Theor Appl Genet (2008) 117:963–975
DOI 10.1007/s00122-008-0836-x

ORIGINAL PAPER

# Cloning and characterization of phosphorus starvation inducible *Brassica napus PURPLE ACID PHOSPHATASE 12* gene family, and imprinting of a recently evolved MITE-minisatellite twin structure

**Kun Lu · You-Rong Chai · Kai Zhang · Rui Wang · Li Chen · Bo Lei · Jun Lu · Xin-Fu Xu · Jia-Na Li**

Received: 23 January 2008 / Accepted: 24 June 2008 / Published online: 21 August 2008


**Abstract** Purple acid phosphatase (PAP) is important for phosphorus assimilation and in planta redistribution. In this study, seven *Brassica napus PAP12* (*BnPAP12*) genes orthologous to *Arabidopsis thaliana PAP12* (*AtPAP12*) are isolated and characterized. NCBI BLASTs, multi-alignments, conserved domain prediction, and featured motif/residue characterization indicate that all *BnPAP12* members encode dimeric high molecular weight plant PAPs. *BnPAP12-1*, *BnPAP12-2*, *BnPAP12-3* and *BnPAP12-7* (Group I) have six introns and encode 469-aa polypeptides structurally comparable to AtPAP12. *BnPAP12-4* and *BnPAP12-6* (Group II) have seven introns and encode 526-aa PAP12s. Encoding a 475-aa polypeptide, *BnPAP12-5* (Group III) is evolved from a chimera of 5′ part of Group I and 3′ part of Group II. Sequence characterization and Southern detection suggest that there are about five *BnPAP12* alleles. Homoeologous non-allelic fragment exchanges exist among *BnPAP12* genes. *BnPAP12-4* and *BnPAP12-6* are imprinted with a *Tourist*-like miniature inverted-repeat transposable element (MITE) which is tightly associated with a novel minisatellite composed of four 36-bp tandem repeats. Existing solely in *B. rapa/oleracea* lineage, this recently evolved MITE-minisatellite twin structure does not impair transcription and coding capacity of the imprinted genes, and could be used to identify close relatives of *B. rapa/oleracea* lineage within *Brassica*. It is also useful for studying MITE activities especially possible involvement in minisatellite formation and gene structure evolution. *BnPAP12-6* is silent in transcription. All other *BnPAP12* genes basically imitate *AtPAP12* in tissue specificity and Pi-starvation induced expression pattern, but divergence and complementation are distinct among them. Alternative polyadenylation and intron retention also exist in *BnPAP12* mRNAs.

Kun Lu, You-Rong Chai and Jia-Na Li have equal contribution to the study.

Database Accession numbers: GenBank accession nos.: EU014612 (*BnPAP12-1*), EU014613 (*BnPAP12-1* mRNA), EU014614 (*BnPAP12-2*), EU014615 (*BnPAP12-2* mRNA), EU014616 (*BnPAP12-3*), EU014617 (*BnPAP12-3* mRNA), EU014618 (*BnPAP12-4*), EU014619 (*BnPAP12-4* mRNA), EU014620 (*BnPAP12-4PM* mRNA), EU014621 (*BnPAP12-5*), EU014622 (*BnPAP12-5* mRNA), EU014623 (*BnPAP12-5PM1* mRNA), EU014624 (*BnPAP12-5PM2* mRNA), EU014625 (*BnPAP12-6*), EU014626 (*BnPAP12-6* putative mRNA), EU014627 (*BnPAP12-7*), EU014628 (*BnPAP12-7* mRNA) and EU014629 (*BnPAP12-7PM* mRNA).

Communicated by C. F. Quiros.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-008-0836-x) contains supplementary material, which is available to authorized users.

K. Lu · Y.-R. Chai (✉) · K. Zhang · R. Wang · L. Chen · B. Lei · J. Lu · X.-F. Xu · J.-N. Li (✉)
Chongqing Rapeseed Engineering Research Center, Southwest University, Tiansheng Road 216#, Beibei, Chongqing 400716, P. R. China
e-mail: chaiyourong1@163.com

J.-N. Li
e-mail: ljn68251950@gmail.com

K. Lu · Y.-R. Chai · K. Zhang · R. Wang · L. Chen · B. Lei · J. Lu · X.-F. Xu · J.-N. Li
Chongqing Key Laboratory of Crop Quality Improvement, Southwest University, Tiansheng Road 216#, Beibei, Chongqing 400716, P. R. China

**Abbreviations**

MITE Miniature inverted-repeat transposable element
MMTS MITE-minisatellite twin structure
PAP Purple acid phosphatase

## Introduction

Acid phosphatases (APases; E.C. 3.1.3.2) catalyze the hydrolysis of activated phosphoric acid esters and anhydrides, and display optimal activity at pH 4–7. Under phosphate-deficient conditions, the secretion of APases increases significantly, which forms a "salvage system" and functions in the production, transport and recycling of phosphate (Pi) (Duff et al. 1994). Among the 44 APases of *Arabidopsis thaliana*, 29 are purple acid phosphatases (PAPs) (Li et al. 2002).

The PAPs comprise of a group of binuclear metal-containing acid hydrolases, and have been found in animals, plants and microbes (Schenk et al. 2005). All plant PAPs contain a binuclear center, Fe(III)–Zn(II) or Fe(III)–Mn(II), in their active sites, and their characteristic of pink to purple colors at about 560 nm results from a charge transfer transition by a tyrosine residue coordinating a ferric ion (Klabunde et al. 1995; Strater et al. 1995).

So far, plant PAPs have been isolated from diverse species, including *A. thaliana*, alfalfa, lupin, potato, rice, soybean, big duckweed, sweet potato, tobacco and tomato (del Pozo et al. 1999; Durmus et al. 1999a, b; Li et al. 2002; Liao et al. 2003; Zimmermann et al. 2004; Xiao et al. 2006). Previous studies have proved that PAPs are multifunctional proteins, which may play important roles in plant growth, development and reproduction, especially under Pi-starvation conditions (Zhu et al. 2005).

In *A. thaliana*, AtPAP17 (AtACP5) and AtPAP12 are the only two members that are well characterized. AtPAP17 has been purified as a 34-kDa monomer and the N-terminal sequence displayed significant similarity to mammal type-5 APases (del Pozo et al. 1999; Li et al. 2002). Its transcription is induced by both Pi-starvation and abscisic acid/salt stress in both root and shoot, but not by paraquat or salicylic acid. It is also induced by oxidative stress and highly expressed in senescent leaf (del Pozo et al. 1999). The transcription of *AtPAP12* (NM_128277) is highest in flowers, intermediate in stems, roots and siliques, and lowest in leaves under Pi-sufficient conditions, and is moderately induced by low-P conditions (Li et al. 2002; Zhu et al. 2005). AtPAP12 has a signal peptide for protein secretion from root to the surrounding medium, and its transcription is induced first in shoots, then in lateral root meristems, and eventually in whole roots in response to Pi deficiency. Unlike *AtPAP17*, *AtPAP12* is induced specifically by low P levels, while salicylic acid, jasmonic acid and other inducers do not activate its promoter (Haran et al. 2000).

Oilseed rape (*Brassica napus*) is the second largest oilseed crop in the world. However, about 8% of the oilseed rape area is Pi-deficient in UK (Skinner and Todd 1998). Similarly, the major oilseed rape production areas in China, especially the Yangtze River belt, are severely affected by insufficient soil phosphate. Since oilseed rape is sensitive to Pi deficiency, it is important to increase P absorption and metabolism efficiency for oilseed rape production. But, little is known about the molecular mechanism of P acquisition in this crop. In this study, the full-length cDNAs and genomic sequences of seven *BnPAP12* genes have been isolated and molecularly characterized. Some tentative *cis*-regulations have been speculated and a recently evolved *Tourist*-like miniature inverted-repeat transposable element (MITE) has been discovered in the *B. rapa*/*oleracea* lineage. The paralogous *BnPAP12* members also show diverged expression patterns on the basis of organ specificity and Pi-starvation induction. These results will help reveal molecular mechanism of P metabolism of *B. napus*, and provide the opportunity to perform transgenic manipulation or molecular marker assisted selection for high P-efficiency genotypes of oilseed rape.

## Materials and methods

### Plant materials, treatment, and nucleic acid extraction

Seeds of 26 Brassicaceae species used in this study were kindly provided by Tohoku University *Brassica* Seed Bank, Japan (http://www.agri.tohoku.ac.jp/pbreed/Seed_Stock_DB/SeedStock-top.html) and Chongqing Rapeseed Technology Research Center, China (Table 1). Inbred line W17 representing *B. napus* was grown under field conditions, while materials of other 25 species were cultivated in light incubators. Roots (Ro), hypocotyls (Hy), cotyledons (Co), stems (St), leaves (Le), buds (Bu), flowers (Fl), silique pericarps (SP), and seeds of 10 (10D), 20 (20D) and 30 days (30D) after flowering were sampled from the line W17, and for the other 25 species, only young leaves were sampled. In order to investigate induced expression patterns under Pi-starvation, seedlings of the line W17 were pot-cultivated

K. Lu · Y.-R. Chai · K. Zhang · R. Wang · L. Chen · B. Lei · J. Lu · X.-F. Xu · J.-N. Li
Key Laboratory of Biotechnology and Crop Quality Improvement of Ministry of Agriculture, Southwest University, Tiansheng Road 216#, Beibei, Chongqing 400716, P. R. China

K. Lu · Y.-R. Chai · K. Zhang · R. Wang · L. Chen · B. Lei · J. Lu · X.-F. Xu · J.-N. Li
College of Agronomy and Biotechnology, Southwest University, Tiansheng Road 216#, Beibei, Chongqing 400716, P. R. China

**Table 1** Plant materials used in this study

| Species | Code in Tohoku University *Brassica* Seed Bank | Chromosome numbers (n) | Species | Code in Tohoku University *Brassica* Seed Bank | Chromosome numbers (*n*) |
|---|---|---|---|---|---|
| *Arabidopsis thaliana* | | 5 | *Brassica gravinae* | Gr-1 | 10 |
| *Brassica carinata* | Ca-115 | 17 | *Brassica macrocarpa* | O-503 | 9 |
| *Brassica juncea* | J-114 | 18 | *Brassica tournefortii* | T-165 | 10 |
| *Brassica napus* | | 19 | *Brassica villosa* | Vill-1 | 9 |
| *Brassica nigra* | Ni-111 | 8 | *Diplotaxis muralis* | DIP-MUR-1 | 21 |
| *Brassica oleracea* | | 9 | *Diplotaxis siifolia* | DIP-SII-4 | 10 |
| *Brassica rapa* | C-146 | 10 | *Eruca sativa* | | 11 |
| *Brassica amplexicaulis* | Am-4 | 11 | *Erucastrum canariense* | EST-CAN-1 | 9 |
| *Brassica barrelieri* | Ba-103 | 10 | *Erucastrum gallicum* | EST-GAL-1 | 15 |
| *Brassica cretica* | Cr-7 | 9 | *Erucastrum laevigatum* | EST-LAE-2 | 7 |
| *Brassica deflexa* | Df-4 | 7 | *Moricandia arvensis* | MOR-ARV-1 | 14 |
| *Brassica erucastrum* | Ec-102 | | *Raphanus sativus* | RAP-SAT-14 | 9 |
| *Brassica fruticulosa* | Fr-103 | 8 | *Sinapis alba* | | 12 |

in full-strength Hoagland solution with a thermo-photoperiod of 25°C for 16 h/18°C for 8 h (light/dark), and 4-week-old seedlings were transferred to a new nutrient solution for Pi-starvation treatment (0.5 μM $KH_2PO_4$) (Hoagland and Arnon 1950). Seedling leaves (SL) and seedling roots (SR) were harvested after 0, 12, 24 h, 2, 4 and 8 days of treatment, respectively, and sampled again after Pi was re-supplied for 4 days. All samples were immediately frozen in liquid nitrogen and stored at −80°C.

Total RNA of each sample was extracted using a CTAB method (Jaakola et al. 2001). RNA aliquots were treated with RNase-free DNase I (TaKaRa) to remove contaminated DNA. Genomic DNA was extracted from leaves of the 26 species using a CTAB method (Saghai-Maroof et al. 1984). Quality and concentration of total RNA and total genomic DNA samples were assessed by agarose gel electrophoresis and spectrophotometry.

## Cloning of full-length cDNAs and genomic sequences of *BnPAP12* genes

Five μg of total RNA from Pi-starvation induced W17 seedling samples was used as the template to synthesize first-strand total cDNA using GeneRacer kit (Invitrogen, USA). Two sense and two antisense polymerase chain reaction (PCR) primers were designed corresponding to specific *AtPAP12* oligonucleotides at conservative sites on the basis of sequence alignment of 29 *AtPAP*s from *Arabidopsis*. Sense primers FPAP12-31 and FPAP12-32 were paired with GeneRacer kit primers 3′P and 3′NP, respectively, for primary and nested amplifications of 3′ RACE (Supplementary Table 1). As for primary and nested amplifications of the 5′ cDNA ends, GeneRacer kit primers 5′P and 5′NP were paired with RPAP12-51 and RPAP12-52, respectively. The 50 μl PCR was set up according to the instruction of the GeneRacer kit. In primary amplification, 1 μl of first-strand total cDNA was used as template, while 0.01 μl of primary PCR product was used in nested amplification. The following PCR program was used for all the four PCRs (MJ PTC-200): 94°C for 2 min, 30 cycles of 94°C for 1 min, 52°C for 1 min and 72°C for 1 min, and succeeded by 72°C for 10 min.

Based on sequencing results of 5′ and 3′ RACE, two 5′-end sense primers (FPAP12-1 and FPAP12-10) and four 3′-end antisense primers (RPAP12-8, RPAP12-9, RPAP12-11 and RPAP12-19) were synthesized and combined into eight primer pairs to amplify full-length cDNAs of *BnPAP12* gene family (Supplementary Table 1). The PCR was performed as the aforementioned procedure to obtain genomic DNA sequences with 0.5 μg total genomic DNA of the line W17.

Target bands were recovered and cloned into the pMD18-T vector (TaKaRa), and multiple PCR-positive colonies especially those with insert-length polymorphism were sequenced using the M13F/M13R primers (Wei et al. 2007). Names of gene members were assigned according to sequenced unique full-length and RACE-end sequences, and member-specific primers were designed based on multi-alignment among them. Primer pairs FPAP12-172S/RPAP12-2MS, FPAP12-3S/RPAP12-3S, FPAP12-43S/RPAP12-4S, FPAP12-5S/RPAP12-56S and FPAP12-64S/RPAP12-56S were used for specific identification of *BnPAP12-2*, *BnPAP12-3*, *BnPAP12-4*, *BnPAP12-5* and *BnPAP12-6*, respectively (Supplementary Table 2). Using gradient PCR with the W17 genomic DNA as template, the highest feasible annealing temperatures of these five primer pairs were determined as 57, 62, 64, 62 and 61°C, respectively. Because of the very high similarity between

*BnPAP12-1* and *BnPAP12-7*, primer pair FPAP12-172S/RPAP12-17S was used to detect the both with an annealing temperature of 58°C. The specificity of the primer pairs was further confirmed by little or no cross-member amplification using sequenced colony plasmids of each member as template. Member-specific identification was adopted to screen colonies in supplementary cloning of some members.

Analyses of open reading frame (ORF) translation, parameter calculation and sequence alignment were performed with Vector NTI Advance 9.0. BLAST analyses, conserved domain (CD) search and protein structure prediction were performed on websites such as NCBI (http://www.ncbi.nlm.nih.gov) and Expasy (http://www.expasy.org). Multiple sequence alignments were performed using Clustal X 1.83, and a phylogenetic tree was constructed by the Neighbor-Joining method with MEGA 3.1 (Thompson et al. 1997; Kumar et al. 2004). The reliability of the tree was measured by bootstrap analysis with 1,000 replicates.

### Detection of distribution of the MITE-minisatellite twin structure among Brassicaceae species

Revealed by multi-alignments of *AtPAP12* and *BnPAP12* genes, *BnPAP12-4* and *BnPAP12-6* were imprinted with two extra inserts (Insert I at $G_{181}$–$A_{358}$ and Insert II at $T_{405}$–$T_{512}$ in both). They formed a MITE-minisatellite twin structure (MMTS), which might be evolved recently within Brassicaceae. To inspect the distribution and origin of the MMTS, conservation-site degenerate-primer pair FMI-SA (5′-GAAGATGAGTTCAAGATCTGACCT-3′)/RMI-SA (5′-CCTTCATGGTT(A/C)CCTTGCGTTA-3′) was designed to amplify the MMTS context region using the total genomic DNA of each of the 26 Brassicaceae species as templates. An annealing temperature of 60°C was used, with other conditions same as in the 3′ and 5′ RACE. MMTS-free amplicants were 219 bp, while MMTS-containing amplicants, 505 bp (219 bp plus a 286-bp MMTS).

### Southern hybridization

Aliquots of total genomic DNA (60 μg per enzyme) of W17 were digested at 37°C for 16–27 h with *Eco*RI, *Eco*RV, *Sac*I, *Xba*I and *Hin*dIII (New England Biolabs, USA), respectively, which did not cut within the probe region of *BnPAP12* members. Fully digested DNA samples were subjected to electrophoresis on 0.8% agarose gel and transferred onto a positively charged nylon membrane (Roche) through capillarity. Using primer pair FPAP12SR/RPAP12SR annealed at 58°C, a 296-bp fragment was amplified by a PCR using *BnPAP12-1* full-length cDNA as template and labeled with Digoxigenin-11-dUTP (Roche) (Supplementary Table 2). Southern hybridization was performed in DIG Easy Hyb (Roche) at 39°C for 16 h, followed by moderate-stringency washing and immunological detection (DIG Wash and Block Buffer Set and DIG Nucleic Acid Detection Kit, Roche).

### RT-PCR detection of the expression patterns of *BnPAP12* genes

One microgram of each total RNA sample was reverse-transcribed (RT) in a 10-μl volume using Oligo dT-Adaptor Primer (RNA PCR Kit (AMV) Ver.3.0, TaKaRa), and 1 μl of the RT product was used as template in a 50-μl PCR. Using primer pair F26S/R26S, a 534-bp region of the *26S rRNA* gene was amplified as internal control (Singh et al. 2004). Annealed at 58°C, primer pair FPAP12SR/RPAP12SR was used to detect the overall expression of the *BnPAP12* gene family. Meanwhile, member-specific primer pairs and corresponding annealing temperatures described above were used to detect the expression patterns of individual *BnPAP12* genes (Supplementary Table 2). *BnPAP12* amplifications were carried out for 31 cycles, and Pi-starvation induced *BnPAP12* overall expression was further detected with a 24-cycle PCR. All RT-PCRs were performed with three replicates. RT-PCR bands of individual *BnPAP12* members in hypocotyl were recovered and T-vector cloned, and three colonies were sequenced for each member.

## Results

### Cloning of full-length cDNAs and genomic sequences of *BnPAP12* gene family

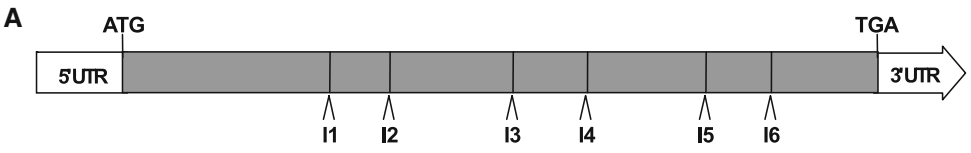
Nested PCR of 3′ RACE produced a band of about 450 bp. After gel recovery, T-vector cloning and *E. coli* transformation, colonies showed polymorphism in insert length. Sequencing of representative colonies yielded 10 different insert sizes of 381, 407, 421, 429, 429, 429, 453, 445, 486 and 559 bp (poly(A) tail was not included), respectively. Alignments showed that they represented four different 3′ cDNA ends, two of which showed alternative polyadenylation sites. Nested PCR product of 5′ RACE showed a band of about 750 bp. Eight representatives of the polymorphic colonies were sequenced and two different 5′ cDNA ends (784 and 624 bp) were obtained. Orthology of these cDNA ends to *AtPAP12* was proved by NCBI BLASTn.

Based on the 5′ and 3′ cDNA ends, eight primer pairs were designed to amplify the full-length *BnPAP12* genes. With total cDNA as template, four specific bands were produced. One of these four bands was about 1.7 kb, two about

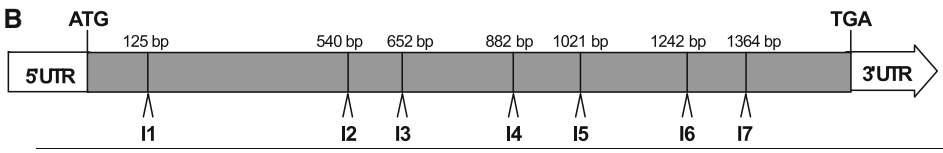
1.8 kb, and one about 1.9 kb, respectively. After sequencing, four unique full-length cDNAs were obtained and named as *BnPAP12-3*, *BnPAP12-4*, *BnPAP12-5* and *BnPAP12-7*. Both *BnPAP12-4* and *BnPAP12-7* had one type of premature mRNA (*BnPAP12-4PM* and *BnPAP12-7PM*) containing unspliced intron(s). *BnPAP12-5* possessed two types of premature mRNAs denoted *BnPAP12-5PM1* and *BnPAP12-5PM2* containing one and three unspliced introns, respectively. Genomic sequence amplifications yielded seven specific bands, five of about 2.4 kb and two of about 2.6 kb. Genomic sequences of *BnPAP12-3*, *BnPAP12-4*, *BnPAP12-5*, *BnPAP12-7* and three novel members (*BnPAP12-1*, *BnPAP12-2* and *BnPAP12-6*) were obtained. Full-length cDNA of *BnPAP12-2* was successfully cloned by member-specific colony screening, while screening for *BnPAP12-6* full-length cDNA failed. As *BnPAP12-1* differed from *BnPAP12-7* by only 2 bp at the cDNA level and they might be allelic (see below), *BnPAP12-7* cDNA was considered to represent both *BnPAP12-1* and *BnPAP12-7*.

## Nucleotide-level characterization of *BnPAP12* genes

Full-length cDNAs and genomic sequences of seven *BnPAP12* members were 1675–1860 and 2326–2644 bp, respectively, 5′ UTR 65–74 bp, ORF (including stop codon TAA) 1410–1581 bp, and 3′ UTR 180–304 bp, respectively (Fig. 1; Supplementary Fig. 1). Start- and stop-codon positions of *BnPAP12* genes were identical to those of *AtPAP12*, except for 18-bp delay of stop codons of *BnPAP12-4*, *BnPAP12-5* and *BnPAP12-6*. *BnPAP12-1*, *BnPAP12-2*, *BnPAP12-3*, *BnPAP12-5* and *BnPAP12-7* resembled *AtPAP12* in number and positions of the six introns. *BnPAP12-4* and *BnPAP12-6* had two additional inserts (Insert I and Insert II), and a 133-bp novel intron (intron 1, $G_{200}$–$G_{332}$) was found within Insert I. Therefore, they contained seven introns, of which introns 2–7 corresponded to introns 1–6 of *AtPAP12* and other *BnPAP12* members, and intron 1 was recently evolved (Supplementary Fig. 1). All the introns follow standard GT…AG splicing boundaries. Interestingly, in *BnPAP12-4* and *BnPAP12-6*, a 36-bp conservative coding unit 5′-TTTACCTGACGACATGCCATTAGACAGCGATGTCTT-3′ at 10 bp downstream of the Insert I was repeated four times. The four tandem repeats (TRs, at $T_{369}$–$T_{404}$, $T_{405}$–$T_{440}$, $T_{441}$–$T_{476}$ and $G_{477}$–$T_{512}$) showed little sequence divergence from each other and formed a minisatellite. Since this unit is not repeated in other *BnPAP12* members and *AtPAP12*, the three extra TRs in *BnPAP12-4* and *BnPAP12-6* form Insert II (Supplementary Fig. 2A).




| Gene name | DNA (bp) | cDNA (bp) | 5'UTR (bp) | Coding region (bp) | Position of introns (bp) | | | | | | 3'UTR (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 1 | 2 | 3 | 4 | 5 | 6 | |
| *BnPAP12-1* | 2435 | 1827 | 67 | 1410 | 455~694 | 807~886 | 1117~1184 | 1324~1416 | 1638~1731 | 1854~1932 | 304 |
| *BnPAP12-2* | 2343 | 1708 | 67 | 1410 | 455~694 | 807~886 | 1117~1184 | 1324~1416 | 1638~1715 | 1838~1913 | 231 |
| *BnPAP12-3* | 2423 | 1766 | 67 | 1410 | 455~657 | 770~867 | 1098~1182 | 1324~1424 | 1646~1735 | 1858~1935 | 289 |
| *BnPAP12-5* | 2326 | 1699 | 67 | 1425 | 455~694 | 807~886 | 1117~1200 | 1340~1417 | 1639~1729 | 1852~1929 | 180 |
| *BnPAP12-5PM1* | | 1766 | 67 | 1194 | | | | | unspliced | | |
| *BnPAP12-5PM2* | | 1928 | 67 | 741 | | | unspliced | | unspliced | unspliced | |
| *BnPAP12-7* | 2431 | 1781 | 67 | 1410 | 455~694 | 807~886 | 1117~1184 | 1324~1412 | 1634~1727 | 1850~1928 | 304 |
| *BnPAP12-7PM* | | 1954 | 67 | 1101 | | | | | unspliced | unspliced | |




| Gene name | DNA (bp) | cDNA (bp) | 5'UTR (bp) | Coding region (bp) | Position of introns (bp) | | | | | | | 3'UTR (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| *BnPAP12-4* | 2644 | 1860 | 74 | 1581 | 200~332 | 748~983 | 1096~1175 | 1406~1471 | 1611~1710 | 1932~2022 | 2145~2222 | 205 |
| *BnPAP12-4PM* | | 2084 | 74 | 126 | unspliced | | | | | unspliced | | |
| *BnPAP12-6* | 2619 | 1835 | 74 | 1581 | 200~332 | 748~983 | 1096~1175 | 1406~1471 | 1611~1710 | 1932~2022 | 2145~2222 | 180 |

**Fig. 1** Basic structure parameters of the seven *BnPAP12* genes. I: intron

*BnPAP12* genes had obviously higher G+C contents in ORFs (44.68–45.35%) and 5′ UTRs (41.89–43.28%) than in 3′ UTRs (27.34–29.76%) and introns (25.00–38.46%). In 5′ UTRs, there was a purine-stretch 5′-AGA-AAAGAAAAGAGAGAGAAG-3′ (Supplementary Fig. 2B). In 3′ UTRs, alternative polyadenylation sites existed right after $C_{2238}$, $G_{2264}$, $T_{2278}$ and $C_{2286}$ of *BnPAP12-2* and $C_{2297}$ and $T_{2314}$ of *BnPAP12-7* (Supplementary Fig. 2C). The 3′ UTRs of *BnPAP12-1*, *BnPAP12-2* and *BnPAP12-7* contained three canonical or non-canonical polyadenylation signals, and other four members contained five. In *BnPAP12-4PM*, *BnPAP12-7PM*, *BnPAP12-5PM1* and *BnPAP12-5PM2*, the ORFs were truncated due to intron-derived stop codons or frameshift.

Based on sequence similarity and gene structure, *BnPAP12* genes could be divided into three groups. *BnPAP12-1*, *BnPAP12-2*, *BnPAP12-3* and *BnPAP12-7* formed Group I and showed 94.1–99.9% identity to each other. In Group II, *BnPAP12-4* was 97.7% identical to *BnPAP12-6*. The singleton *BnPAP12-5* was in Group III since its 5′ half was similar to that of Group I and its 3′ half to that of Group II (Supplementary Fig. 1). NCBI BLASTn indicated that mRNAs of *BnPAP12* genes shared 87% local identity with *AtPAP12*. When pairwise-aligned, *BnPAP12-4*, *BnPAP12-6* and the rest of the members showed 77.3, 76.9 and 85.6–87.4% of full-coding-region identity to *AtPAP12*, and 5′-UTR identity (68.2–71.6%) was higher than 3′-UTR identity (45.2–47.3%).

## Distribution of the MMTS among Brassicaceae species

PCR results showed that eight of the assessed 26 Brassicaceae species contained the MMTS (Fig. 2). Of the six U-triangle species, diploids *B. rapa* (AA) and *B. oleracea* (CC) and allopolyploids *B. napus* (AACC), *B. carinata* (BBCC) and *B. juncea* (AABB) possessed the MMTS for simultaneous amplification of the 219-bp and the 505-bp bands from them, but *B. nigra* (BB) did not contain the MMTS for amplification of just a single 219-bp band. *B. macrocarpa*, *B. villosa* and *B. cretica* also contained the MMTS, whereas the rest *Brassica* species did not. The MMTS was not detected in all other genera such as *Arabidopsis*, *Diplotaxis*, *Erucastrum*, *Eruca*, *Moricandia*, *Raphanus* and *Sinapis*.

## Southern hybridization analysis

After moderately stringent hybridization and immunological detection in Southern analysis, *Eco*RI and *Hin*dIII digestions both resulted in five distinct hybridization bands, while four bands were detected by *Eco*RV and *Xba*I digestions. Hybridization bands of *Sac*I digestion were not clear enough, but at least two bands could be distinguished (Fig. 3).

## Conservation and features of inferred BnPAP12 proteins

Like *AtPAP12*, mature mRNAs of *BnPAP12-1*, *BnPAP12-2*, *BnPAP12-3* and *BnPAP12-7* encoded proteins of 469 amino acids (aa). Due to extra coding in Inserts I and II and 18-bp delay of the stop codon, the inferred BnPAP12-4 and BnPAP12-6 proteins had 526 amino acids and were 57-aa larger than other members. The *BnPAP12-5* encoded 475 aa containing six additional amino acids at the C-terminus like in BnPAP12-4 and BnPAP12-6, while it lacked the internal 51-aa insertion. BnPAP12 proteins possessed theoretical molecular weights (MWs) of 53.95–60.39 kDa and isoelectric points (pIs) of 5.90–6.26. ORFs of *BnPAP12-4PM*, *BnPAP12-5PM1*, *BnPAP12-5PM2* and *BnPAP12-7PM* encode truncated polypeptides of 42, 397, 246 and 366 aa, respectively (Table 2).

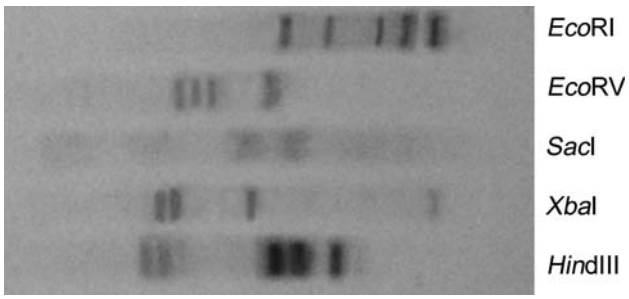



**Fig. 3** Southern hybridization detection of *BnPAP12* members in *B. napus*

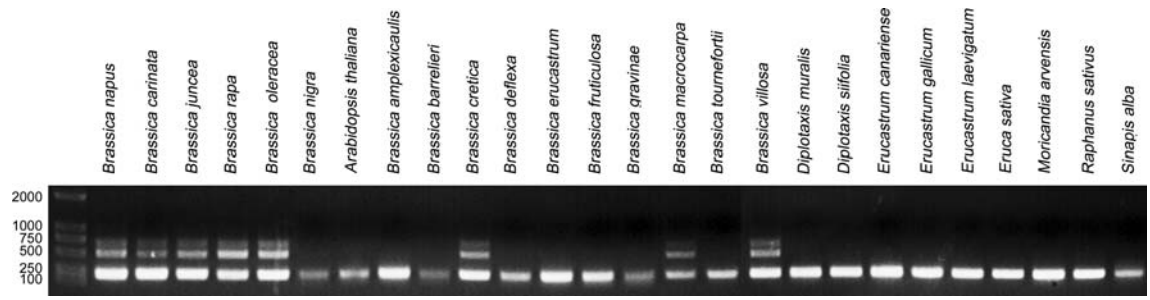



**Fig. 2** PCR detection of distribution of the MMTS among 26 Brassicaceae species

**Table 2** Predicted primary parameters of inferred BnPAP12 proteins

| BnPAP12 proteins | Amino acid numbers (aa) | MW (kDa) | pI | Domain | | Signal peptide | | Number of phosphorylation sites | Number of N-glycosylation sites |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Name | Position | Length (aa) | Cleavage site | | |
| BnPAP12-1 | 469 | 54.09 | 6.26 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 36 | 3 |
| BnPAP12-2 | 469 | 54.07 | 6.07 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 36 | 3 |
| BnPAP12-3 | 469 | 53.95 | 6.25 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 36 | 3 |
| BnPAP12-4 | 526 | 60.32 | 5.99 | Metallophos | $Y_{212}$–$Y_{411}$ | 28 | $CDG_{28}$-$G_{29}I$ | 40 | 2 |
| BnPAP12-4PM | 42 | 4.64 | 9.68 | No | No | 28 | $CDG_{28}$-$G_{29}I$ | 2 | 0 |
| BnPAP12-4TC | 430 | 50.11 | 8.02 | Metallophos | $Y_{194}$–$Y_{393}$ | No | No | 30 | 2 |
| BnPAP12-5 | 475 | 54.69 | 5.90 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 37 | 3 |
| BnPAP12-5PM1 | 397 | 46.11 | 7.31 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 29 | 3 |
| BnPAP12-5PM2 | 246 | 28.29 | 5.62 | Metallophos | $Y_{161}$–$I_{246}$ | 28 | $CDG_{28}$-$G_{29}I$ | 15 | 2 |
| BnPAP12-6 | 526 | 60.39 | 5.95 | Metallophos | $Y_{212}$–$Y_{411}$ | 28 | $CDG_{28}$-$G_{29}I$ | 41 | 2 |
| BnPAP12-7 | 469 | 54.07 | 6.26 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 36 | 3 |
| BnPAP12-7PM | 366 | 42.60 | 6.28 | Metallophos | $Y_{161}$–$Y_{360}$ | 28 | $CDG_{28}$-$G_{29}I$ | 27 | 3 |

BnPAP12 proteins showed 76.5–88.1% identity and 79.6–91.3% positive to AtPAP12 (NP_180287), and 58.8–75.9% identity and 67.7–85.2% positive to non-Cruciferous PAPs such as sweet potato IbPAP3 (CAA07280) and IbPAP1 (AAF19821), tobacco NtPAP19 (BAC55156) and NtPAP4 (BAC55154), etc. Phylogenetic analysis and sequence alignment divided BnPAP12 proteins into three groups (Fig. 4). Group-I proteins (BnPAP12-1, BnPAP12-2, BnPAP12-3 and BnPAP12-7) shared 97.4–99.8% identity and 98.7–100% positive with one another, and Group-II proteins (BnPAP12-4 and BnPAP12-6) shared 98.7% identity and 99.4% positive. But the members between these two groups shared only 83.7–87.8% identity and 85.2–88.4% positive. Group-III (BnPAP12-5) was an intermediate between Groups I and II, with 93.0–93.8% identity and 94.8–95.3% positive to Group-I and 86.5–87.8% identity and 87.8–88.4% positive to Group-II. All the BnPAP12 members formed a small branch, which further clustered with AtPAP12 to form a distinct Brassicaceae PAP12 branch. All other analyzed PAP proteins from *Arabidopsis* and other plants showed distinctly farther distances to BnPAP12 proteins.

NCBI conserved domain (CD) search detected a pfam00149 (metallophos) conserved domain located at $Y_{212}$–$Y_{411}$ of BnPAP12-4 and BnPAP12-6 and at $Y_{161}$–$Y_{360}$ of other BnPAP12 proteins (Supplementary Fig. 3). BnPAP12 proteins were as conserved as other PAPs at the seven metal binding residues and five blocks of conserved motifs in the metallophos domain (Strater et al. 1995; Li et al. 2002).

SignalP 3.0 predicted that BnPAP12 proteins would contain a signal peptide most likely cleaved at $G_{28}$–$G_{29}$ (Supplementary Fig. 3), and TargetP 1.1 predicted them to be secretory proteins. Predicted by TMpred, all proteins had a strong N-terminal transmembrane helix ($L_7$–$G_{27}$). NetPhos 2.0 predicted 35–40 significant phosphorylation sites in each member (17–21 for S, 5–6 for T, and 12–14 for Y). NetNGlyc 1.0 predicted that all BnPAP12 proteins had 2–3 significant N-glycosylation sites.

Conclusively, BnPAP12 proteins are typical secretory plant PAP proteins with highest similarities to AtPAP12.

## Secondary and tertiary structures of BnPAP12 proteins

Analyzed by SOPMA, BnPAP12 proteins had 42.43–46.39% random coils, 25.16–26.81% extended strands, 18.25–21.96% alpha helices, and 8.56–9.59% beta turns. Besides the N-terminal signal peptide helix, the major helices were in the about 180-aa C-terminal region. The seven metal binding sites were all composed of random coils (Supplementary Fig. 4).

Predicted by Swiss-Model, the tertiary structures of BnPAP12 proteins were very similar to one another and showed 72% identity to IbPAP3 (PDB ID code 1XZW; Fig. 5). Each BnPAP12 monomer (BnPAP12-1 as an example) contained two domains. The N-terminal domain ($D_{43}$–$D_{160}$, red) was mainly composed of two sandwiched $\beta$ sheets, each formed by three anti-parallel extended strands ($\beta$1, $\beta$2 and $\beta$7; $\beta$3, $\beta$8 and $\beta$9). One edge of the sandwich was shielded by the other three extended strands ($\beta$4–$\beta$6). The C-terminal domain ($D_{161}$–$D_{458}$, green) was the major domain and was composed of two large sandwiched $\beta$–$\alpha$–$\beta$–$\alpha$–$\beta$ folds ($\beta$10–$\alpha$1–$\beta$11–$\alpha$2–$\beta$13; $\beta$13–$\beta$14–$\alpha$6–$\beta$15–$\beta$16–$\alpha$7–$\beta$17–$\beta$18) that were connected by the three continuous helices $\alpha$3–$\alpha$4–$\alpha$5. Sandwiched folds formed the core and active center and each contained three parallel strands

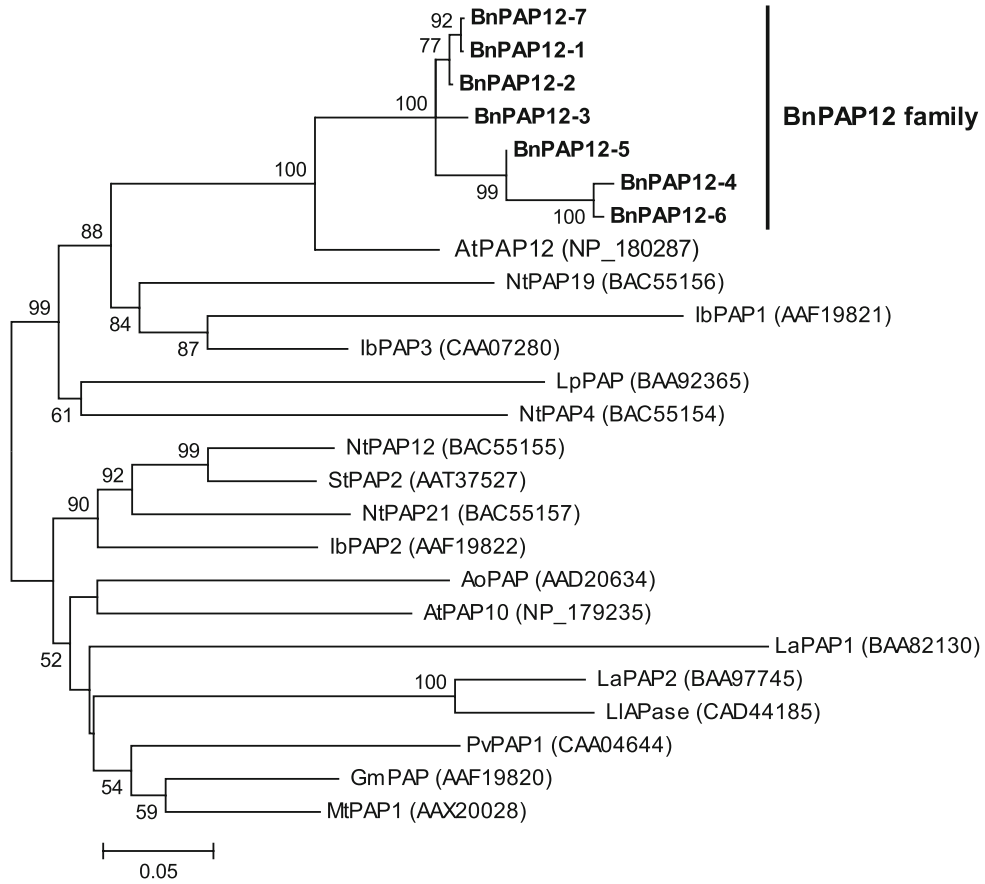



**Fig. 4** Phylogenetic relationships of inferred BnPAP12 proteins and other plant PAPs. *Anchusa officinalis* AoPAP (AAD20634); *A. thaliana* AtPAP10 (NP_179235) and AtPAP12 (NP_180287); *Glycine max* GmPAP (AAF19820); *Ipomoea batatas* IbPAP1 (AAF19821), IbPAP2 (AAF19822) and IbPAP3 (CAA07280); *Lupinus albus* LaPAP1 (BAA82130) and LaPAP2 (BAA97745); *Lupinus luteus* LlAPase (CAD44185); *Landoltia punctata* LpPAP (BAA92365); *Medicago truncatula* MtPAP1 (AAX20028); *Nicotiana tabacum* NtPAP4 (BAC55154), NtPAP12 (BAC55155), NtPAP19 (BAC55156) and NtPAP21 (BAC55157); *Phaseolus vulgaris* PvPAP1 (CAA04644); *Solanum tuberosum* StPAP2 (AAT37527). This tree is constructed by Neighbor-Joining method with *p*-distance. The number for each interior branch is the percent bootstraps value (1000 replicates), and only values greater than 50% are shown. The scale bar indicates the estimated number of amino acid substitutions per site

($\beta$10–$\beta$12; $\beta$15–$\beta$16 and $\beta$18). The binuclear metal center of BnPAP12 proteins was Fe(III)–Mn(II). The Fe ion was coordinated by $D_{168}$, $Y_{200}$ and $H_{358}$, while $N_{234}$, $H_{319}$ and $H_{356}$ coordinated the Mn ion. The two metal ions were bridged by the carboxy group of $D_{197}$. Most of the known plant PAPs contained cysteines around position 340–370 and were able to form a disulfide bridge between two subunits. Existence of $C_{378}$ between $\beta$18 and $\beta$19 indicated that BnPAP12 proteins could form homodimers (Olczak et al. 2003).

## Tissue specificity and Pi-starvation induced expression of *BnPAP12* genes

Stems had the highest overall expression levels of *BnPAP12* family, followed by buds, silique pericarps, hypocotyls, flowers, roots, leaves, 10-day seeds, cotyledons and 20-day seeds, whereas 30-day seeds had the lowest expression (Fig. 6). *BnPAP12-6* transcript was not detectable in any organ, and *BnPAP12-1/7* only had weak expression in hypocotyls. Transcripts of *BnPAP12-2* and *BnPAP12-3* were observed in all of the 11 organs. *BnPAP12-2* was expressed intensively in stems, buds, flowers and silique pericarps, moderately in hypocotyls and 10-day seeds, and weakly in other organs. *BnPAP12-3* expression was strong in stems, buds, silique pericarps, flowers, hypocotyls and leaves, moderate in roots and 10-day seeds, and weak in cotyledons, 20-day seeds and 30-day seeds. Expression of *BnPAP12-4* was highest in stems and roots, strong in all other vegetative organs and early reproductive organs (buds and flowers), very weak in silique pericarps, and nearly undetectable in seeds. *BnPAP12-5* expression was high in hypocotyls and stems, moderate in leaves and flowers, weak in buds, silique pericarps, roots and cotyledons, and non-detectable in seeds. Cloning and sequencing of RT-PCR bands of individual *BnPAP12* members in

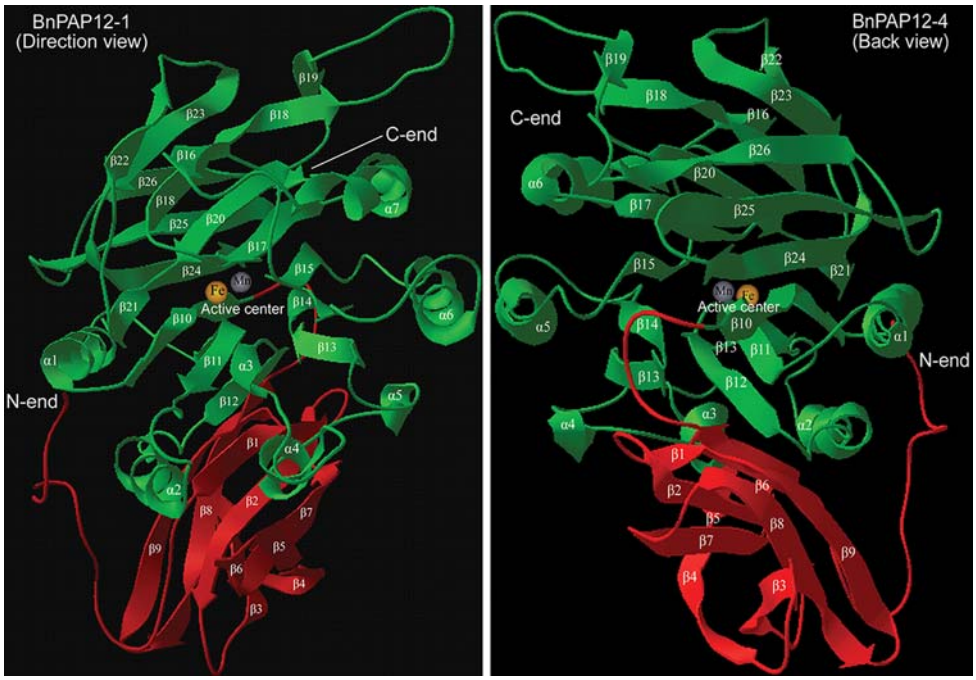



**Fig. 5** Ribbon diagrams of BnPAP12-1 and BnPAP12-4 monomer predicted by SWISS-MODEL and made with Swiss-PdbViewer. Tertiary structures of BnPAP12-1 and BnPAP12-4 are direction view and back view respectively. The two metal irons are shown as spheres (Fe−yellow, Mn−gray). The loop between the strands $\beta$18 to $\beta$19 contains the cysteine residue forming the disulfide bridge to the other subunit. Additional interactions between both subunits occur between the helices $\alpha$7 and between the loops from residues 253 to 260. The N-terminal domain is located at the bottom of the figure and colored red. Labeling is according to Supplementary Fig. 3

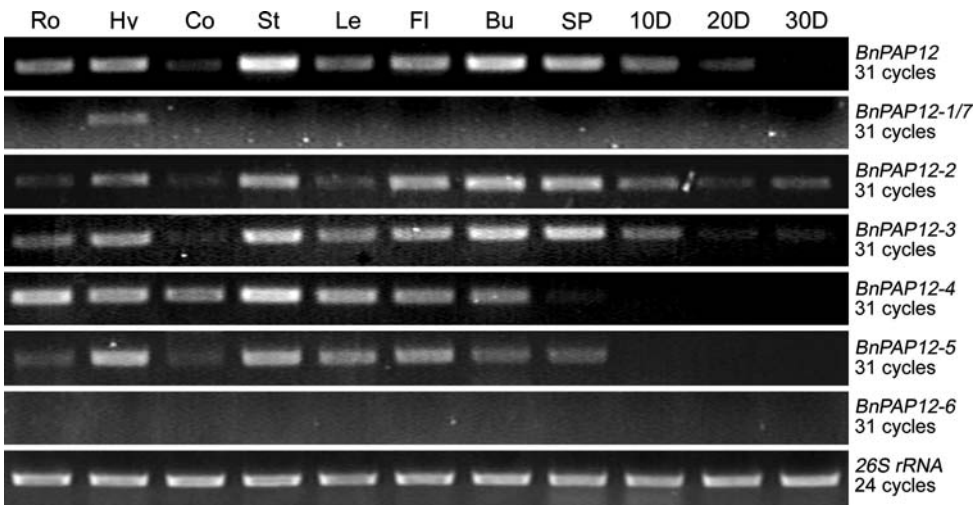



**Fig. 6** RT-PCR detection of transcription levels of *BnPAP12* members in various organs of *B. napus*. Ro: root; Hy: hypocotyl; Co cotyledon; St: stem; Le: leaf; Fl: flower; Bu: bud; SP: silique pericarp; 10D: 10-day seeds; 20D: 20-day seeds; and 30D: 30-day seeds

hypocotyls did not find cross-amplification, validating the RT-PCR specificity of each member.

Similar patterns of Pi-starvation induced overall *BnPAP12* expression were revealed between 24 and 31 cycles of amplifications and between SL and SR. *BnPAP12* overall expression in SL and SR continuously increased after 12 h of Pi starvation, reaching maximal levels after about 4 days of induction. However, it declined to near basal levels after 4 days of Pi-resupply (Fig. 7).

Integration of expression patterns of individual *BnPAP12* members coincided with the aforementioned overall expression pattern, but different members responded distinctly to Pi-starvation (Fig. 7). *BnPAP12-6* was not expressed in both organs under any Pi-starvation induction, while other members responded actively.

In SL, basal expression under Pi-sufficient condition was mainly conditioned by *BnPAP12-1/7*, *BnPAP12-4* and *BnPAP12-3*, and transcription of all the active members was induced to maximal levels and maintained under Pi-starvation condition. But different members responded differentially in regard to the speed. *BnPAP12-4* needed only 12-h induction to reach near maximal expression level. *BnPAP12-5* and *BnPAP12-3* needed only 1 day, whereas *BnPAP12-1/7* and *BnPAP12-2* needed 2 and 4 days, respectively. After 4 days of Pi-resupply, the transcription droped to basal (*BnPAP12-2* and *BnPAP12-2*), near basal

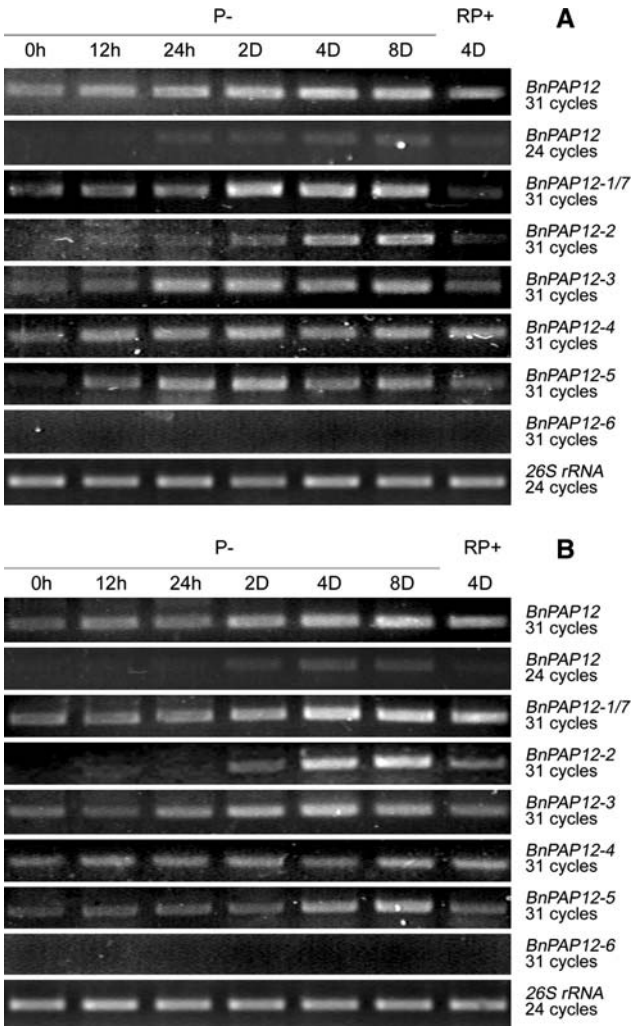



**Fig. 7** Pi-starvation induced expression of *BnPAP12* genes in SL (A) and SR (B) Total RNA was isolated from stressed SL and SR at 0 h, 12 h, 24 h, 2 days, 4 days and 8 days of Pi-starvation treatment (P-), and then sampled again 4 days after Pi-resupply (RP+). The entire experiments were repeated three times

(*BnPAP12-4* and *BnPAP12-5*), or even below basal (*BnPAP12-1/7*) levels.

In SR, *BnPAP12-1/7*, *BnPAP12-3*, *BnPAP12-4* and *BnPAP12-5* contributed to the basal expression. Induction trends of individual members in SR were similar to those in SL, but a little slower. It needed about 2–4 days, 2–4 days, 1–2 days, 12 h and 4 days of Pi-starvation for *BnPAP12-1/7*, *BnPAP12-2*, *BnPAP12-3*, *BnPAP12-4* and *BnPAP12-5*, respectively, to have a distinct increase in expression, and most of them reached maximal levels after 4 days of induction. After 4 days of Pi-resupply, transcription of all the active members in SR declined but was still obviously higher than the basal levels.

Of all the *BnPAP12* members, *BnPAP12-4* was more constitutive and stable than others, with relatively higher basal expression in various organs (except siliques) and less induced increment especially in SR.

## Discussion

### Possible triplication and homoeologous inter-allelic exchange of *BnPAP12*

Comparative mapping and sequencing have disclosed extensive collinearity between *Arabidopsis* and *Brassica* species for the chromosome segments (Kowalski et al. 1994; Quiros et al. 2001). The ancestor of the tribe Brassiceae has tripled its genome after divergence with *Arabidopsis* ancestor (Cavell et al. 1998). Resulted from the fusion of *B. oleracea* and *B. rapa* progenitors, the genome of *B. napus* is more than six times the *A. thaliana* genome (Johnston et al. 2005). It is suggested that *B. napus* might have about 2–8 orthologous genes corresponding to each single-copy gene from *A. thaliana* (Cavell et al. 1998). Roughly conforming to the triplication hypothesis, this study has isolated seven *BnPAP12* genes orthologous to the single *AtPAP12* gene.

The seven *BnPAP12* genes might represent at least five unique alleles. First, Southern hybridization resulted in a maximum of five different bands, and none of the *BnPAP12* members had cutting sites of the five enzymes in the probe region (Fig. 3). Second, it is of little doubt that *BnPAP12-1* to *BnPAP12-5* represent five unique alleles based on sequence similarity and gene structure, while it is not conclusive whether *BnPAP12-6* and *BnPAP12-7* also represent independent alleles or are allelic to two of the former five alleles (Supplementary Fig. 1). *BnPAP12-1* shares 99.6% identity to *BnPAP12-7*, and they differ from each other by only 7 bp in intron 4 and 2 bp in exon 5. Interestingly, careful scrutiny indicated that two inter-allelic exchange events occurred among Group I genes in the region containing this 7 bp, leading to two local donor/acceptor pairs: *BnPAP12-1/BnPAP12-2* and *BnPAP12-3/BnPAP12-7*. So this 7-bp difference is caused by inter-allelic exchange instead of successive point mutations. *BnPAP12-1* and *BnPAP12-7* either are from homologous chromosomes (allelic) or signify a recent duplication event (non-allelic). Exchange among more divergent alleles of *BnPAP12* genes also exists. *BnPAP12-5* (Group III), though regarded as an independent allele, is evolved from a chimera of two structurally distant groups (Group I and II). Besides, the 1–1885-bp region of *BnPAP12-6* is just 1-bp different from *BnPAP12-4*, and its 1886–2619-bp region is completely identical to *BnPAP12-5*. Thus, *BnPAP12-6* might be allelic to *BnPAP12-4* or *BnPAP12-5*. *BnPAP12-6* is silent in any RT-PCR detection, implying that it is recessive to its allelic partner. The fusion process of *B. rapa* and *B. oleracea* is accompanied by complicated chromosomal rearrangements (Lagercrantz 1998). Our results indicate that rearrangements caused by homoeologous inter-allelic exchange have been involved in the reconstruction of the *Brassica* genomes.

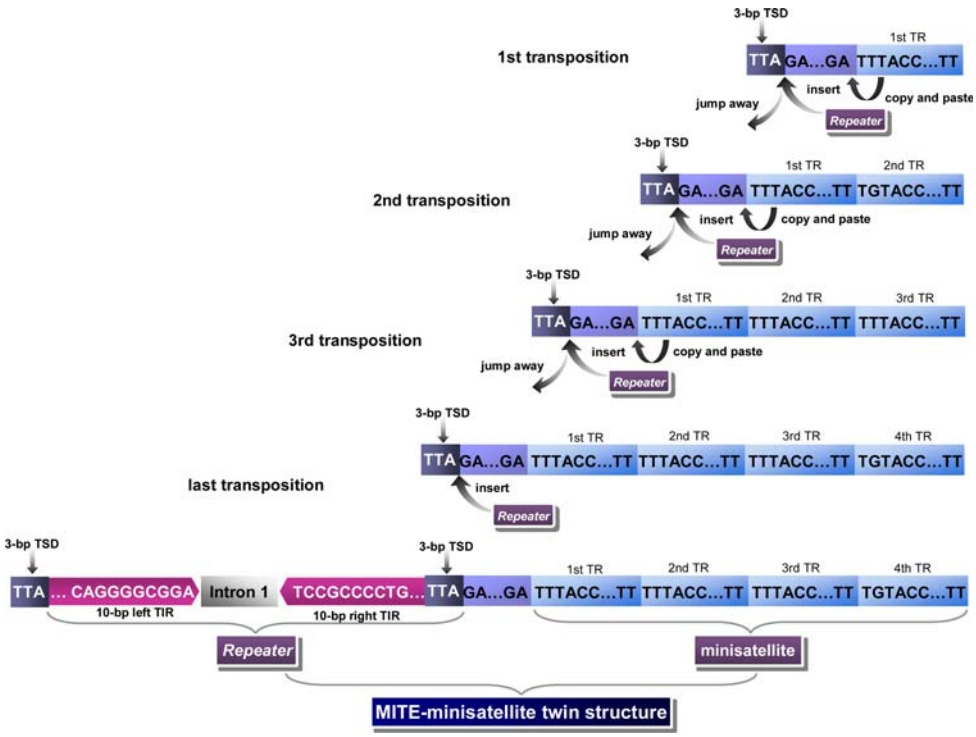



**Fig. 8** Possible origin of the MMTS. Each transposition event contains five possible steps. In step 1, *Repeater* recognized the TSD TTA; it step 2, *Repeater* inserted into genomic sequence just after TSD; in step 3, *Repeater* recognized the TR sequence; in step 4, *Repeater* copied the TR sequence and pasted the clone just prior to the progenitor TR; in step 5, *Repeater* jumped away from the inserted locus. Based on base divergence, the 3rd TR generated first, then the 2nd and the 4th. In the last transposition, *Repeater* retained, forming the MMTS within *BnPAP12-4* and *BnPAP12-6*

## Recently evolved novel MMTS in *B. rapa/oleracea* lineage and its value

The MITEs constitute a particular type of transposable elements (TEs), with characteristics of small size (less than 500 bp) and conserved terminal inverted repeats (TIRs) (Casacuberta and Santiago 2003). They are also characterized by high copy number, A/T-richness, individual target-site preference, lack of coding capacity and stable secondary structure. Based on the target-site duplications (TSDs), MITEs have been divided into *Tourist*-like (3-bp TSDs, TTA/TAA) and *Stowaway*-like groups (2-bp TSDs, TA) (Feschotte et al. 2002).

In this study, the most interesting structures are the two inserts within *BnPAP12-4* and *BnPAP12-6*. The recently evolved intron 1 ($G_{200}$–$G_{332}$) within Insert I is flanked by two 10-bp TIRs ($C_{185}AGGGGCGGA_{194}$ and $T_{339}CCGCCCCTG_{348}$) and typical TSDs of *Tourist*-like MITEs ($T_{178}TA_{180}$ and $T_{356}TA_{358}$), suggesting that the $G_{181}$–$A_{358}$ region may be a novel 178-bp *Tourist*-like MITE (Fig. 8; Supplementary Fig. 2A). Since a minisatellite containing four tandem repeats (TRs) of a 36-bp unit tightly follows this MITE with a spacer of only 10 bp, the MITE is named as *Repeater*. *Repeater* is A/T-rich (60.67%) and has a potential to form stable secondary structures ($\Delta G = -20.94$ kcal/mol, predicted by Quikfold, http://www.bioinfo.rpi.edu/applications/hybrid/quikfold.php).

Different from common MITEs, *Repeater* is tightly associated with a minisatellite to form a novel MMTS. Current minisatellite formation hypotheses cannot explain well the origin of the minisatellite reported in this case, thus its relation to transposition events of *Repeater* can be assumed (Charlesworth et al. 1994). It is unknown whether the MMTS has a high copy number in *B. napus* genome, and its evolutionary mechanism is also unclear; but it is known that it has been originated quite recently in *B. rapa/oleracea* lineage after its divergence with *B. nigra* and before the *rapa-oleracea* split, since only *B. rapa* and *B. oleracea*, their progeny species *B. napus*, *B. juncea* and *B. carinata*, as well as their very close relatives *B. macrocarpa*, *B. villosa* and *B. cretica* contain this structure (Fig. 2).

The 2nd, 3rd and 4th TRs of the minisatellite have 2, 10 and 2 bp of difference from the 1st one, respectively, which is conserved in *AtPAP12* and *BnPAP12* members. Thus, the four TRs are formed by 2–3 independent repeating events, which might have been caused by 2–3 times of transposition. Each transposition event might involve five possible steps (Fig. 8; Supplementary Fig. 2A). This newly evolved MMTS can be used to identify very closely related species of *B. rapa/oleracea* lineage and distinguish them from other *Brassica* species. This special structure also provides a model to explore the activity of *Repeater*-type MITEs, especially its possible involvement in minisatellite formation and gene structure evolution.

### Divergence of organ specificity and Pi-starvation induced expression among *BnPAP12* genes

Like *AtPAP12*, *BnPAP12* genes are expressed strongly in vascular and early reproductive organs, implying that they most likely participate in Pi transfer under normal conditions (Zhu et al. 2005; Fig. 6). Promoter of *AtPAP12* is specifically activated by low P levels, and induced *AtPAP12* expression is first evident in leaves after 2 days of Pi-starvation, and then in roots about 2 days later (Haran et al. 2000). Trends of Pi-starvation induced expression of *BnPAP12* genes imitate *AtPAP12*. Combining the information from gene and protein structures, *BnPAP12* genes are assumed to possess biological function similar to *AtPAP12* under both normal and Pi-starvation conditions.

Though sharing high sequence similarities to each other, *BnPAP12* members show fast divergence and complementation in organ specificity. Specifically expressed in hypocotyls, SL and SR, *BnPAP12-1/7* is believed to be involved in vascular Pi transfer during seed germination and early seedling growth. *BnPAP12-2* and *BnPAP12-3* are the only members showing expression in developing seed, but *BnPAP12-2* is more specific for Pi assimilation in reproductive organs, while *BnPAP12-3* is important for Pi assimilation in both reproductive organs and leaves. *BnPAP12-4* distinguishes itself from other members for its intensive expression in roots and cotyledons besides in other organs except siliques. It is assumed to be involved in root Pi absorption, cotyledon Pi export and Pi transfer to sink organs via the vascular system. Insertion of the MMTS in the "standard" exon 1 does not impair transcription and coding abilities of *BnPAP12-4*, but it needs to be clarified whether the MMTS has any contribution to the intensive expression of *BnPAP12-4* in root and cotyledon as well as to its rapid induction and slow declining. *BnPAP12-5* is as if specified for transfer of Pi to vegetative organs and early-stage reproductive organs via vascular system. Inferred BnPAP12-6 protein shows no structural defect, but *BnPAP12-6* is not transcribed in any organ under any Pi condition, possibly because of dominant allelic suppression by *BnPAP12-4* or *BnPAP12-5*.

*BnPAP12* members also show differences in Pi-starvation induced expression. In SL, *BnPAP12-4* transcription is induced to peak level after 12 h, followed by *BnPAP12-5* (24 h) and *BnPAP12-3* (24 h), while *BnPAP12-1/7* and *BnPAP12-2* need 2 and 4 days, respectively. In SR, induction of *BnPAP12-4* transcription to peak level also needs only 12 h, while all other members need about 4 days. Besides, induction of *BnPAP12-4* is the most durable, since after 4 days of Pi-resupply its transcription shows little decrease in both SL and SR. Expression of all other members declines to basal or near basal levels in SL and to certain extents in SR after 4 days of Pi-resupply.

### Possible *cis*-regulations of some *BnPAP12* genes

In sequencing, it was found that four intron-retaining mRNAs encoded premature proteins BnPAP12-4PM, BnPAP12-5PM1, BnPAP12-5PM2 and BnPAP12-7PM (Fig. 1). Lacking almost the whole structure of a PAP, the 42-aa BnPAP12-4PM is certainly non-functional. BnPAP12-5PM1 and BnPAP12-7PM both lack the C-terminal 112 aa but contain the whole pfam00149 conserved domain as compared with mature proteins. Clarification of the influence of this truncation on the catalytic ability still needs experiment clues. Lacking the 232-aa C-terminal region including 117 aa of the pfam00149 conserved domain, BnPAP12-5PM2 is possibly completely impaired in function. In higher eukaryotes, alternative splicing is a kind of post-transcriptional regulation mechanism to increase transcriptome and proteome diversification, and intron retention may be the most ancient form (Ast 2004). Isolation of flavonoid pathway genes from *B. napus* in our laboratory shows that alternative splicing is generally observed on regulatory genes, e.g. *BnTT2-2PM* in *BnTT2* gene family (Wei et al. 2007). *BnPAP12* genes are structural genes, but alternative splicing with the form of intron retention also exists. However, it is not clear whether this is a kind of active *cis*-regulation for limiting the over-induction of *BnPAP12* mRNAs, or it is just a result of leaky splicing.

In the 5′ UTRs of *BnPAP12* genes, the 21–23 bp purine-stretches prior to the start codon are quite different from the corresponding sequence in *AtPAP12* and show distinct directional evolution (Supplementary Fig. 2B). Whether these structures represent a type of purine-rich start codon context deserves characterization. In the 3′ UTRs, there are 3–5 canonical or non-canonical polyadenylation signals in *BnPAP12* genes, and their relationship to alternative polyadenylation (5 sites in *BnPAP12-2* and 3 sites in *BnPAP12-7*) needs further study (Supplementary Fig. 2C).

**Acknowledgments** We are indebted to Professor Takeshi NISHIO, Laboratory Plant Breeding and Genetics, Graduate School of Agriculture, Tohoku University, Sendai, 981-8555, Japan, for kind provision of stock seeds of Brassicaceae species, and to Dr Genyi Li from University of Manitoba and Dr Guoping Chen from Chongqing University for critical reading of the manuscript. This research was supported by the National High Technology Research and Development Program of China (2006AA10A113 and 2006AA100106).

## References

Ast G (2004) How did alternative splicing evolve? Nat Rev Genet 5:773–782

Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1–11

Cavell AC, Lydiate DJ, Parkin IAP, Dean C, Trick M (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. Genome 41:62–69

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

del Pozo JC, Allona I, Rubio V, Leyva A, de la Pena A, Aragoncillo C, Paz-Ares J (1999) 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

Duff SMG, Sarath G, Plaxton WC (1994) The role of acid phosphatases in plant phosphorus metabolism. Physiol Plant 90:791–800

Durmus A, Eicken C, Sift BH, Kratel A, Kappl R, Huttermann J, Krebs B (1999a) The active site of purple acid phosphatase from sweet potatoes (*Ipomoea batatas*) metal content and spectroscopic characterization. Eur J Biochem 260:709–716

Durmus A, Eicken C, Spener F, Krebs B (1999b) Cloning and comparative protein modeling of two purple acid phosphatase isozymes from sweet potatoes (*Ipomoea batatas*). Biochim Biophys Acta 1434:202–209

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable element: where genetics meets genomics. Nat Rev Genet 3:329–341

Haran S, Logendra S, Seskar M, Bratanova M, Raskin I (2000) Characterization of *Arabidopsis* acid phosphatase promoter and regulation of acid phosphatase expression. Plant Physiol 124:615–626

Hoagland DR, Arnon DL (1950) The water culture method for growing plants without soil. Calif Agric Exp Sta Circ 347:32

Jaakola L, Pirttil AM, Halonen M, Hohtola A (2001) Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit. Mol Biotechnol 19:201–203

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in Brassicaceae. Ann Bot (Lond) 95:229–235

Klabunde T, Strater N, Krebs B, Witzel H (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

Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138:499–510

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Li DP, Zhu HF, Liu KF, Liu X, Leggewie G, Udvardi M, Wang DW (2002) Purple acid phosphatases of *Arabidopsis thaliana*: comparative analysis and differential regulation by phosphate deprivation. J Biol Chem 277:27772–27781

Liao H, Wong FL, Phang TH, Cheung MY, Li WY, Shao G, Yan X, Lam HM (2003) *GmPAP3*, a novel purple acid phosphatase-like gene in soybean induced by NaCl stress but not phosphorus deficiency. Gene 318:103–111

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* Indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

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

Quiros CF, Grellet F, Sadowski J, Suzuki T, Li G, Wroblewski T (2001) *Arabidopsis* and *Brassica* comparative genomics sequence: structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. Genetics 157:1321–1330

Schenk G, Gahan LR, Carrington LE, Mitic N, Valizadeh M, Hamilton SE, de Jersey J, Guddat LW (2005) Phosphate forms an unusual tripodal complex with the Fe–Mn center of sweet potato purple acid phosphatase. Proc Natl Acad Sci USA 102:273–278

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Skinner RJ, Todd AD (1998) Twenty-five years of monitoring pH and nutrient status of soils in England and Wales. Soil Use Man 14:162–169

Singh K, Raizada J, Bhardwaj P, Ghawana S, Rani A, Singh H, Kaul K, Kumar S (2004) 26S rRNA-based internal control gene primer pair for reverse transcription-polymerase chain reaction-based quantitative expression studies in diverse plant species. Anal Biochem 335:330–333

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25:4876–4882

Wei YL, Li JN, Lu J, Tang ZL, Pu DC, Chai YR (2007) Molecular cloning of *Brassica napus TRANSPARENT TESTA 2* gene family encoding potential MYB regulatory proteins of proanthocyanidin biosynthesis. Mol Biol Rep 34:105–120

Xiao K, Harrison M, Wang ZY (2006) Cloning and characterization of a novel purple acid phosphatase gene (*MtPAP1*) from *Medicago truncatula* Barrel Medic. J Int Plant Biol 48:204–211

Zhu HF, Qian WQ, Lu XZ, Li DP, Liu X, Liu KF, Wang DW (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, Regierer B, Kossmann J, Frossard E, Amrhein N, Bucher M (2004) Differential expression of three purple acid phosphatases from potato. Plant Biol (Stuttg) 6:519–528