred into plastic pots (diameter:15.5cm, 6seedlings per pot, and 10 pots per species) filled with quartz sand (not containing P) and grown under glasshouse conditions(temperature of 22°±2°C, relative humidity of 70%–75%, under 12-h-daylight photoperiod).Once the seedlings had grown for7 days, to them we applied a normal (P+, containing 2 mM Pi) or a P-deficient (P–) 4×diluted Hoagland's solution for 30 days, with n = 30 seedlings (5 pots) under each treatment. Then, 5 seedlings from each treatment were collected in liquid nitrogen and stored at –80°C until further use. The remaining seedlings (n=25) were used for phenotypic observations, and a statistical analysis was performed on measured root traits (total root length, average diameter, and lateral root number) and the dry weight of areal part.

**Statistical analysis of root measurements and dry weight of areal part:** Total root length, average diameter, and lateral root number of *Aegilops* seedlings were assessed in at least three independent experiments. For this, the roots of the seedlings were cleaned after the two treatments, and then digitally scanned and saved them as root images. These three root traits were measured using the plant root analyzer program, WinRHIZO (Hangxin, Guangzhou, China). The dry weight of areal part was taken on an electronic balance (accuracy: 0.0001 g, FA1004; Fangrui, Shanghai, China), after first oven-drying the areal part of the seedlings. To compare the means of the two treatments, Student's *t*-tests were performed in all cases.

**RNA isolation and cDNA synthesis:** Total RNA was extracted from *Aegilops* seedlings with the RNAprep pure plant kit (TIANGEN, Redwood City, CA, USA)and RNA quality assessed by an NanoDrop 2000 (Thermo Scientific, USA).The cDNA was prepared from 1mg of total RNA, by using the Prime Script™ II 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China).

**Cloning the purple phosphatase gene(s)*PAP1s*:** The cloning primers (Table 1) were designed in DNAMAN 6.0software (Lynnon Biosoft, USA), according to the cDNA sequence of the wheat *TaPAP1* gene (NCBI accession number: AB518866) and *BrachypodiumPAP1* gene (NCBI accession number: XM_003558353).The cDNAs of *Aegilops* seedlings (P deficiency treatment) were used as templates for the PCR amplifications, done in a 25-μL reaction system consisting of 2 μL of cDNA template, 2.5 μL of 10×PCR Buffer, 0.5 μL of dNTP (2 mmol•$L^{-1}$), 1.5 μL of each upstream and downstream primers (10 mmol•$L^{-1}$), 0.2 μL of Taq DNA polymerase (5 U•$\mu L^{-1}$), and topped with dd$H_2O$ to 25 μL. The PCR reaction program went as follows: pre-denaturation at 95°Cfor 3 min; then, 35 cycles at 94°C for 30 s, 56°Cfor 30 s, and 72°Cfor 1 min, followed by an extension at 72°Cfor10 min, and storage at 4°C. The amplified product was detected using1% agarose gel. Then, the purified product was ligated to pMD18-T vector and transferred into *E.coli* DH5α for sequencing.

**Analysis of CDS sequences and protein characteristics:** CDS sequences were used for ORF prediction, preformed with the ExPASy software tool (http://us.expasy.org/tools/dna.html). Different protein parameters, such as amino acid composition, isoelectric point (PI), and hydropathicity index, were estimated for AeVePAP1 and AeCrPAP1, in DNAMAN software (Nicolas, 2001).

**Secondary structure and 3D structure prediction:** Prediction of amino acid composition, α-helices, β-strand and coils was estimated for the AePAP1 protein, using DNAMAN software. The 3D protein structure prediction was carried out using the SWISS-MODEL tool (https://swissmodel.expasy.org/). Signal peptide analysis was estimated using theSignalP4.1 server (http://www.cbs.dtu.dk/services/SignalP-4.1/) while the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM) was used for transmembrane predictions. Subcellular localization analysis was implemented in the PSORT II server (https://psort.hgc.jp/form2.html).

**Phylogenetic analysis of homologous proteins:** Multiple sequence alignment of AeVePAP1 and AeCrPAP1 with the known PAP1s of other plant species was performed using Clustal W server (https://embnet.vital-it.ch/software/ClustalW.html), under its default parameters. For the phylogenetic relationship analysis, the neighbor-joining method and a Poisson model were used, in MEGA v6.0 software. The internal node stability of the phylogenetic tree was assessed by the bootstrap value of 1000 replicates.

**Table 1. List of primers.**

| Primers | Sequences (5'-3') | Target | Used for |
|---|---|---|---|
| PAP1-F | CGGCCAGCAGCCATGGCGAG | AePAP1 | Gene clone |
| PAP1-R | CCTCATTYTKCAGTGACATA | AePAP1 | Gene clone |
| RtPAP1-F | ATGACTACAGGGGTGATGCG | AePAP1 | Realtime PCR |
| RtPAP1-R | GTCTTTGGGGTGAGTCCAGT | AePAP1 | Realtime PCR |
| Actin-F | CTCCCTCACAACAACCGC | ACTIN | Realtime PCR |
| Actin-R | TACCAGGAACTTCCATACCAAC | ACTIN | Realtime PCR |

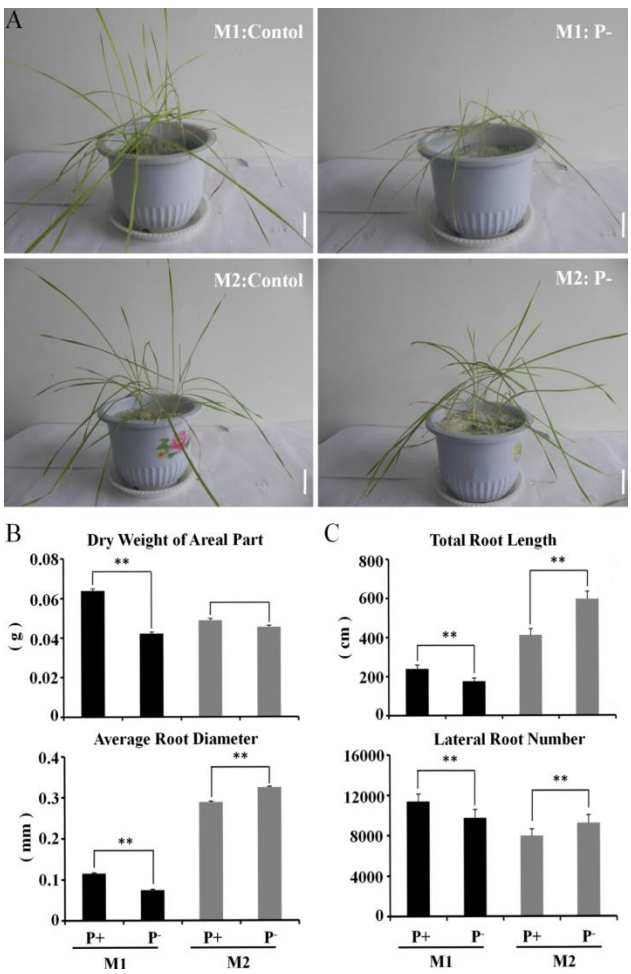


Fig. 1. The phenotypes of two *Aegilops* species under phosphorus starvation treatment.
(A) The phenotypes of 40-day-old *Aegilops* seedlings grown in pot with phosphorus deficiency (P-) solution treatment for 30 days and normal (P+, contain 2 mM Pi) solution as control. M1: *Aegilops ventricosa*; M2: *Aegilops crassa*. Scale bar, 1 cm.
(B) The dry weight of areal part, average root diameter of seedlings grown as described in A (mean ± s.d., n ≥ 30). The asterisks indicate a statistically significant difference (*t*-test, ∗ $p<0.05$, ∗∗ $p<0.01$). The data are from three repeat experiments. M1: *Aegilops ventricosa*; M2: *Aegilops crassa*.
(C) The total root length, lateral root number of seedlings grown as described in A (mean ± s.d., n ≥ 30). The asterisks indicate a statistically significant difference (*t*-test, ∗ $p<0.05$, ∗∗ $p<0.01$). The data are from three repeat experiments. M1: *Aegilops ventricosa*; M2: *Aegilops crassa*.

**Quantitative real-time PCR:** The 7-day-old seedlings of two *Aegilops*, which were sown in plastic pots filled with quartz sand, were treated with the P-deficient solution (P–) for 6 hour and a normal solution (P+, contain 2 mM Pi) treatment that served as the control. Then, their leaf and root tissues were snap-frozen in liquid nitrogen prior to RNA extractions, which was carried out with the RNAprep Pure Plant Kit (TIANGEN). The Poly(dT) cDNA was prepared from 1 μg of total RNA with the reverse transcriptase PrimeScriptTMRTase (TaKaRa), then analyzed on a 7300 Real Time PCR System (Applied Biosystems, Foster City, CA, USA) with SYBR Premix Ex Taq II (TaKaRa) according to the manufacturer's instructions. Transcripts of *PAP1* gene(s) were quantified using specific primer pairs designed in Primer Premier 5.0 software (Table 1). Every individual reaction was performed with three biological replicates.

## Results

**P-starvation tolerance evaluation of *A. ventricosa* and *A. crassa*:** To evaluate the P-starvation tolerance of the two *Aegilops* species, a P-deficient solution was applied to *Aegilops* seedlings. Significant alterations to the areal part's phenotype were observed in *A. ventricosa* seedlings (Fig. 1A), and the dry weight of the areal part (Fig. 1B), average root diameter (Fig. 1B), total root length (Fig. 1C), and lateral root number (Fig. 1C) all decreased significantly in *A. ventricosa* seedlings under the P deficiency treatment. By contrast, the *A. crassa* seedlings' areal parts were evidently unchanged by the treatment (Fig. 1A, 1B), and its average root diameter (Fig. 1B), total root length (Fig. 1C), and lateral root number (Fig. 1C) all increased significantly in those starved of P. These phenotype results strongly suggested that the P-starvation tolerance of *A. crassa* was stronger than that of *A. ventricosa*.

**Amplification and cloning of *AePAP1* from *Aegilops*:** To analyze the different sites of the P catalytic key enzyme gene-*PAP1*in *A. ventricosa* and *A. crassa*, their seedlings' total RNA was extracted, and the synthesized cDNA from the extracted total RNA was used as a template for PCR amplification of the *AePAP1* CDS sequence. The amplified amplicon of 1.008 kb was detected on 1% agarose gel, near the1-kb band of DNA marker (Fig. 2). The colony-PCR of recombinant cells also showed a specific single band (1 kb) equal to the size of *AePAP1* CDS.

Our sequencing results showed that the cDNA sequences of the cloned genes were highly homologous with the wheat *PAP1* gene (*TaPAP1*, accession number: AB51886), and the shared identity between the two *AePAP1s* and *TaPAP1* reached 96.13%. So, the two cloned cDNA sequences from *A. ventricosa* and *A. crassa* were accordingly named *AeVePAP1* and *AeCrPAP1*, respectively.

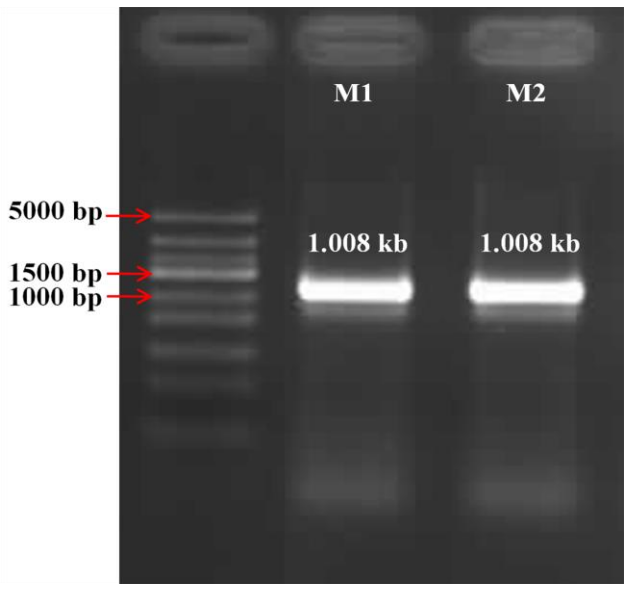


Fig. 2. PCR amplification of *PAP1* cDNA sequences from *Aegilops*. Lane 1: Marker; Lane 2: *AeVePAP1* from M1; Lane 3: *AeCrPAP1* from M2. M1: *Aegilops ventricosa*; M2: *Aegilops crassa*.

**Sequence analysis of *AeVePAP1* and *AeCrPAP1*:** Sequencing of the cloned gene *AePAP1* revealed that both the CDS of *AeVePAP1* (from *A. ventricosa*) and that of *AeCrPAP1* (from *A. crassa*) were 1.008 kb. The start codons of the two *AePAP1* CDS sequences were ATG, and their stop codons were TGA (Fig. 3). Alignment between these two genes revealed them to be only 3.87% dissimilar overall, each having 42 SNPs in total, including 20 synonymous SNPs and 22 non-synonymous SNPs (Fig. 3; Fig. 5).

**Physicochemical properties and secondary, 3D structure prediction:** Both species' *AePAP1* genes encoded 335 amino acids; the shared identity value of AeVePAP1 and AeCrPAP1 was 94.03%, for which 19 amino acid residues differed from each other. Both AeVePAP1 and AeCrPAP1 had leucine (>9%) as the highest, and cysteine as the lowest, amino acid residue. The AeVePAP1 had a molecular weight of 37.891 kD (Table 2), whose PI value was 6.08 (Table 2), whereas the AeCrPAP1protein had a higher molecular weight (38.201kD) and more acid PI (5.35) (Table 2). The hydrophobicity analysis showed the number of negative peaks exceeding the number of positive peaks (Fig. 4A, 4C), indicating that both AeVePAP1 and AeCrPAP1 were hydrophilic soluble proteins.

According to the secondary structure prediction, the rate of helices, strands, and coils was accounted for 11.5%, 38.5%, and 50% (Table 2), while those of AeCrPAP1 accounted for 10.2%, 38.8% and 51%, respectively (Table 2). The 3D structure prediction of the AeVePAP1 and AeCrPAP1 protein showed that the two proteins each have metal ion-binding center (Fig. 4B, 4D), indicative of metallic proteins.

```
AeVePAP1.seq ATGGCGAGGAGTTCCATGGCCGTGGTGCTCGCCGTCCTGGCCGTGGCGGCGCTCCGGTGCGCGCC   65
AeCrPAP1.seq ----------------------c---------------------------------------   65

AeVePAP1.seq GGCCGCCGCGGAGCTCCCGCTGGTGGAGCACCCGGCCAAGAACGACGGGTCGCTCAGCCTGCTCG  130
AeCrPAP1.seq ------------------------------------------------------------  130

AeVePAP1.seq TCATCGGGGACTGGGGCCGCAACGGCACCTACAACCAATCCAGGGTCGCGGAGCAGATGGGGAAG  195
AeCrPAP1.seq ------ac--------------------------------------a---------------  195

AeVePAP1.seq GTCGGGGAGAAGCTGGACATCGACTTCGTGGTCTCCACCGGCGACAACTTCTACGAGAACGGCCT  260
AeCrPAP1.seq -at-a-tg--g-------------------------------------------g-------  260

AeVePAP1.seq CACGGGCGTCCATGACCAGCAGTTCGAGGAATCCTTCACCAATATATACACCGCCAAGAGCTTGC  325
AeCrPAP1.seq ----------------------------------------c---------c---------  325

AeVePAP1.seq AGAAGCCGTGGTACCTCGTTCTCGGAAACCATGACTACAGGGGTGATGCGCTCGCACAGCTTGAT  390
AeCrPAP1.seq ----------t--------t---------------c----c--------t-----------c  390

AeVePAP1.seq CCGGTCATGCGCAAGCTCGATGAGCGATTCGTTTGCATGAGATCATTCGTTGTCAATGCAGAGAT  455
AeCrPAP1.seq ------------------------------------------a-----------------  455

AeVePAP1.seq CGTGGAGTTCTTCTTCGTCGACACCACTCCATTCCAACTCAAGTACTGGACTCACCCCAAAGACA  520
AeCrPAP1.seq ----------------a-------t-----------g--------------------t-------  520

AeVePAP1.seq GTCACTACGACTGGAGAGGAGTGGCGCCTCGAAAAAACTACATAGCTAATCTGCTGAAGGATTTG  585
AeCrPAP1.seq -------t-------------------------g-------------------------------  585

AeVePAP1.seq GATGAGGCAATGAAGAAATCAACTGCAAAGTGGAAGATTGCTATTGGGCATCATACCATCAGGAG  650
AeCrPAP1.seq ----------------------------------------c---------------------  650

AeVePAP1.seq CGTCAGTGATCATGGGGACACCGAGGAGCTCCTGCAATTGCTTCTTCCAGTCCTCAAGGTTAACG  715
AeCrPAP1.seq ----------------------a-------------------------------------  715

AeVePAP1.seq GCATCGACTTCTACATCAATGGGCACGACCACTGCCTGGAACACATTAGCAGCAGAGACAGCCCA  780
AeCrPAP1.seq ------------------------------------------------------------t---  780

AeVePAP1.seq ATCCAATACTTCACTAGCGGAGGCGGTTCAAAAGCATGGAGGGGAGTCTACCAGCCAAATGATGA  845
AeCrPAP1.seq ------------------------------------------------------------  845

AeVePAP1.seq TAAGCTCCAGTTCTTTTATGATGGGCAAGGGTTCATGTCCCTCCAGCTAAACCAGGACCAAGCTG  910
AeCrPAP1.seq ------------------------------------------------------------  910

AeVePAP1.seq ACTTCATCTTTTATGATGTTTCTGGGAAAGTCTTGTACGAATGGAGCTCGCGCAAAATAAACTAC  975
AeCrPAP1.seq ----------------------------------------g----c-----a-----c----c--  975

AeVePAP1.seq CTTCCACCCTCCATCCATGTCACTGCAGAATGA                              1008
AeCrPAP1.seq t-c-ag---------t----------------                              1008
```

Fig. 3. CDS sequence alignment of two *Aegilops AePAP1* genes.
The two *AePAP1* genes were cloned from seedlings roots of *Aegilops ventricosa* and *Aegilops crassa*, which were treated with phosphorus deficiency solution (P-) for 30 days. The sequence alignments were performed using DNAMAN software with default parameters.

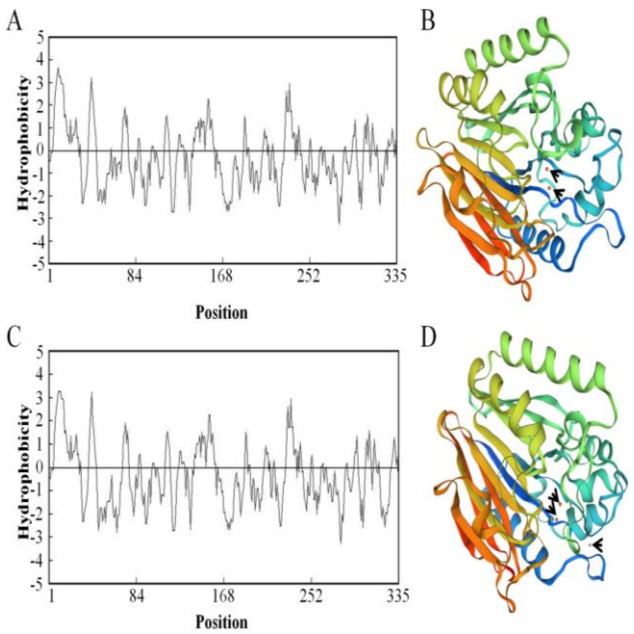


Fig. 4. Physicochemical properties analysis and structure prediction of AePAP1 proteins.
(A) The hydrophobicity analysis of AeVePAP1 protein was using DNAMAN software with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the hydrophobicity value of AA.
(B) The 3D structure prediction of AeVePAP1 protein was carried out using SWISS-MODEL tool. The metal ions were labeled by black arrows.
(C) The hydrophobicity analysis of AeCrPAP1 proteins was using DNAMAN software with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the hydrophobicity value of AA.
(D) The 3D structure prediction of AeCrPAP1 protein was carried out using SWISS-MODEL tool. The metal ions were labeled by black arrows.

**Multiple sequence alignment and motif analysis:** To study the conserved regions and structural characteristics of the AeVePAP1 and AeCrPAP1 proteins, their conserved motifs and domains were investigated. Alignment of the above PAP1 protein sequences and TaPAP1(accession number: BAI78301) revealed they all possessed the characteristic seven metal-binding residues (D, D, Y, N, H, H, H) pf PAP metalloesterase and contained its five conserved motifs [DxG/GDx2Y/GNH (E,D)/Vx2H/GHxH] (Fig. 5). This suggested both AeVePAP1 and AeCrPAP1 belong to the PAPs family in *Aegilops*.

**Protein transmembrane and signal peptide analysis:** To predict the AeVePAP1 and AeCrPAP1 protein functioning, their corresponding transmembrane region and signal peptide sequence were analyzed. These results showed the AeVePAP1 has a transmembrane region (Fig. 6A) and a signal peptide of between 1 and 30 amino acids (Fig. 7A); this indicated the AeVePAP1 protein was not only a membrane protein but also a secreted protein; the AeCrPAP1 likewise had a signal peptide sequence between 1 and 30 amino acids (Fig. 7B), but no transmembrane region (Fig. 6B), indicating the AeCrPAP1 protein was a secreted protein only. Consistent with the above, the subcellular localization prediction confirmed that AeVePAP1 and AeCrPAP1 mainly functioned in the extracellular region (Table 3).

**Phylogenetic relationship analysis:** Purple acid phosphatases (PAPs) belong to a family of binuclear metalloenzymes that exhibit diverse biological functions in all known plant species. We investigated the evolutionary distance and relationship of AeVePAP1 and AeCrPAP1vis-à-vis its orthologs' genes. All PAPs were divided into two clades, wherein AeVePAP1 and AeCrPAP1 were clustered with other small (about 35 kD) mammal-like PAPs (i.e. CAC09923). This phylogenetic tree showed AeVePAP1 closely related to PAP1 from the proximate species *T. aestivum*, yet AeCrPAP1 was more closely related to TaPAP1 than AeVePAP1 (Fig. 8). This revealed that AeVePAP1 and AeCrPAP1had the same most recent common ancestor for the PAP1 gene as *T. aestivum*.

**Expression analysis of gene*AeVePAP1* and *AeCrPAP1*:** To evaluate whether the *AeVePAP1* and *AeCrPAP1*were induced by phosphate (P) starvation, a qRT-PCR was performed on *Aegilops* leaves and roots subjected to phosphate (P) starvation (6 h). This detected no changed transcript levels of*AeVePAP1* and *AeCrPAP1* between P-starvation treatment (P–) and control (P+) (Fig. 9), indicating the *AeVePAP1* and *AeCrPAP1* were constitutively expressed genes; hence, we speculated starvation mainly induced the AeVePAP1 and AeCrPAP1 proteins to move into the extracellular region. We also found that the expression levels of *AeVePAP1* and *AeCrPAP1*were higher in roots than leaves, prompting us to posit the functioning of *AeVePAP1* and *AeCrPAP1* was mainly concentrated in roots.

**Table 2. The data of physicochemical properties, protein secondary structure of *AeVePAP1* and *AeCrPAP1*.**

| | Number of amino acids | Molecular weight | Theoretical pI | Helices number | Strands number | Coils number |
|---|---|---|---|---|---|---|
| *AeVePAP1* | 335 | 37891 | 6.08 | 6 | 20 | 26 |
| *AeCrPAP1* | 335 | 38201 | 5.35 | 5 | 19 | 25 |

**Table 3. The data of subcellular location prediction of AeVePAP1 and AeCrPAP1.**

| | Extracellular, including cell wall | Golgi | Vacuolar | Endoplasmic reticulum | K-NN prediction value |
|---|---|---|---|---|---|
| *AeVePAP1* | 66.7% | 11.1% | 11.1% | 11.1% | K=9/23 |
| *AeCrPAP1* | 66.7% | 11.1% | 11.1% | 11.1% | K=9/23 |

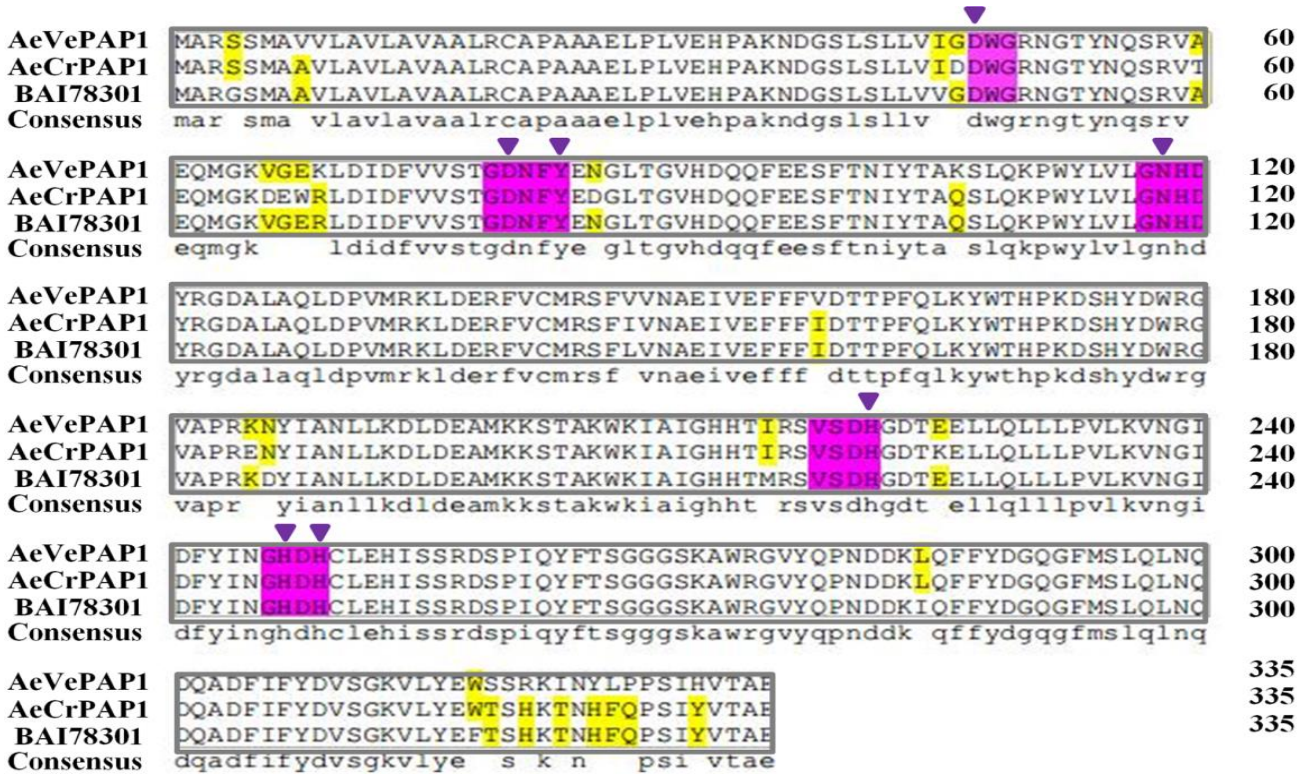


Fig. 5. Sequence alignment of the AePAP1s by ClustalW.
PAP1 protein sequences from *Aegilops ventricosa*,*Aegilops crassa* and *Triticum aestivum* (BAI78301) are shown. The five consensus motifs [DxG/GDx2Y/GNH (E,D)/Vx2H/GHxH] of PAP1 family were identified (Strater *et al*., 1995) and shown boxed in and are shown boxed in pink. The characteristic PAP metalloesterase seven metal-binding residues (D,D, Y, N, H, H, H) were labeled by purple triangles and the different AA sites among the three PAP1 sequences were labeled by yellow.

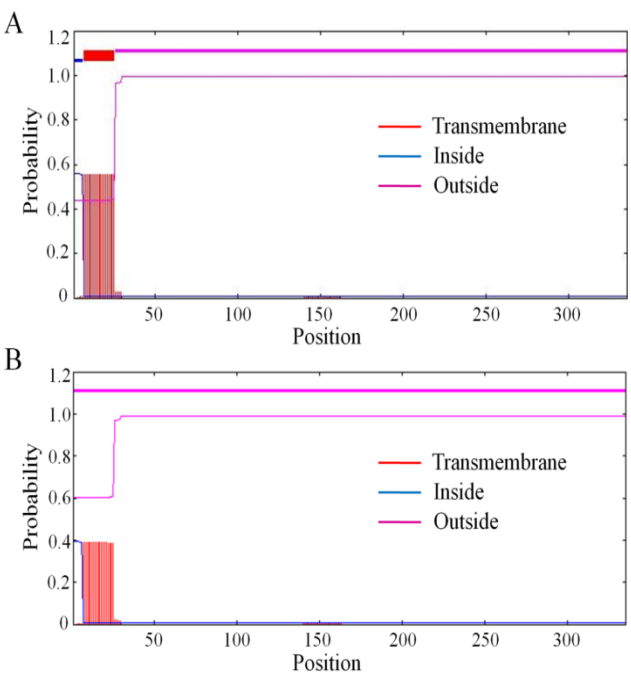


Fig. 6. The transmembrane region analysis of AePAP1s.
(A) The transmembrane region analysis of AeVePAP1 protein was using the TMHMM 2.0 server with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the probability value.
(B) The transmembrane region analysis of AeCrPAP1 protein was using the TMHMM 2.0 server with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the probability value.

## Discussion

*Aegilops tauschii*, a proximate species of wheat, is apt at adjusting to stressful conditions in nature and therefore is inherently tolerant to a variety of abiotic stresses (Liu *et al.*, 2015). Therefore, it is an excellent genetic resource in which to identify candidate genes involved in plants' P starvation response. Yet, surprisingly, the study in P-starvation tolerant evaluations in *Aegilops* remains limited. In this study, there spective tolerance of two *Aegilops* species to P-starvation was evaluated, and comprehensive phenotypic observations proved that this tolerance was stronger of *A. crassa* than *A. ventricosa*. Then, the *PAP1* gene, which encodes a key regulator of the P-starvation response, was cloned from both *Aegilops* genotypes and their sequence characteristics and evolutionary differences were analyzed. Our results lay a timely foundation for identification of P-starvation tolerant germplasm and further studies of *PAP1* functions that could be practically applied for improving the P-utilization efficiency of crops.

We found the *AeVePAP1* and *AeCrPAP1*all encoded LMW PAPs not unlike the mammal PAPs and AtPAPs in the Arabidopsis subgroup Ⅲb (Tran *et al*., 2010), results consistent with those of the phylogenetic tree analysis. Many studies have found that PAPs responded to conditions of P starvation, but most PAPs induced by the stress belonged to HMW PAPs (Pozo *et al*., 2010; Tran *et al*., 2010; Zimmermann *et al*., 2004), leaving few LMW PAPs responsive to P starvation at the transcriptional level in plants (Pozo *et al*., 2010). Like most LMW PAPs, the AeVePAP1

and AeCrPAP1proteins studied here were not induced by P starvation, but they all contained the predicted N-terminal secretion signal and therefore could participate in extracellular P scavenging, P mobilization from the rhizosphere, or plant cell wall regeneration. Research in the last 20 years has provided detailed knowledge about plant P-starvation response genes and their mechanisms. It appears that PAPs are crucially involved in the adjustment of P-starvation tolerance trait of plants (Li *et al*., 2002; Tran *et al*., 2010). Moreover, overexpression of *PAPs* in *Arabidopsis* and soybean augmented these plants' P utilization efficiency (Wang *et al*., 2009; Xiao *et al*., 2006; Zhang *et al*., 2015). In view of the above reasoning, the CDS of *AePAP1s* stands as a potentially valuable candidate gene for the future development of P-starvation tolerant transgenic crop plants.

To locate the key evolution sites of *PAP1*, we compared the CDS of *AePAP1s* in the two different *Aegilops* genotypes, it is found that these are 22 non-synonymous SNPs between *AeVePAP1* and *AeCrPAP1*. Though not located in any catalytic domains, these non-synonymous SNPs are possibly related to subfunctions of *PAP1s* and these may contribute to the pronounced P-starvation tolerance difference between *Aegilops* species. Moreover, the difference concerning the protein transmembrane region between AeVePAP1 and AeCrPAP1 also supports our above interpretation that posits the subfunctioning of *PAP1s*. A recent phylogenetic study found that both AeVePAP1 and AeCrPAP1 were closely related to PAP1 from wheat proximate species, indicating that PAP1s in wheat proximate species are highly homologous; that finding is consistent with *A. tauschii* being the progenitor of the wheat D genome (Luo *et al*., 2017). However, the AeCrPAP1was closer to the PAP1 of wheat than was AeVePAP1, indicating that the *AeCrPAP1* from *A. crassa* could be preferentially applied to improve P-starvation tolerance in wheat.

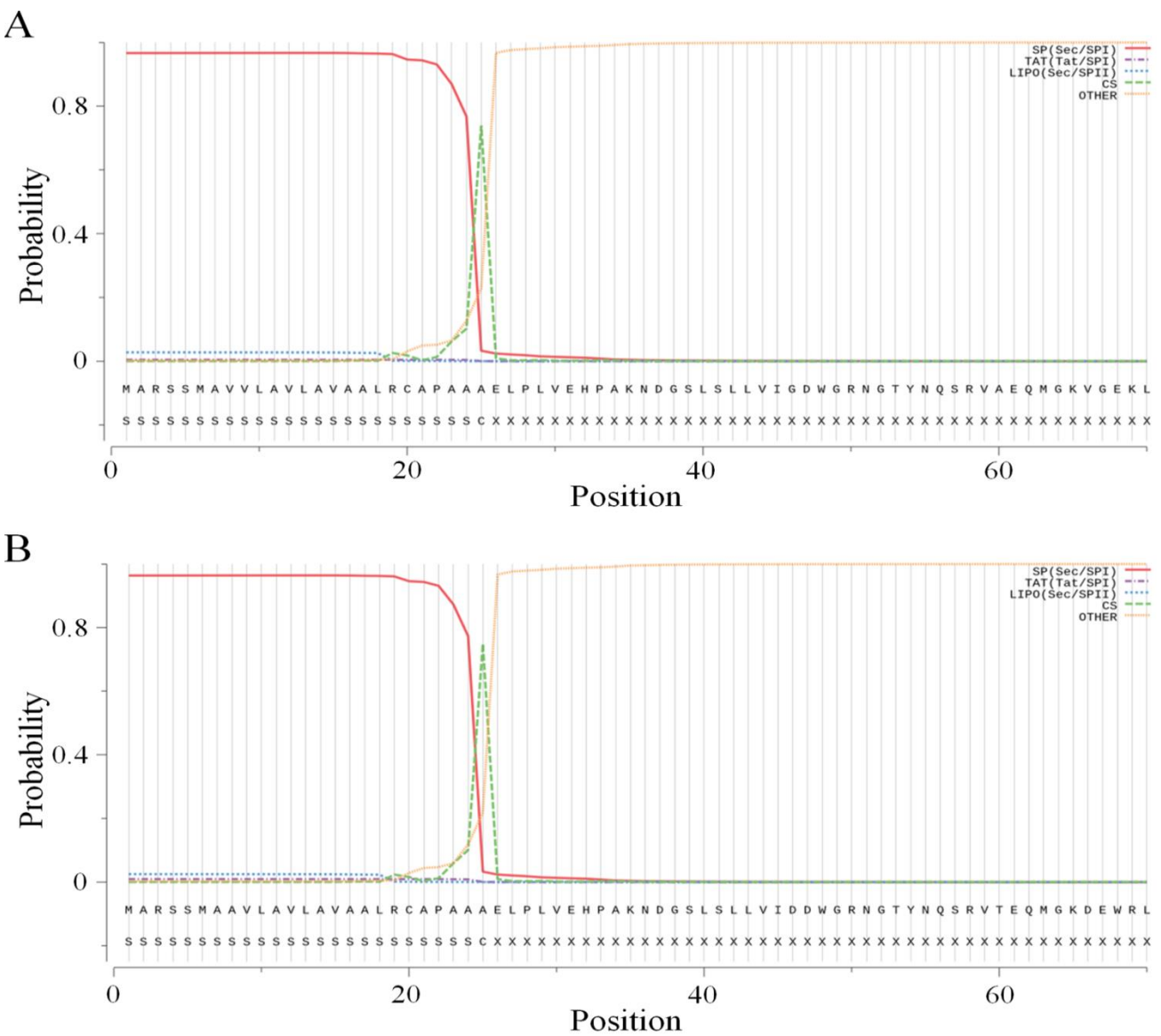


Fig. 7. The signal peptide analysis of AePAP1s.
(A) The signal peptide analysis of AeVePAP1 protein was estimated by using the SignalP4.1 server with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the probability value.
(B) The signal peptide analysis of AeCrPAP1 protein was estimated by using the SignalP4.1 server with default parameters. The x-coordinate represents the amino acid (AA) position of protein sequence and the y-coordinate represents the probability value.

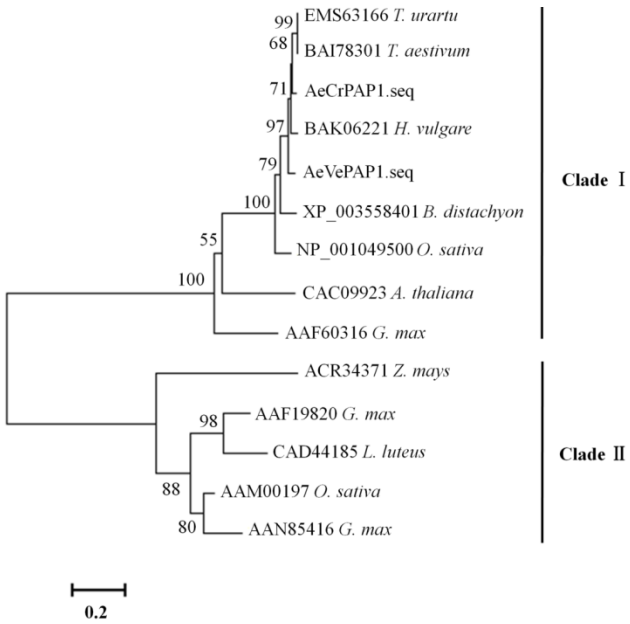


Fig. 8. Phylogenetic analysis of *AePAP1s* with *PAP* from other plants.
Phylogenetic tree of PAPs from *Arabidopsis thaliana*, *Glycine max*, *Lupinus luteus*, *Oryza sativa*, *Zea mays*, *Brachypodium distachyon*, *Hordeum vulgare*, *Triticum urartu* and *Triticum aestivum*. The neighbor-joining tree was constructed by MEGA 6.0 software and the all PAPs were divided two clade and named clade Ⅰ and clade Ⅱ.

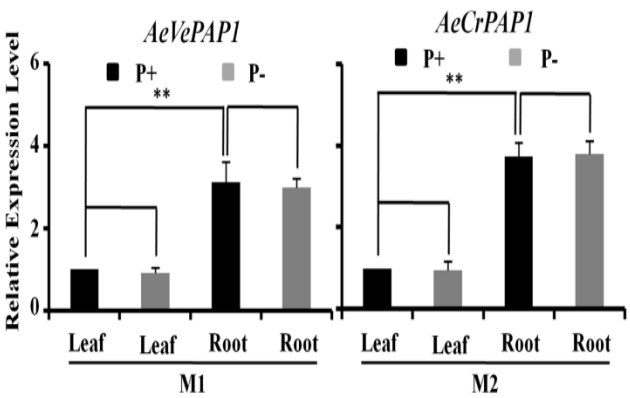


Fig. 9. Expression analysis of *AeVePAP1* and *AeCrPAP1* under phosphorus starvation treatment.
Quantitative RT-PCR experiments measured the relative expression levels of *PAP1* in leaf or root tissue of 10-day-old *Aegilops* after treatment with phosphorus deficiency (P-) solution treatment for 6 h and normal (P+, contain 2 mM Pi) solution as control. All results are the mean ± s.d. The asterisks indicate a statistically significant difference (*t*-test, $* p<0.05$, $** p<0.01$). Individual reactions were performed in three biological replicates and three technical replicates and the experiments were repeated at least twice. M1: *Aegilops ventricosa*; M2: *Aegilops crassa*.

## Conclusion

*A. crassa* was more tolerant than *A. ventricosa* against imposed P-starvation stress conditions. To clarify the involved P-starvation tolerance mechanism, the key P-metabolism catalyzing enzyme genes *AePAP1s* from each species were analyzed. The qRT-PCR experiment proved that *AeVePAP1* and *AeCrPAP1*were not induced by P-starvation, similar to most LMW PAPs, so we focused on analyzing their differences in sequence, structure, and evolution. Our bioinformatics analysis found 42 SNPs, including 22 non-synonymous SNPs, in the CDS, but the transmembrane region and the evolutionary relationship also differed significantly between *AeVePAP1* and *AeCrPAP1*. Our study provides fundamental data for elucidating the P-starvation tolerance mechanism of *Aegilops*, and it provides germplasm resources and candidate genes for improving the P-starvation tolerance of crops.

## Acknowledgement

We acknowledge the financial support of Fundamental Scientific Research Program of Northwest Agriculture and Forestry University (QN2011003) and Cyrus Tang Breeding Foundation.

## References

Dionisio, G., C.K. Madsen, P.B. Holm, K.G. Welinder, M. Jorgensen, E. Stoger, E. Arcalis and H. Brinchpedersen. 2011. Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice. *J. Plant Physiol*., 156: 1087-1100.

Duff, S.M.G., G. Sarath and W.C. Plaxton. 2010. The role of acid phosphatases in plant phosphorus metabolism. *Physiol. Plant*., 90: 791-800.

Fang, Z., C. Shao, Y. Meng, P. Wu and M. Chen. 2009. Phosphate signaling in Arabidopsis and *Oryza sativa*. *J. Plant Sci*., 176: 170-180.

Greiner, R., K.D. Jany and M.L. Alminger. 2000. Identification and properties of myo-inositol hexakisphosphate phosphohydrolases (phytases) from barley (*Hordeum vulgare*). *J. Cereal Sci*., 31: 127-139.

Hayakawa, T., Y. Toma and I. Igaue. 1989. Purification and characterization of acid phosphatases with or without phytase activity from rice bran. *J. Agric. Biol. Chem*., 53: 1475-1483.

Hegeman, C.E. and E.A. Grabau. 2001. A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *J. Plant Physiol*., 126: 1598-1608.

Hou, E.Q., C.R. Chen, Y.Q. Luo, G.Y. Zhou, Y.W. Kuang, Y.G. Zhang, M. Heenan, X.K. Lu and D.Z. Wen. 2018. Effects of climate on soil phosphorus cycle and availability in natural terrestrial ecosystems. *Global Change Biol*., 8: 3344-3356.

Klabunde, T., N. Sträter, B. Krebs and H. Witzel. 1995. Structural relationship between the mammalian Fe (III)-Fe(II) and the Fe(III)-Zn(II) plant purple acid phosphatases. *Febs Lett*., 367: 56-60.

Kuang, R., K.H. Chan, E. Yeung and B.L. Lim. 2009. Molecular and biochemical characterization of *AtPAP15*, a purple acid phosphatase with phytase activity, in *Arabidopsis*. *J. Plant Physiol*., 151: 199-209.

Lebansky, B.R., T.D. Mcknight and L.R. Griffing. 1992. Purification and characterization of a secreted purple phosphatase from soybean suspension cultures. *J. Plant Physiol*., 99: 391-395.

Li, D.P., H.F. Zhu, K.F. Liu, X. Liu, G. Leggewie, M.K. Udvardi and D.W. Wang. 2002. Purple acid phosphatases of *Arabidopsis thaliana*. Comparative analysis and differential regulation by phosphate deprivation. *J. Biol. Chem*., 277: 27772-27781.

Liu, Y.X., L. Wang, M. Deng, Z.Y. Li, Y.L. Lu, J.R. Wang, Y.M. Wei and Y.L. Zheng. 2015. Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. *Theor. Appl. Genet*., 128: 2203-2212.

Luo, M., Y.Q. Gu, D. Puiu, H. Wang, S. Twardziok, K.R. Deal and J. Dvořák. 2017. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, 551: 498-502.

Nakano, T., T. Joh, E. Tokumoto and T. Hayakawa. 1999. Purification and characterization of phytase from bran of *Triticum aestivum* L. cv. Nourin #61. *Food Sc. Technol. Res*., 5: 18-23.

Nicolas, L.N. 2001. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics*, 17: 1226-1227.

Olczak, M. and W. Wtorek. 2010. Two subfamilies of plant purple acid phosphatases. *Physiol. Plantarum*., 118: 491-498.

Olczak, M., B. Morawiecka and W. Watorek. 2003. Plant purple acid phosphatases - genes, structures and biological function. *Acta Biochim. Pol*., 50: 1245-1256.

Pozo, J.C.D., I. Allona, V. Rubio, A. Leyva and J. Paz-Ares. 2010. A type 5 acid phosphatase gene from *Arabidopsis thaliana* is induced by phosphate starvation and by some other types of phosphate mobilising/oxidative stress conditions. *Plant J*., 19: 579-589.

Raghothama, K.G. 1999. Phosphate acquisition. *Ann. Rev. Plant Biol*., 50: 665-693.

Ravichandran, S., S. Stone, B. Benkel and B. Prithiviraj. 2013. Purple acid phosphatase 5 is required for maintaining basal resistance against pseudomonas syringae in Arabidopsis. *BMC Plant Biol*., 13: 107-107.

Rehman, S.U. 2019. Major nutrient fluxes and water use efficiency of winter vegetables under peri-urban farming of north-western Punjab, Pakistan. *Pak. J. Bot.*, 51: 1753-1759.

Richardson, A.E. 2009. Regulating the phosphorus nutrition of plants: molecular biology meeting agronomic needs. *Plant Soil*., 322: 17-24.

Robinson, W.D., I. Carson, S. Ying, K. Ellis and W.C. Plaxton. 2012. Eliminating the purple acid phosphatase AtPAP26 in *Arabidopsis thaliana* delays leaf senescence and impairs phosphorus remobilization. *New Phytol*., 196: 1024-1029.

Schenk, G., T.W. Elliott, E. Leung, L.E. Carrington, N. Mitić, L.R. Gahan and L.W. Guddat. 2008. Crystal structures of a purple acid phosphatase, representing different steps of this enzyme's catalytic cycle. *BMC Struct. Biol*., 8: 6-6.

Strater, N., T. Klabunde, P. Tucker, H. Witzel and B. Krebs. 1995. Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site. *Science*, 268: 1489-1492.

Tran, H.T., B.A. Hurley and W.C. Plaxton. 2010. Feeding hungry plants: the role of purple acid phosphatases in phosphate nutrition. *Plant Sci*., 179: 14-27.

Vance, C.P., C. Uhdestone and D.L. Allan. 2003. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytol*., 157: 423-447.

Wang, X.R., Y.X. Wang, J. Tian, B.L. Lim and X.L. Yan. 2009. Overexpressing *AtPAP15*enhances phosphorus efficiency in soybean. *Plant Physiol*., 151: 233-240.

Wu, F., G.C. Lin, J.W. Wu, J. Yao, F. Zhou and D.G. Jiang. 2013. Effects of planting *AtPAP15* transgenic soybeans on soil nutrients and enzyme activities in rhizosphere. *Acta Pedologica Sinica*., 3: 600-608.

Xiao, K., J.T. Gu, M. Harrison and Z.Y. Wang. 2006. Expression characteristics of *MtPAP1* and its exotic expression in Arabidopsis affecting organic phosphorus absorption of plants. *J. Plant Physiol. Mol. Biol*., 32: 99-106.

Xiao, K., M.J. Harrison and Z.Y. Wang. 2005. Transgenic expression of a novel *M. truncatula* phytase gene results in improved acquisition of organic phosphorus by Arabidopsis. *Planta*., 222: 27-36.

Zhang, Y., X. Wang and D. Liu. 2015. Arabidopsis phosphatase under-producer mutants *pup1* and *pup3* contain mutations in the *AtPAP10* and *AtPAP26* genes. *Plant Signal. Behav*., 10: 1559-2324.

Zhu, H.F., W.Q. Qian, X.Z. Lu, D.P. Li, X. Liu, K.F. Liu and D.W. Wang. 2005. Expression patterns of purple acid phosphatase genes in Arabidopsis organs and functional analysis of *AtPAP23* predominantly transcribed in flower. *Plant Mol. Biol*., 59: 581-594.

Zimmermann, P., B. Regierer, J. Kossmann, E. Frossard, N. Amrhein and M. Bucher. 2004. Differential expression of three purple acid phosphatases from potato. *Plant Biol*., 6: 519-528.

(Received for publication 30 November 2020)