High mature grain phytase activity in the Triticeae has evolved by duplication followed by neofunctionalization of the purple acid phosphatase phytase (PAPhy) gene

Claus Krogh Madsen, Giuseppe Dionisio, Inger Bæksted Holme, Preben Bach Holm and Henrik Brinch-Pedersen

Aarhus University, Faculty of Science and Technology, Dept. of Molecular Biology and Genetics, Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark

* To whom correspondence should be addressed. E-mail: Henrik.BrichPedersen@agrsci.dk

Received 19 November 2012; Revised 3 March 2013; Accepted 8 April 2013

Abstract

The phytase activity in food and feedstuffs is an important nutritional parameter. Members of the Triticeae tribe accumulate purple acid phosphatase phytases (PAPhy) during grain filling. This accumulation elevates mature grain phytase activities (MGPA) up to levels between ~650 FTU/kg for barley and 6000 FTU/kg for rye. This is notably more than other cereals. For instance, rice, maize, and oat have MPGs below 100 FTU/kg. The cloning and characterization of the PAPhy gene complement from wheat, barley, rye, einkorn, and Aegilops tauschii is reported here. The Triticeae PAPhy genes generally consist of a set of paralogues, PAPhy_a and PAPhy_b, and have been mapped to Triticeae chromosomes 5 and 3, respectively. The promoters share a conserved core but the PAPhy_a promoter have acquired a novel cis-acting regulatory element for expression during grain filling while the PAPhy_b promoter has maintained the archaic function and drives expression during germination. Brachypodium is the only sequenced Poaceae sharing the PAPhy duplication. As for the Triticeae, the duplication is reflected in a high MGPA of ~4200 FTU/kg in Brachypodium. The sequence conservation of the paralogous loci on Brachypodium chromosomes 1 and 2 does not extend beyond the PAPhy gene. The results indicate that a single-gene segmental duplication may have enabled the evolution of high MGPA by creating functional redundancy of the parent PAPhy gene. This implies that similar MGPA levels may be out of reach in breeding programs for some Poaceae, e.g. maize and rice, whereas Triticeae breeders should focus on PAPhy_a.

Key words: Barley, Brachypodium, einkorn, gene duplication, PAPhy, phytase, purple acid phosphatase, Triticeae, wheat.

Introduction

Phytases (myo-inositol hexakisphosphate phosphohydrolase; EC 3.1.3.8 and EC 3.1.3.26) are defined as phosphatases which can initiate the sequential hydrolysis of the major storage form of phosphorus in plant seeds, phytate (InsP_6, myo-inositol 1,2,3,4,5,6-hexakisphosphate). This provides bio-available phosphate, inositol phosphates, and eventually inositol required for a range of cellular activities (Brinch-Pedersen et al., 2002). Expression or activation of phytase and the concomitant hydrolysis of InsP_6 has been observed during germination in many plant species (Duff et al., 1994). The digestive tracts of non-ruminants possess negligible phytase activity and the activity provided by food and feedstuffs is often insufficient for efficient hydrolysis of the InsP_6 content. Unable to be absorbed, InsP_6 is excreted along with chelated cations. InsP_6 is therefore considered the single, most important antinutritional compound for the bioavailability of a range of essential minerals including Zn^{2+}, Ca^{2+}, and Fe^{2+} (Brinch-Pedersen et al., 2007).
Endogenous plant phytases can only be beneficial to nutrition if they are present at the stage of consumption. This usually means in the mature seeds. Among cereals, the Triticeae tribe grains exhibit a remarkable high mature grain phytase activity (MGPA) (Eeckhout and De Paepe, 1994). Triticeae phytases are therefore known to be beneficial during bread making and in the digestive tract of pigs and humans (Pointillart et al., 1987; Sandberg and Andersson, 1988; Türk et al., 1996). Indexing MGPA against wheat (Triticum aestivum L.; 100%) places barley (Hordeum vulgare L.) at 49%, rye (Secale cereale L.) at 43% and triticate (\( \times \) Triticosecale, Wittmack) at 141%. In comparison, the non-Triticeae cereals oat (Avena sativa L.), maize (Zea mays L.) and rice (Oryza sativa L.) index at 3.5, 5.7, and 6.0%, respectively (Eeckhout and De Paepe, 1994; Holme et al., 2001). Moreover, there is also variation within species, in wheat up to 2-fold differences between cultivars (Ram et al., 2010).

The cereal phytase complement has been found to consist of the multiple inositol polyphosphate phosphatase (MINPP) phytases and the purple acid phosphatase phytases (PAPhy) (Dionisio et al., 2007, 2011). For PAPhy, 12 wheat, barley, maize, and rice cDNAs expressed in the grain have been described. In a phylogenetic analysis including 42 plant purple acid phosphatases, the isolated PAPhy genes grouped in a single clade (PAP type 1). The wheat and barley PAPhy could further be grouped as PAPhy_a isoforms, predominantly expressed during grain filling and as PAPhy_b isoforms, predominantly expressed during germination. The isoforms can also be distinguished by the C terminal but are otherwise of very similar sequence. Biochemical characterization of the recombinant proteins demonstrated similar enzyme kinetic parameters except for a slightly higher pH optimum for the examined PAPhy_a and a different preference for metal cofactors. The PAPhy_a protein accumulates in the protein storage vacuoles of the aleurone cells in close proximity to the major InPs deposits (Dionisio et al., 2011).

The MINPPs are also present in different isoforms and they are also expressed both during grain development and germination (Dionisio et al., 2007). However, purification of phytases from wheat bran and dry and germinating barley grains, in hindsight and based on molecular mass and N-terminal amino acid sequence, revealed the PAPhy as the major contributor to the MGPA in wheat and barley (Nakano et al., 1999; Greiner et al., 2000). This was further substantiated when it was shown that a transformation mediated duplication of the HvPAPhy_a gene in barley doubled the MGPA (Holme et al., 2012).

The present study addresses the characterization of genes encoding the Triticeae PAPhy and uncovers their evolutionary history towards a high MGPA. Fifteen Triticeae PAPhy genes from the five species wheat, barley, rye, einkorn, and Aegilops tauschii are cloned. It is demonstrated by chromosomal mapping in wheat and barley that the PAPhy genes can be assigned to two paralogous loci encoding the PAPhy_a and PAPhy_b isoforms. In rye, the PAPhy_a locus has undergone an additional duplication. The promoters responsible for the differentiated expression of the two loci were examined and the evolutionary origin of the paralogues is discussed in the context of the sequenced grasses rice, maize, sorghum, and Brachypodium (Brachypodium distachyon (L.) P. Beauv).

### Materials and methods

#### Plant materials used for cloning

Cultivars/accessions for two rye, five Ae. tauschii, three wheat, barley, einkorn, and Brachypodium were used for PAPhy cloning and/or phytase assays (Supplementary Table S1, available at JXB online). Plants were grown in the greenhouse according to Brinch-Pedersen et al. (2000).

#### In silico sequence analysis

Sequence information was retrieved from the databases of Gramene (www.gramene.org/) and NCBI (www.ncbi.nlm.nih.gov/) as well as the raw 454 wheat sequence reads from CerealsDB (www.cerealsdb.uk.net/Index_Home.html). Alignments and trees were carried out using the CLC-workbench (CLC bio). Promoters were analysed for cis-acting regulatory elements with the PlantCARE webtool (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Graphic representations of sequences were drawn with Vector NTI (Invitrogen). Synteny was analysed with the Narcisse webtool (www1.clermont.inra.fr/umr1095/narcisse_cereals/cgi-bin/narcisse15.cgi). Dotblots were drawn with the EMBOSS Polymod webtool (http://emboss.bioinformatics.nl/cgi-bin/emboss/polymod).

#### Genomic library constructions, screening, and gene cloning

Genomic libraries of wheat cultivar Skagen and rye cultivar Picasso were generated using the Lambda Fix II/XhoI Partial Fill-In Vector Kit (Agilent Technologies) according to the manufacturer’s instructions. The initial libraries were estimated to have 2.8–7.2 and 4.5–11.5-times genome coverage, respectively. The libraries were amplified once on NZY agar at a density of 300 pfu cm\(^{-2}\). Screening probes were labeled with 20 μCi \(^{32}\)P using PCR. Probe one was amplified using the forward primer 5'-CTTGAGGCTCGG GACGAAGTG-3' and the reverse primer 5'-GAGAGGGACCCC GCCCTCC-3'. The primers amplified a 479-bp fragment of the HvPAPhy_b cDNA. Probe two was amplified using the forward primer 5'-TGGACCTGCCGGACAC-3' and the reverse primer 5'-CGAAAGCGCCGATGAGC-3', amplifying a 770-bp fragment of the SrPAPhy_b1 gene.

For screening, the amplified libraries were plated out on 240 x 240 mm NZY agar at a density of 600 pfu cm\(^{-2}\). Plaque lifts were preformed with Hybond N+ membranes (GE Healthcare) followed by alkaline denaturation and UV crosslinking. Prehybridization (2 hours) and hybridization (overnight) were performed in Church’s buffer (0.25 M sodium phosphate buffer, pH 7.2, with 7% SDS and 0.17 mg ml\(^{-1}\) salmon sperm DNA) at 65 °C. The membranes were washed twice at 65 °C and once at room temperature with SSPE buffer. X-ray films were exposed with the membranes at −80 °C for 2–5 days and the film was subsequently developed and analysed. λ preparations were performed with the Lambda midi kit (Qiagen), according to the manufacturer’s instructions.

Sequencing of the λ clones facilitated the design of primers in conserved regions 3–400 bp upstream of the start codon and 50–250 bp downstream of the stop codon of the PAPhy genes (Supplementary Table S2). PCR was performed using the Herculase II polymerase in a reaction mixture with 6% DMSO (Agilent Technologies) according to the manufacturer’s instructions. Amplicons were cloned in the pCR4Blunt-TOPO vector (Invitrogen) and sequenced. At least two clones from independent PCR reactions were sequenced.

Inverted PCR was used to extend known sequences in the 5’-direction. The templates were prepared by digestion of cereal genomic DNA with restriction endonucleases, BamHI, XhoI, or ScaI followed by inactivation of the enzyme and ligation with T4 DNA ligase in a final
DNA concentration of 2 ng μl⁻¹. Templates were generated from barley and einkorn genomic DNA. Inverted PCR consisted of a reaction with a primary primer set (5'-GAATTCCAGATGGCGGC-3' and 5'-CTGCCTGGAGGCCACAG-3') followed by a second nested PCR with a secondary primer set (5'-TACAGGCAGCTACCCCT-3' and 5'-ACCGGTGCGCCCTGTG-3'). PCR, cloning, and sequencing were preformed as described above.

Chromosomal mapping

Wheat chromosomal mapping was performed using the Chinese Spring nullisomic-tetrasicomic lines described by Kimber and Sears (1979). Out of original 42 possible nullisomic-tetrasicomic lines, two were missing in the present set of lines, the nullisomic (N) 2A tetrasomic (T) 2B and the N4BT4D lines. Their absence did not compromise the mapping because there are two nullisomic lines for each chromosome.

For chromosomal mapping in barley, the addition lines consisting of individual barley chromosomes (from cv. Betzes) translocated to wheat (cv. Chinese spring) was used (Islam et al., 1981; Islam and Shepherd, 2000). A part of each PAPhy isogene was PCR amplified using a specific set of primers (Supplementary Table S3) and Taq polymerase.

GUS constructs

The construct pCLEAN-G185-PAPhy_a (Holme et al., 2012) was modified to serve as a GUS construct. The PAPhy_a coding open reading frame and terminator was replaced by the UidA open reading frame followed by the NOS terminator. The cloning was performed with In-Fusion technology (Zhu et al., 2007), ensuring that the start codon context was protected.

The vector backbone and promoter of the construct was amplified using the primers Cis to GUS Fw (5'-TGGGATGCGTACAACTG-3') and Cis to GUS Rv (5'-GTGGATGTGGTTGTTTGGCATTG-3'). The UidA and NOS terminator was amplified from a donor construct using the primers Fw with overhang (5'-AGCAAAACATCAACATTTACGGCTTCTCATGAAACAAAC-3') and GUS Rv with overhang (5'-GGAACGTCGACTCGACTATGACCCAGTTACGAAATTTCC-3'). The resulting construct was named pCLEAN-G185-wt-proGUS.

A potential regulatory motif was identified in silico (see Results). In order to evaluate the motif in vivo, a GUS construct where the motif had been obstructed was created. This was done by taking 20 bp including the 14-bp motif from the pCLEAN-G185-PAPhy_a construct and process them trough a nucleotide randomizer (http://molbioi.ru/eng/scripts/01_16.html) with settings to preserve the nucleotide ratios of the original sequence. The resulting sequence, 5'-gctagcagacttagcag-3', was only identical to the original in three nucleotide positions and did not contain any regulatory elements known by PlantCARE. The original 20bp in pCLEAN-G185-wt-proGUS was replaced by the randomized sequence as described by Zhu and co workers (2007). The primers were Kill triad Fw (5'-gctagcagacttagcagCTAGCGGTCAAACTTTG-3') and Kill triad Rv (5'-ctagcagacttagcagCTAGCAAAATGTTAGCTTGAATTAAGAGG-3'). The resulting construct was pCLEAN-G185-KOTriad-proGUS.

Transient expression

Immature barley (cv. Golden Promise) endosperms, 15–20 days after pollination, were isolated and placed on 0.25% phytagel pH 5.8 containing (1') 4.4 g MS medium (Duchefa) supplemented with 10.5 g myo-inositol, 1.5 g ammonium nitrate, 1 g casein hydrolysate, 0.7 g l-proline, 1.5 mg 2,4-dichlorophenoxyacetic acid, 30 g maltose, and 72.88 g mannitol. The endosperms were bombarded in a DuPont PDS 1000 Helium Biolistic Delivery system using 1100-psi rupture discs and 1 μm gold particles coated with plasmid DNA as described by the manufacturer (BioRad). The endosperms were incubated in the dark at 23 °C overnight before histochemical staining as described by (Jefferson et al., 1987).

Phytase assay

Mature cereal grains were assayed for phytase activity essentially as Engelen et al. (1994). Grains were ground in a mortar. Protein was extracted from flour using 10 ml of 25 mM sodium acetate buffer (pH 5.5) with 0.1 M calcium chloride per gram and shaking for 60 minutes. After centrifugation, 30 μl was diluted to 800 μl with extraction buffer containing sodium phytate to a final concentration of 5 μM. Phytate hydrolysis proceeded for 60 minutes at 37 °C. The reaction was stopped with 800 μl of freshly prepared 25 g l⁻¹ ammonium heptamolybdate with 0.25% ammonia/6.9 g l⁻¹ ammonium vanadate with 1% nitric acid/65% nitric acid (15:15:10). The mixture was centrifuged for 5 minutes at 3800 g before absorbance was measured at 415 nm. Blanks were made by substituting the protein extract with extraction buffer. The concentration of phosphate was calculated according to a standard curve. All samples were measured in technical duplicates.

Results

Cloning of the PAPhy gene complement of selected Triticeae

Screening of two λ phage libraries lead to the identification of four wheat and two rye unique PAPhy clones. The clones were partially sequenced in order to localize the coding sequence as well as the promoter, introns, and terminator sequences. Wheat TaPAPhy_a2 and rye ScPAPhy_al λ clones contained only part of the coding sequences. In addition, a barley PAPhy λ clone isolated previously was analysed in more detail in the present study (Holme et al., 2012).

Introns and exons were annotated after alignment with the known cDNA clones and identification of the in-frame start and stop codons (Fig. 1). One interruption of the coding sequence on the TaPAPhy_a2 λ clone was only 8 bp and did not follow the GT–AG rule for introns. PCR screening of nine wheat cultivars (including Chinese spring) showed that the 8-bp interruption was only present in the cultivars Skagen and Bob White (results not shown). Hence, this interruption was annotated as an insertion rather than an intron. All other interruptions of the coding sequences followed the GT–AG rule as expected for typical introns.

A PCR-based cloning strategy was designed as supplement to the library screenings. Sequence information from the λ clones and known cDNA sequences was used to design primers in conserved regions 3–400 bp upstream of the start codon and 50–250 bp downstream of the stop codon. In wheat, all genes already known from the library screenings were isolated again (Fig. 1). However, two additional genes, termed TaPAPhy_b1 and TaPAPhy_b2, were identified. In rye, the full coding sequence of ScPAPhy_al was obtained and an additional gene, ScPAPhy_a2, was discovered. From barley, HvPAPhy_b1-2 was identified. One PAPhy_a gene and one PAPhy_b gene were amplified from both einkorn and Ae. tauschii. The PAPhy 5'-flanking regions of the PCR clones were extended by inverted PCR and by consulting raw 454 reads of wheat available at CerealsDB. Using these techniques, the 5'-flanking regions were extended from 277 bp in...
**TaPAPhy_a2** to a maximum of 713 bp in **HvPAPhy_b1-2**. In summary, three **PAPhy_a** and three **PAPhy_b** genes could be isolated from wheat. One **PAPhy_a** and one **PAPhy_b** gene were cloned from each of barley, einkorn, and *Ae. tauschii*. In rye, two **PAPhy_a** genes and one **PAPhy_b** gene were isolated. The clones are described in Table 1 with the details of their accession numbers, size, and gene coverage.

**Gene structure, copy number, and chromosomal mapping**

The cloned **PAPhy** genes contain either four or five introns (Fig. 1). Pairwise alignments of the genomic wheat and barley clones and the corresponding cDNA clones previously cloned and classified as **PAPhy_a** and **PAPhy_b** by Dionisio et al. (2011) revealed that the **PAPhy_a** genes possessed four introns and the **PAPhy_b** genes possess five introns. The four introns of **PAPhy_a** were positioned as introns 1, 2, 4, and 5 of **PAPhy_b**. Barley is known to express one **PAPhy_a** transcript and two **PAPhy_b** transcripts (Dionisio et al., 2011). The known **PAPhy_a** transcript corresponds to the **λ** clone of **HvPAPhy_al**, but only one **PAPhy_b** gene was successfully cloned by PCR. However, this gene could account for both transcripts by differential splicing at the 3′-border of the third intron, where two alternative AG dinucleotides facilitate the inclusion or omission of a GAG codon. Because of the differential splicing of the barley **PAPhy_b** transcripts, this gene is referred to as **HvPAPhy_b1-2** and the two transcripts as **HvPAPhy_b1** and **HvPAPhy_b2**. The three **TaPAPhy_a** and **TaPAPhy_b** genes isolated from Chinese Spring suggest that the two groups are a pair of paralogues with homoeologous loci on each of the three subgenomes A, B, and D. Moreover, only one **PAPhy_a** gene and one **PAPhy_b** gene could be isolated from the diploid carriers of the A and D genomes (einkorn and *Ae. tauschii*, respectively) and from barley (diploid).

The wheat and barley **PAPhy_a** and **PAPhy_b** genes were chromosomally mapped to confirm or disprove the assumption of a single pair of paralogues (Supplementary Figs. S1 and S2 and Supplementary Table S2). The three **TaPAPhy_a** genes were mapped with one gene on each of the three subgenomes A, B, and D. The **PAPhy_a** genes were located on chromosome 5 in all wheat subgenomes and barley. The wheat **TaPAPhy_b** genes were distributed in a similar manner, with one locus on each of the three homoeologous subgenomes. In wheat and barley, the genes mapped to chromosome 3 (Table 1).

**The PAPhy genes of sequenced grasses**

The **TaPAPhy_al** coding sequence was blasted against the sequenced genomes of rice, maize, sorghum, and
Table 1. Summary of isolated clones, the chromosomal location of their corresponding genes and the extent to which the coding sequences and flanking regions have been sequenced

| Species         | Clone type | Accession | Gene       | Chromosome | 5’- end of start codon (bp) | Exons and introns (bp) | 3’-end of stop codon (bp) | Extension                        |
|-----------------|------------|-----------|------------|------------|-----------------------------|------------------------|-----------------------------|---------------------------------|
| Wheat           | λ, PCR     | JF838306  | TaPAPhy_a1 | 5D         | 2090                         | 2003                   | 721                         |                                  |
|                 |            | JF838305  |            |            | 309                          | 2003                   | 76                          |                                  |
|                 | λ, PCR     | JF838308  | TaPAPhy_a2 | 5A         | 2373                         | 1119                   | –                           |                                  |
|                 |            | JF838307  |            |            | 324                          | 1975                   | 76                          |                                  |
|                 | λ, PCR     | JF838310  | TaPAPhy_a3 | 5B         | 3837                         | 1982                   | 703                         |                                  |
|                 |            | JF838309  |            |            | 261                          | 1982                   | 94                          |                                  |
|                 | PCR        | JF838311  | TaPAPhy_b1 | 3D         | 391 (697)                    | 2081                   | 192                         | 306 bp by 454 read lcl|GKTG12C01AVAXX                  |
|                 |            | JF838312  | TaPAPhy_b2 | 3B         | 378 (655)                    | 2072                   | 203                         | 277 bp by 454 read lcl|F3OD4V4H01C3ZB5                  |
|                 | λ, PCR     | JF838314  | TaPAPhy_b3 | 3A         | 1883                         | 2088                   | 209                         |                                  |
|                 |            | JF838313  |            |            | 361                          | 2087                   | 196                         |                                  |
| Einkorn         | PCR        | JF838315  | TmPAPhy_a1 |            | 316 (627)                    | 1972                   | 76                          | 311 bp by inverted PCR         |
| Aegilops tauschi| PCR        | JF838316  | TmPAPhy_b1 |            | 364                          | 2058                   | 197                         |                                  |
|                 |            | JF838317  | AePAPhy_a1 |            | 307                          | 2008                   | 76                          |                                  |
|                 | PCR        | JF838318  | AePAPhy_b1 |            | 391                          | 2085                   | 193                         |                                  |
| Barley          | λ, PCR     | FR851293  | HvPAPhy_a1 | 5H         | 2728                         | 2246                   | 878                         |                                  |
|                 |            | JF838323  | HvPAPhy_b1 | 3H         | 319 (1032)                   | 2066                   | 202                         | 713 bp by inverted PCR         |
|                 | PCR        | JF838320  | ScPAPhy_a1 |            | 1493                         | 149                    | –                           |                                  |
| Rye             | λ, PCR     | JF838319  | ScPAPhy_a2 |            | 504                          | 1968                   | 102                         |                                  |
|                 |            | JF838322  | ScPAPhy_b1 |            | 2723                         | 2016                   | 698                         |                                  |

Brachypodium in the LATESTGP database (Gramene.org). The search sensitivity was set at ‘near exact matches’. Rice, maize, and sorghum were found to possess one PAPhy locus, whereas Brachypodium possesses two loci. Contig views showing the TaPAPhy_a1 coding sequence aligned to the chromosome and predicted genes are shown in figure 2. According to the Narcisse webtool, conserved synteny exists between Brachypodium chromosome 1, barley chromosome 5, maize chromosome 1, rice chromosome 3, and sorghum chromosome 2. This suggests that the PAPhy genes on these chromosomes are orthologues. There was no conserved synteny between Brachypodium chromosomes 1 and 2. Likewise, there was no conserved synteny between barley chromosomes 3 and 5 (results not shown).

All-against-all dotblots were generated using 20 and 200 Kb chromosome sequences centred around the start codon of the PAPhy genes in the sequenced grasses (Fig. 3). The best results were obtained with word size 11 for the 20-kb blots and word size 14 for the 200-kb blots. The sequences were orientated according to the PAPhy genes. The genes were clearly recognizable as a diagonal line from the centre and upwards to the right in all 20-kb blots. The synteny extended beyond the PAPhy gene in 20-kb blots of Brachypodium chromosome 1 vs. rice and in rice vs. sorghum. In 200-kb blots, the PAPhy genes were only just visible as an enlarged dot in the centre. Nearby regions with more extensive identity were visible as diagonals or tracks of enlarged dots in the blots of Brachypodium chromosome 1 vs. rice and sorghum, sorghum vs. rice as well as sorghum vs. maize.

Taken together, this confirms the hypothesis that the Brachypodium chromosome 1 gene and the single-copy PAPhy genes of rice, maize, and sorghum are orthologues. Dotblot confirmation regarding inclusion of the gene on Triticaceae chromosome 5 in this group is pending because current sequence data is insufficient.

The Brachypodium chromosome 2 segments did not form clear diagonals or tracks beyond the PAPhy gene with any other sequence, although enlarged dots were present. A multiple alignment of the two Brachypodium 20-kb sequences and the predicted coding sequences of the two PAPhy genes was generated. The genomic sequences aligned from the start codon until 43/52 bp before the stop codons and they were 53.5% identical in this region. Three smaller regions of identity (72 bp/97.22%, 51 bp/86.27%, and 30 bp/83.33%) were identified downstream of the PAPhy genes. The first region was found only 100 bp downstream of the PAPhy gene on chromosome 1 but 5051 bp downstream of the gene on chromosome 2. All introns of the Brachypodium chromosome 1 gene were also present in the gene on chromosome 2. This gene on the other hand had two additional introns, numbers 2 and 6.

The TaPAPhy_a1 and TaPAPhy_b1 genomic and coding sequences were added to the alignment. The two introns specific to the gene on Brachypodium chromosome 2 were found in both wheat genes and all wheat introns were also found on the Brachypodium chromosome 2 gene. The Brachypodium
Fig. 2. Contiq views of the PAPhy loci in sequenced grasses. Boxes, exons; lines, introns; bright red, significant alignments of the TaPAPhy_a1 coding sequence to the genomes; dark red, predicted genes.

Fig. 3. All-against-all dotblots of 20 and 200 kb sequences centred around the PAPhy start codon.
coding sequences were aligned to each other and were found to be 78.1% identical.

The Triticeae PAPhy core promoter and untranslated regions

The longest available sequences representing the individual Triticeae PAPhy genes were aligned and adjusted to include the beginning of the coding sequence and the 5′-flanking region (Supplementary Fig. S3). One in-frame ATG codon was conserved and aligned in the beginning of the predicted coding sequences of all 15 PAPhy genes. The four bases preceding the ATG agrees with the consensus start motif found for monocots G(A/C)(G/A)GC(A/C/G)(G/A)(C/A)(G/C) AUGGCCG (Nakagawa et al., 2008). This codon was therefore assigned as the start codon.

All 15 PAPhy genes were represented in the alignment up to position –601 bp (relative to the start codon). CAP-trapper-generated cDNA clones of TaPAPhy_a2 and TaPAPhy_b2 are available (Genbank AK331144.1 and AK331704.1) (Kawaura et al., 2009). The CAP trapmer method ensures full transcript length clones (with an additional 5′-guanidine) so the clones facilitated the identification of conserved motives with positions and sequences corresponding to the TATA box and the near upstream element. Most genomic clones contained two adjacent TATA boxes and one contained two near upstream elements (Fig. 1 and Supplementary Fig. S3).

Transcription is expected to begin 25–39 bp downstream of the TATA box (Joshi, 1987). Genomic vs. cDNA pairwise alignment demonstrated that the full transcript length TaPAPhy_a2 cDNA clone starts 35 bp downstream of the TATA box closest to the start codon (TATA box 1). The context of this site is conserved and it fits well with the ‘cap signal’ (T/G/C(T/G/I/C)(T/C/A)(C/G)(I/C)/C) reported by (Bucher, 1990). The adjacent alignment position is a better match to the cap signal for the PAPhy_b genes. The two positions were assigned as the transcriptional start sites when TATA box 1 is used (Supplementary Fig. S3).

The TaPAPhy_b2 clone on the other hand have been transcribed via the second TATA box (TATA box 2) since it begins 38 bp downstream at alignment position –206. This position is only covered by the Triticum and Aegeilops PAPhy_b genes, which appear to have an insertion here when they are compared to the PAPhy_a genes and to the barley and rye PAPhy_b sequences. The transcriptional start site of genes without the insertion is expected to be closer to the start codon because of the strict positional constraints of the TATA box and the transcriptional start site. A good candidate is alignment position –186. This residue is conserved in all PAPhy genes and its consensus context CCAAGCGA (position –186 underlined) fits well with the cap signal.

TATA box 1 aligns across all PAPhy genes at positions –115 to –121 bp and it is surrounded by a 121-bp block of high conservation, even between PAPhy_a and PAPhy_b (Supplementary Fig. S3). The position of TATA box 2 was also conserved, although it is missing from TaPAPhy_b1 and AePAPhy_b1. Moreover, as seen for TATA box 1, box 2 is surrounded by a segment of significant conservation across all PAPhy genes. However, for TATA box 2, the conserved region is much shorter (58 bp) and less conserved (Supplementary Fig. S3). The distance between the TATA boxes is 90 +/− 28 bp in all promoters except ScPAPhy_a2.

In conclusion, the Triticeae PAPhy genes have a conserved core promoter usually consisting of two TATA boxes approximately 90 bp apart and adjacent transcriptional start sites. The PAPhy_b 5′-UTR is 60–70 bp longer than the PAPhy_a 5′-UTR when transcripts are derived from the same TATA box. Pairwise alignment of the Brachypodium promoters (1000 bp 5′-flanking regions) showed that they do not possess a conserved core promoter (results not shown).

The differentiated Triticeae PAPhy_a and PAPhy_b promoters

There was no significant conservation between paralogous promoters upstream of the core promoter. However, orthologous promoters were still highly conserved until alignment position –650 and –910 for PAPhy_a and PAPhy_b, respectively (Supplementary Fig. S3). Blasting of the 5′- and 3′-flanking regions against the NCBI nucleotide collection revealed no known genes. However, λ clones of TaPAPhy_a2, TaPAPhy_a3, TaPAPhy_b3 and ScPAPhy_a1 contained regions which returned several matches typically annotated as transposons (Table 2). Moreover, comparing the ScPAPhy_a2 to other PAPhy_a genes reveals a 219-bp insertion between the two TATA boxes (Fig. 1 and Supplementary Fig. S3). The size and presence of flanking short direct repeats suggest the insertion is a short interspersed element.

The breakdown of intra-orthologue conservation coincides with the occurrence of transposon-like sequences (Table 2 and Supplementary Fig. S3). In the further investigation of the PAPhy promoters, it was assumed that cis-acting regulatory elements responsible for the differentiated expression pattern of PAPhy_a and PAPhy_b are found upstream of the core promoter and downstream of the intra-orthologue unconserved region. These assumptions defined osthologous-specific promoter sequences of 103–157 bp and 121–259 bp for PAPhy_a and PAPhy_b, respectively (Supplementary Fig. S3, red and blue boxes). These sequences were screened for cis-acting regulatory elements.

Five motifs were common in the PAPhy_a promoters (Table 3). The RY element was the only omnipresent element, but all the elements were present in at least half of the promoters and many cases of absence were only single-nucleotide sequences that could reflect variations of the element. The observations were drawn to scale using the TaPAPhy_a2 5′-flanking region as template (Fig. 4A).

The overlapping motifs of GCN4 and Skn-1 and the preceding base are identical to the odd base C-box ATGAGTCAT reported by De Pater et al. (1994). Hence, the two elements and the RY element can be interpreted as two elements overlapping in three nucleotides and spanning a total of 14 bp. The significance of the combined odd base C-box and RY element was examined by transient expression of GUS driven by the barley PAPhy_a promoter in its native form (construct pCLEAN-G185-wt-proGUS) and with the two elements
randomized (construct pCLEAN-G185-KO-proGUS). The pCLEAN-G185-wt-proGUS construct yielded several blue dots (Fig. 5) whereas the pCLEAN-G185-KO-proGUS did not yield any blue dots (results not shown). The two Brachypodium promoters (1000 bp 5′-flanking regions) were also screened for the double element but it was not present.

Three cis-acting regulatory elements were common in the PAPhy promoters (Table 3). The observations were drawn to scale using the TaPAPhy_b3 5′-flanking region as template (Fig. 4B).

**Table 2.** Matches to the 5′-flanking regions of the λ clones

| λ clone            | Match position (relative to the start codon) | Maximum identity (%) | Annotation at match                        | Match accession   | Origin of match |
|--------------------|---------------------------------------------|----------------------|--------------------------------------------|--------------------|-----------------|
| TaPAPhy_a2         | –507 to –2360                                | 86                   | None, long terminal repeat retrotransposons at similar matches | AY494981.1         | T. turgidum     |
| TaPAPhy_a3         | –800 to –3837                                | 95                   | Long terminal repeat                       | FN564428.1         | Wheat (3B)      |
| TaPAPhy_b3         | –714 to –1883                                | 96                   | None, flanked by repeat regions            | FN564434           | Wheat (3B)      |
| ScPAPhy_a1         | –698 to –1447                                | 82                   | Retrotransposon, gypsy                     | FN564428           | Wheat (3B)      |

**Table 3.** cis-Acting regulatory elements in PAPhy promoters

The occurrence is given as the number of orthologous promoters with the element compared with the number of sequences represented at the particular position in the alignment.

| Motif          | Sequence | Description in PlantCARE                                                                 | Origin according to PlantCARE | Occurrence |
|----------------|----------|----------------------------------------------------------------------------------------|-------------------------------|------------|
| PAPhy_a        | TGATC    | cis-Acting regulatory element involved in the MeJA-responsiveness                      | Barley                        | 4 of 6     |
| GNC4           | TGAGTCA  | cis-Regulatory element involved in endosperm expression                                | Rice                          | 5 of 8     |
| Skn1           | GTCAT    | cis-Acting regulatory element required for endosperm expression                        | Rice                          | 7 of 8     |
| RY-element     | CATCGAT  | cis-Acting regulatory element involved in seed-specific regulation                     | Sunflower                     | 8 of 8     |
| PAPhy_b        | ABRE     | cis-Acting element involved in the abscisic acid responsiveness                       | Arabidopsis thaliana          | 5 of 7     |
| GARE           | AACGTA   | Gibberellin-responsive element                                                         | Brassica oleracea             | 5 of 5     |
| TGACG          | TGACG    | cis-Acting regulatory element involved in the MeJA-responsiveness                      | Barley                        | 6 of 7     |

A multiple alignment of the genomic sequences of all 15 Triticeae PAPhy genes was created in CLC workbench using the settings gap open cost 10, gap extension cost 1 and end gap cost ‘as any other’. Sequences from the PCR clones were used when available. Alternatively, sequences from λ clones were shortened to lengths similar to the PCR clones and used for the alignment. The initial alignment was adjusted to include only the section between the first T of TATA box 1 and the T in the near upstream element.

**Phylogeny**

Fig. 4. Schematic representation of the regulatory elements in the (A) PAPhy_a and (B) PAPhy_b –300 and –376 bp 5′-flanking regions.
A UGPMA tree was constructed from this alignment (Fig. 6A). The PAPhy_a and PAPhy_b genes form two distinct groups across the species. The PAPhy genes from einkorn (A genome) and Ae. tauschii (D genome) group most closely with the wheat genes located on the A and D subgenomes, respectively. In both the PAPhy_a and PAPhy_b groups, the einkorn and wheat A subgenome PAPhy genes are more diverged than Ae. tauschii and wheat D subgenome PAPhy genes. The divergence between the two PAPhy_a genes in rye is comparable to the divergence between einkorn and the wheat PAPhy genes of the A subgenome. Rye and barley genes fall outside a close Triticum/Aegilops grouping for both the PAPhy_a and PAPhy_b groups.

Arguably, the tree is heavily affected by the additional intron present in the PAPhy_b genes. In order to investigate whether the tree would be different without the introns, a multiple alignment of the coding sequences was generated (Fig. 6B). The same settings were used but additional coding sequences of the PAPhy genes from sequenced grasses were included in order to put the PAPhy_a and PAPhy_b groups into a broader context. When using this approach, the rye PAPhy_a genes group with their Triticum counterparts. However, apart from this, the Triticeae PAPhy genes group as before. ScPAPhy_a1, ScPAPhy_a2, HvPAPhy_a1, and TaPAPhy_a3 coding sequences are 27 bp shorter than other PAPhy_a genes due to a polymorphism introducing a stop codon. When the coding sequences are shortened to equal length, they group almost as seen for the genomic clones (Fig. 6C). Moreover, the Brachypodium PAPhy located on chromosome 2 groups with the PAPhy_b genes. The Brachypodium PAPhy located on chromosome 1 is the most diverged of all genes included. The rice, sorghum, and maize PAPhy sequences fall outside the two groups formed by the Triticeae PAPhy genes and the maize and sorghum PAPhy genes group closely together.

Phytase activity in Triticeae and Brachypodium

The MGPA of the plant material used for cloning and Brachypodium accessions was measured. Additional cultivars/accessions of each species were measured to provide biological replication. The samples were greenhouse grown as described or primary seedbank material in the case of AC 1548, AE 1068, AE 938, GRA 787 and GRA 788. The values differed significantly between the materials. The highest MGPA (6019 FTU/kg) was measured in Ae. tauschii accession NGB90403 and the lowest (736 FTU/kg) was measured in wheat cultivar Skagen. The MGPA of Brachypodium was in the same range as the Triticeae.

Discussion

The PAPhy_a and PAPhy_b paralogues

Chromosome mapping demonstrated that the PAPhy genes of wheat and barley consists of a set of paralogous genes located on chromosomes 3 and 5 in each of the homoeologous genomes. This is in good agreement with a previous division of PAPhy proteins and transcripts into two isoforms, PAPhy_a and PAPhy_b (Dionisio et al., 2011). The gene on chromosome 5 generates the transcripts termed PAPhy_a and the gene on chromosome 3 generates the PAPhy_b transcripts. It is reasonable to assume a similar chromosomal localization in the diploid carriers of the A and D genomes, i.e. einkorn and Ae. tauschii, respectively.

Rye possesses as the only investigated member of the Triticeae a duplicated PAPhy_a gene. The exact nature of this duplication remains to be investigated but the phylogenetic trees indicate that the two ScPAPhy_a genes are as divergent as the PAPhy genes of wheat subgenome A and those of einkorn. This is notably more divergent than the PAPhy genes of wheat subgenome D and Ae. tauschii. The
wheat A subgenome and einkorn diverged between 0.5 and 1 million years ago whereas the hexaploidization which separated the wheat D subgenome from *Ae. tauschii* occurred only 8000 years ago (Huang et al., 2002). Divergence of the *Secale* and *Triticum* genera occurred around 7 million years ago and rye was domesticated 10,000–12,000 years ago (Salamini et al., 2002; Huang et al., 2009). This suggests that the rye *ScPAPhy_a* duplication occurred long after speciation but prior to domestication. It is also possible that two different wild *Secale* species each have contributed a locus to domesticated rye.

*Brachypodium* is the closest relative of the Triticeae which has been fully sequenced and assembled. The *Brachypodium* set of *PAPhy* paralogues is located on chromosomes 1 and 2. These chromosomes share orthologous linkage blocks with the Triticeae chromosomes 5 and 3, respectively (International Brachypodium Initiative, 2010). The presence of orthologous linkage blocks on chromosomes carrying the paralogous *PAPhy* genes in *Brachypodium* and in Triticeae suggests that they originate from the same gene duplication event. This is supported by the high degree of shared intron/exon structure and the close phylogenetic relationship between the *PAPhy_b* genes and the *Brachypodium* chromosome 2 gene. On the other hand, the gene on *Brachypodium* chromosome 1 is the most distant member of the clade. However, accelerated divergence can be caused by the functional redundancy introduced by the gene duplication itself (Zhang, 2003). This is supported by the dotblot analysis which show that the *Brachypodium* chromosome 1 *PAPhy* is the orthologue of the *PAPhy* genes in rice, maize, and sorghum. The dotblot and

**Fig. 6.** UGPMA trees of: (A) the genomic Triticeae *PAPhy* sequences from TATA box 1 to the first near upstream elements; (B) full-length coding sequences from *PAPhy* genes isolated in this study and sequenced grasses; (C) coding sequences from *PAPhy* genes isolated in this study and sequenced grasses trimmed to equal lengths. Bootstrap values are shown at branch nodes.
alignment generated from the two *Brachypodium* sequences demonstrated that the duplication is of the same approximate size as the gene itself. Small duplications may result from re-roposition but this would usually result in the loss of introns and certainly the loss of the core promoter (Zhang, 2003). The core promoter is preserved in the Triticeae and introns are preserved in both *Brachypodium* and Triticeae. The duplication must therefore be of the segmental type.

Other sequenced grasses (rice, maize, and sorghum) contain only one *PAPhy* gene. This raises the question of gene duplication vs. gene loss. All the Triticeae *PAPhy* genes and the *Brachypodium* chromosome 2 gene were phylogenetically closer to each other than to the *PAPhy* genes of maize, rice, and sorghum. The duplication is therefore younger than the divergence of the species. Further, it is older than the divergence of *Brachypodium* and the Triticeae since they share it. This indicates duplication between 54 and 32 million years ago (International *Brachypodium* Initiative, 2010). The most ancient gene must be that, which is shared with all the examined grasses, i.e. the *PAPhy_a*/*Brachypodium* chromosome 1 gene.

The divergent promoters of *PAPhy_a* and *PAPhy_b*

Analysis of the 5’-flanking region of eight Triticeae *PAPhy_a* and seven Triticeae *PAPhy_b* genes revealed a conserved core promoter with two TATA boxes. The two full-length transcripts demonstrate that both TATA boxes are active albeit in different genes. The high conservation of both TATA boxes and their associated transcriptional start sites suggest that both TATA boxes are generally active.

There was no significant conservation between *PAPhy_a* and *PAPhy_b* upstream of the core promoter, but orthologues showed conservation for an additional 150–200 bp. The upstream sequences of the paralogous promoters of *PAPhy_a* and *PAPhy_b* contained cis-acting regulatory elements consistent with the differentiated expression reported previously (Dionisio et al., 2011). Thus, *PAPhy_b* contained the gibberellic acid response element typical for hydrolases expressed during germination in response to gibberellic acid (Ritchie and Gilroy, 1998). For *PAPhy_a*, the promoters contain the overlapping motifs of the odd base C box and the RY-element. The same motif can also be described as a triad consisting of the GCN4 motif, the skn1 motif, and the RY-element. These motives have been described in promoters driving expression of storage protein genes during seed formation and grain filling. For instance, the RY element has been found and proven functional in promoters of dicot storage proteins such as legumins or conglycinin (Bäumlein et al., 1992; Fujiwara and Beachy, 1994). Similarly the odd base C box has been found in lectin promoters, hordeins and rice glutelins (Müller and Knudsen, 1993; De Pater et al., 1994; Takaiwa et al., 1996). This study’s reporter gene experiment shows that this double motif or triad is critical for the expression of *PAPhy_a* during grain filling in barley.

**Phytase activity in the wider Triticeae tribe and Brachypodium**

The present data agrees with previously reported values for wheat, barley, and rye (Eeckhout and De Paepe, 1994; Viveros et al., 2000; Steiner et al., 2007). A collection encompassing einkorn, *Ae. tauschii* and *Brachypodium* accessions was also assayed (Fig. 7). As far as is known, these are the first published MGPA values for the three species. The MGPA measurements of einkorn and *Ae. tauschii* provide further evidence of the association of high MGPA with the Triticeae tribe but the *Brachypodium* data demonstrate that the trait is not restricted to the Triticeae. Rather, it is associated with the duplication of the *PAPhy* gene which is shared by *Brachypodium* and the

---

**Fig. 7.** Phytase activities of mature seeds.
Chromosomal mapping of the Primers for amplifying CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. The expression pattern of the Brachypodium 

Bäumlein H, Nagy I, Villarroel R, Inzé D, Wobus U. 1992. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. The Plant Journal 2, 233–239.

Brinch-Pedersen H, Olesen A, Rasmussen SK, Holm PB. 2000. Generation of transgenic wheat (Triticum aestivum L.) for constitutive accumulation of an Aspergillus phytase. Molecular Breeding 6, 195–206.

Brinch-Pedersen H, Sorensen LD, Holm PB. 2002. Engineering crop plants: getting a handle on phosphate. Trends in Plant Science 7, 118–125.

Brinch-Pedersen H, Borg S, Tauris B, Holm PB. 2007. Molecular genetic approaches to increasing mineral availability and vitamin content of cereals. Journal of Cereal Science 46, 308–326.

Bucher P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of Molecular Biology 212, 563–578.

De Pater S, Katagiri F, Kijne J, Chua NH. 1994. bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. The Plant Journal 6, 133–140.

Dionisio G, Holm PB, Brinch-Pedersen H. 2007. Wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.) multiple inositol phospholipase C promoter elements derived from 502 unrelated promoter sequences. Journal of Molecular Biology 212, 563–578.

Duff SMG, Sarath G, Plaxton WC. 1994. The role of acid phosphatases in plant phosphorus metabolism. Physiologia Plantarum 90, 791–800.

Eckhoff W, De Paepe M. 1994. Total phosphorus, phytate-phosphorus and phytase activity in plant feedstuffs. Animal Feed Science and Technology 47, 19–29.

Engelen A, van der Heeft F, Randsdorp P, Smit EL. 1994. Simple and rapid determination of phytase activity. Journal of AOAC International 77, 760–764.

Fujiwara T, Beachy RN. 1994. Tissue-specific and temporal regulation of a β-conglycinin gene: roles of the RY repeat and other cis-acting elements. Plant Molecular Biology 24, 261–272.

Greiner R, Jany K, Larsson Alminger M. 2000. Identification and properties of myo-inositol hexakisphosphate phosphohydrolases (phytases) from barley (Hordeum vulgare). Journal of Cereal Science 31, 127–139.

Holme IB, Dionisio G, Brinch-Pedersen H, Wendent T, Madsen CK, Vincze E, Holm PB. 2012. Cisgenic barley with improved phytase activity. Plant Biotechnology Journal 10, 237–247.

Huang H, Shi P, Wang Y, Luo H, Shao N, Wang G, Yang P, Yao B. 2009. Diversity of beta-propeller phytase genes in the intestinal contents of grass carp provides insight into the release of major phosphorus from phytate in nature. Applied and Environmental Microbiology 75, 1508–1516.

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex
and the evolutionary history of polyploid wheat. Proceedings of the National Academy of Sciences, USA 99, 8133–8138.

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463, 763–768.

Islam AKMR, Shepherd KW. 2000. Isolation of a fertile wheat-barley addition line carrying the entire barley chromosome 1H. Euphytica 111, 145–149.

Islam AKMR, Shepherd KW, Sparrow DHB. 1981. Isolation and characterization of euplasmic wheat-barley chromosome addition lines. Heredity 46, 161–174.

Jefferson RA, Kavanagh TA, Bevan MW. 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO Journal 6, 3901–3907.

Joshi CP. 1987. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Research 15, 6643–6653.

Kawaura K, Mochida K, Enju A, et al. 2009. Assessment of adaptive evolution between wheat and rice as deduced from full-length common wheat cDNA sequence data and expression patterns. BMC Genomics 10, 271.

Kimber G, Sears E. 1979. Uses of wheat aneuploids. Basic Life Science 13, 427–443.

Lucca P, Hurrell R, Potrykus I. 2001. Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. Theoretical and Applied Genetics 102, 392–397.

Müller M, Knudsen S. 1993. The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. The Plant Journal 4, 343–355.

Nakagawa S, Niimura Y, Gojobori T, Tanaka H, Miura Ki. 2008. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. Nucleic Acids Research 36, 861–871.

Nakano T, Joh T, Tokumoto E, Hayakawa T. 1999. Purification and characterization of phytase from bran of Triticum aestivum L. cv. Nourin #61. Food Science and Technology Research 5, 18–23.

Pointillart A, Fourdin A, Fontaine N. 1987. Importance of cereal phytase activity for phytate phosphorus utilization by growing pigs fed diets containing triticale or corn. Journal of Nutrition 117, 907–913.

Ram S, Verma A, Sharma S. 2010. Large variability exits in phytase levels among Indian wheat varieties and synthetic hexaploids. Journal of Cereal Science 52, 486–490.

Ritchie S, Gilroy S. 1998. Gibberellins: regulating genes and germination. New Phytologist 140, 363–383.

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. Nature Reviews Genetics 3, 429–441.

Sandberg AS, Andersson H. 1988. Effect of dietary phytase on the digestion of phytate in the stomach and small-intestine of humans. Journal of Nutrition 118, 469–473.

Steiner T, Mosenthin R, Zimmermann B, Greiner R, Roth S. 2007. Distribution of phytase activity, total phosphorus and phytate phosphorus in legume seeds, cereals and cereal by-products as influenced by harvest year and cultivar. Animal Feed Science and Technology 133, 320–334.

Takaishi F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K. 1996. Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Molecular Biology 30, 1207–1221.

Türk M, Carlsson NG, Sandberg AS. 1996. Reduction in the levels of phytate during wholemeal bread making; effect of yeast and wheat phytases. Journal of Cereal Science 23, 257–264.

Viveros A, Centeno C, Brenes A, Canales R, Lozano A. 2000. Phytase and acid phosphatase activities in plant feedstuffs. Journal of Agricultural and Food Chemistry 48, 4009–4013.

Zhang J. 2003. Evolution by gene duplication: an update. Trends in Ecology and Evolution 18, 292–298.

Zhu B, Cai G, Hall EO, Freeman GJ. 2007. In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. Biotechniques 43, 354–359.