Characterization and Transposon Mutagenesis of the Maize (*Zea mays*) *Pho1* Gene Family

M. Nancy Salazar-Vidal[1], Edith Acosta-Segovia[1], Nidia Sanchez-Leon[1], Kevin R. Ahern[2], Thomas. P. Brutnell[3], and Ruairidh J. H. Sawers[1,*],

1 Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Unidad de Genómica Avanzada, Centro de Investigaciones y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Irapuato C.P. 36821, Guanajuato, México
2 Boyce Thompson Institute for Plant Research, Ithaca, New York 14853-1801, U.S.A.
3 Donald Danforth Plant Science Center, St. Louis, Missouri 63132, U.S.A.

*** rsawers@langebio.cinvestav.mx**

Abstract

Phosphorus is an essential nutrient for all plants, but is one of the least mobile, and consequently least available, in the soil. Plants have evolved a series of metabolic and developmental adaptations to increase the acquisition of phosphorus and to maximize the efficiency of use within the plant. In Arabidopsis (*Arabidopsis thaliana*), the PHO1 protein regulates and facilitates the distribution of phosphorus within the plant. To investigate the role of PHO1 in maize (*Zea mays*), we searched the B73 reference genome for homologous sequences and identified four genes that we designated *ZmPho1;1*, *ZmPho1;2a*, *ZmPho1;2b* and *ZmPho1;3*. *ZmPho1;2a* and *ZmPho1;2b* are the most similar to *AtPho1*, and represent candidate co-orthologs that we hypothesize to have been retained following a whole genome duplication. Tissue- and phosphate-specific differences in the accumulation of *ZmPho1;2a* and *ZmPho1;2b* transcripts suggest functional divergence. The presence of phosphate-regulated anti-sense transcripts derived from both *ZmPho1;2a* and *ZmPho1;2b*, suggest the possibility of regulatory crosstalk between paralogs. To characterize fully functional divergence between *ZmPho1;2a* and *ZmPho1;2b*, we conducted a Ds transposon mutagenesis and describe here the generation of novel insertion alleles.

Introduction

Phosphorus (P) is an essential nutrient for all plants and a limitation on productivity in many agricultural systems [1]. Current levels of agricultural phosphorus inputs are recognized to be both unsustainable and environmentally undesirable [2]. Rational strategies to improve P efficiency in agricultural systems demand a greater understanding of P relations in crop plants, both in terms of P uptake from the soil and P translocation and use within the plant.

The protein PHO1 has been characterized from Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) to play a key role both in the export of inorganic P (Pi) into the xylem apoplast for translocation [3] and in the modulation of long-distance signals

underlying the P-deficiency response [4]. The Arabidopsis *Atpho1* mutant exhibits severe Pi deficiency in the shoots but normal Pi content in the roots, displaying associated symptoms of phosphate deficiency, including reduced growth rate, thinner stalks, smaller leaves, very few secondary inflorescences, delayed flowering and elevated levels of anthocyanin accumulation [3]. The ortholog of *AtPHO1* in rice has been designated *OsPho1;2* [5]. Disruption of *OsPHO1;2* results in phenotype very similar to that of the *Atpho1* mutant [5], suggesting that the two genes are functionally equivalent. Indeed, expression of *OsPHO1;2* in the *Atpho1* background will partially complement the mutant phenotype [6]. A feature distinguishing the rice *OsPHO1;2* gene from its Arabidopsis ortholog is the production of a *cis*-Natural Antisense Transcript (*cis*-NAT) [5] [7]. *Cis*-NATs are defined as RNA transcripts generated from the same genomic loci as their associated sense transcript, but on the opposite DNA strand, thus generating a transcript containing a region with perfect complementarity to the sense transcript [8]. Accumulation of *cis*-NAT$_{OsPHO1;2}$ increases under P limitation promoting translation of the *OsPHO1;2* sense RNA by enhancing the efficiency of polysome loading [9].

The PHO1 protein contains two domains: the N-terminal hydrophilic SPX domain (named for the yeast proteins Syg1 and Pho81, and the human Xpr1) and the C-terminal hydrophobic EXS domain (named for the yeast proteins ERD1 and Syg1 and the mammalian Xpr1) [10]. The SPX domain is subdivided into three well-conserved sub-domains, separated from each other by regions of low conservation. SPX domain containing proteins are key players in a number of processes involved in P homeostasis, such as fine tuning of Pi transport and signaling by physical interactions with other proteins [7]. Following the SPX domain there are a series of putative membrane-spanning α-helices that extend into the C-terminal EXS domain [10]. In AtPHO1, the EXS domain is crucial for protein localization to the Golgi/*trans*-Golgi network and for Pi export activity. In addition, the AtPHO1 EXS domain is involved in the modulation of long-distance root-to-shoot signaling under P limitation [4].

Despite the importance of maize as a staple crop and the dependence of maize production on large-scale input of phosphate fertilizers, the molecular components of maize P uptake and translocation remain poorly characterized [11]. Although it has been possible to identify maize sequences homologous to known P-related genes from other species, functional assignment has been based largely on patterns of transcript accumulation. With the development of accessible public-sector resources, it is now feasible to conduct reverse genetic analyses in maize. Here, we aim to extend the molecular characterization of maize P response by generating mutant alleles of maize *Pho1* genes using endogenous *Activator/Dissociation* (*Ac/Ds*) transposable elements. The (*Ac/Ds*) system consists of autonomous *Ac* elements that encode a transposase (TPase) and non-autonomous *Ds* element that are typically derived from *Ac* elements by mutations within the TPase gene. Lacking TPase, *Ds* elements are stable, unless mobilized by TPase supplied in trans by an *Ac*. *Ac/Ds* elements move via a cut-and-paste mechanism [12], with a preference for transposition to linked sites [13] that makes the system ideal for local mutagenesis [14]. To exploit the system for reverse genetics, *Ac* and *Ds* elements have been distributed throughout the genome and placed on the maize physical map, providing potential "launch pads" for mutagenesis of nearby genes [15] [16].

In this study, we identify four maize *Pho1* like genes in the maize (var. B73) genome, including two (*ZmPho1;2a* and *ZmPho1;2b* that we consider co-orthologs of *AtPHO1*. We provide experimentally determined gene structures of *ZmPho1;2a* and *ZmPho1;2b* and characterize transcript accumulation in the roots and shoots of seedlings grown in P-replete or P-limiting conditions. We present novel insertional alleles of *ZmPho1;2a* and *ZmPho1;2b* generated using the *Ac/Ds* transposon system.

Materials and Methods

Identification of maize *Pho1* genes

The AtPHO1 cDNA sequence (GenBank ID: AF474076.1) was used to search the maize working geneset peptide database (www.maizesequence.org) in a BLASTX search performed under the default parameters. Identified maize sequences were in turn used to reciprocally search *Arabidopsis thaliana* (www.phytozome.net). Four sequences were identified with a high level of similarity to AtPHO1: GRMZM5G891944 (chr 3:28,919,073-28,923,871); GRMZM2G466545 (chr 4:171,946,555-171,952,268); GRMZM2G058444 (chr 5:215,110,603-215,115,635) and GRMZM2G064657 (chr 6:122,577,593-122,582,074). The putative protein sequences were confirmed to contain canonical PHO1 domain structure by PFam analysis (pfam.sanger.ac.uk) and NCBI conserved domains search (www.ncbi.nlm.nih.gov).

Amplification of full length *Pho1;2* cDNAs

Total RNA was extracted using Trizol-chloroform from the roots of 10-day-old B73 seedlings grown under phosphate limiting conditions (sand substrate; fetilized with 0mM PO_4). 1.5μg of RNA was used to synthesize cDNA with oligo(dT) primer and *SuperScript III* Reverse Transcriptase (Invitrogen, in Guanajuato, Mexico) in a reaction volume of 20μl. PCR amplification of full length *Pho1;2* cDNAs was performed using using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) under the following cycling conditions: initial incubation at 95°C for 3min, followed by 35 cycles of 94°C for 30sec, 61°C for 30sec and 68°C for 30sec. Primers used are shown in the Table 1.

Analysis of *Pho1;2* transcript accumulation

Total RNA was extracted using Trizol-Chloroform from the roots of 10-day-old B73 seedlings grown under P replete (sand substrate; fetilized with 1mM PO_4) or P limiting (sand substrate; fertilized with 0mM PO_4) conditions. cDNA was synthesized as described above. PCR amplification of was performed under the following cycling conditions: initial incubation at 95°C for 5min, followed by 32 cycles of 95°C for 30sec, 63°C for 30sec and 72°C for 1min. Primers are shown in Table 1. Maize ubiquitin (GRMZM2G419891) was used as a control. Products were analyzed on 1.5% agarose gels.

Table 1. Primers used in this study for CDS amplification

Target	Forward Primer 5'-3'	Reverse Primer 5'-3'	D
pho1;2a	MS103-TCGCGGAGGATGGTGAAGTTCT	MS062-GTGTGTCCATTCCTGGAACTCT	Fu
pho1;2a	MS002-AGGTGGCCATGAAGTACCTG	MS056-CCTGCATTGCTCTCCAGTAGTAA	Ex
pho1;2b	MS013-ATCCCACGATGGTGAAGTTCT	MS077-ACAATTCTCAATCGACCACTAGC	Fu
pho1;2b	RS138-AACTCCGTCTCGGTATGGTGGAGTCT	RS139-TGACCAGAACGCCTCATGTTATACCC	Ex
cis-NAT$_{Pho1;2a}$	MS121-GAGCATCCTGATTCCATATCTACC	MS120-CCCGTAATGGAAGCTTTTACTG	Fl
cis-NAT$_{Pho1;2b}$	MS131-CGTACTGCTGATCGCATGCATA	MS132-AGTACGTGATCAGTGATCTACACTC	Fl
ZmUbq	RS164-CTACAACATTCAGAAGGAGAGCAC	RS165-TCTGCAAGGGTACGGCCATCC	

Transposon mutagenesis

The strategy for *Ac/Ds* mutagenesis was as previously described [15]; [16]. Genetic stocks were maintained in the T43 background, a color-converted W22 stock carrying

r1-sc::m3, a *Ds6-like* insertion in the *r1* locus that controls anthocyanin accumulation in aleurone and scutellar tissues (Alleman and Kermicle 1993). The frequency of purple spotting in the aluerone as a result of somatic reversion of *r1-sc::m3* was used to monitor *Ac* activity (McClintock1951). Donor *Ac* and *Ds* stocks were selected from existing collections [15] [16]: the element *Bti31094::Ac* is placed on the B73 physical map 650.8Kb from *ZmPho1;2a*; the element *I.S06.1616::Ds* is inserted in Intron 13 of *ZmPho1;2b*, and was subsequently designated *ZmPho1;2b-m1::Ds*. To generate a testcross population for mutagenesis of *ZmPho1;2a*, 207 indiviudals homozygous for *Bti31094::Ac* were crossed as females by T43, and rare finely spotted progeny kernels selected for screening. To remobilize the *Ds* element *I.S06.1616::Ds* within *ZmPho1;2b-1*, homozygous *ZmPho1;2b-1::Ds* individuals carrying the unlinked stable transposase source *Ac-Immobilized Ac-im* (Conrad2005) were used as males to pollinate T43, and coarsely spotted progeny kernels were selected for screening.

To identify novel *Ac/Ds* insertions into *pho1;2a* and *pho1;2b*, selected kernels were germinated in the greenhouse and DNA isolated from pools of 18 seedlings. The candidate gene space was explored by PCR performed with a range of gene specific primers used in combination with outward-facing *Ac/Ds* primers. All primers are listed in Table 2. PCR reactions contained 400ng DNA and .25μM of each primer. Reactions for *pho1;2a* were performed using Platinum Taq DNA Polymerase High Fidelity (Invitrogen; in NY, USA,) under the following cycling conditions: denaturation at 94^{o}c for 4 min; 30 cycles of 94^{o}c for 30 sec, 60c for 30 sec, 72^{o}c for 3min 30 sec;final extension at 72^{o}c for 10 min. Reactions for *pho1;2b* were performed using Kappa Taq DNA polymerase (Kapa Biosystems; in Guanajuato, Mexico) under the following cycling conditions: denaturation at 95^{o}c for 5 min; 35 cycles of 95^{o}c for 30 sec, 58^{o}c for 30 sec, 72^{o}c for 3min 30 secs; final extension at 72^{o}c for 5 min. Positive pools were re-analyzed as individuals, following the same cycling conditions. PCR reactions were analyzed on 1.5% agarose gels. Products from positives individuals were purified using QIAquick PCR Purification Kit (Qiagen), ligated into pGEM T-easy vector (Promega) and sequenced.

Footprint alleles generation was carried on using homozygous individuals for *pho1;2a-m1::Ac/pho1;2a-m1::Ac* as males and testcrossed by T43. Rare non-spotted progeny kernels were selected. Putative *Ac* excision was confirmed by PCR across the site using primers shown in Table 2, and amplified with Kappa Taq DNA polymerase (Kapa Biosystems; in Guanajuato, Mexico) under the following cycling conditions: denaturation at 95^{o}c for 5 min; 35 cycles of 95^{o}c for 30 sec, 58^{o}c for 30 sec, 72^{o}c for 3min 30 secs; final extension at 72^{o}c for 5 min. PCR productS from each individuals were purified using QIAquick PCR Purification Kit (Qiagen), ligated into pGEM T-easy vector (Promega, in Guanajuato, Mexico) and sequenced.

Table 2. Primers used in this study for genotyping

Target	Forward Primer 5'-3'	Reverse Primer 5'-3'	Descr
pho1;2a	MS069-ACCTTTCTACACTGCCTGTACC	MS044-GTTCGACCTACCTAACATGGACT	1st fra
pho1;2a	MS043-GTTTGTACGTACCCATGCCGTAT	MS048-GGAAGGGAAGTACCTTGTCAGAG	2nd fr
pho1;2a	MS047-CTCCCTCAATGTGAAGGCTTT	MS062-GTGTGTCCATTCCTGGAACTCT	3rd fr
pho1;2b	MS085-CTCATTTGTTTCCAGTTTCTCTCC	N/A	1st fra
pho1;2b	MS013-ATCCCACGATGGTGAAGTTCT	MS016-GTGCTGGAGAAGTCGAAGATG	2nd fr
pho1;2b	MS047-CTCCCTCAATGTGAAGGCTTT	MS077-ACAATTCTCAATCGACCACTAGC	3rd fr
pho1;2a'-m1.1	MS052-GCATCCTAATAAAGCCTGGAAGA	MS124-CACACTCATCATCTGAACAAAGCAAG	flanki

Results

The maize genome contains four *PHO1* homologs

To identify maize *Pho1* genes, we searched the B73 reference genome (B73 reference genome v3; www.maizeGDB.org) to identify gene models whose putative protein products exhibit a high degree of similarity to the *Arabidopsis* protein AtPHO1 (Table 3). We identified four such maize gene models, and, on the basis of similarity to previously annotated rice genes Secco, 2010, designated these *ZmPho1;1* (GRMZM5G891944), *ZmPho1;2a* (GRMZM2G466545), *ZmPho1;2b* (GRMZM2G058444) and *ZmPho1;3* (GRMZM2G064657). To investigate orthology among *Arabdidopsis* and grass *PHO1* genes, we identified additional sequences from sorghum (*Sorghum bicolor*) and canola (*Brassica rapa*), and performed a multiple sequence alignment of the putative protein sequences to generate a distance tree (Fig. 1).

In our analysis, we included from *Arabidopsis* only the proteins AtPHO1 and AtPHO1;H1, leaving aside a large clade of divergent functionally distinct PHO1 proteins that are specific to dicotyledonous plants Stefanovic, 2007. We used a PHO1 protein from the moss *Physcomitrella patens* to root the tree. Our analysis supported the previously reported divergence of PHO1 and PHO1;H1 clades, dating from before divergence of monocotyledonous and dicotyledonous plants Secco, 2010; Stefanovic, 2007. Within the PHO1;H1 clade, we observed a duplication event specific to the grasses in our analysis. As a result, the three grass species each contain two co-orthologs of AtPHO1;H1 – encoded by the genes annotated *Pho1;1* and *Pho1;3*. We observed also an expansion of the PHO1;H1 clade in canola, although this expansion is lineage specific, and there is no suggestion that Pho1;H1 was not a single gene at the base of this clade. The PHO1 clade itself contains the products of single-copy *Pho1/Pho1;2* sequences in all species in our analysis, with the exception of a lineage-specific duplication in maize. As a consequence, the duplicated maize genes *ZmPho1;2a* and *ZmPho1;2b* are both considered to be orthologous to *AtPho1*.

Table 3. Homologous *Pho1* genes

***At* Gene**	***Os* Gene**	***Zm* Gene**	***Zm* Gene Model**	***Zm* Position**
AtPHO1	*OsPho1;2*	*ZmPho1;2a*	GRMZM2G466545	chr 4:171,946,555-171,952,268
AtPHO1	*OsPho1;2*	*ZmPho1;2b*	GRMZM2G058444	chr 5:215,110,603-215,115,635
AtPHO1;H1	*OsPho1;1*	*ZmPho1;1*	GRMZM5G891944	chr 3:28,919,073-28,923,871)
AtPHO1;H1	*OsPho1;3*	*ZmPho1;3*	GRMZM2G064657	chr 6:122,577,593-122,582,074

ZmPho1;2a and *ZmPho1;2b* show features of syntenic paralogs retained following to genome duplication

The high degree of sequence similarity between *ZmPho1;2a* and *ZmPho1;2b* suggests that they result from a recent gene duplication event. This interpretation is supported by the observation that *Pho1;2* is a single copy sequence in both rice and sorghum. It has been hypothesized that the last whole genome duplication (WGD) event in maize occurred between 5 and 12 million-years-ago, sometime after divergence from the sorghum lineage, as the result of polyploidization [17]. Following WGD, the maize genome has returned to a diploid organization, and the majority of duplicate sequences have been lost through a process known as fractionation [17]. In certain cases, however, both paralogs have been retained. To assess whether *ZmPho1;2a* and *ZmPho1;2b* might represent a pair of paralogs retained following WGD, we investigated synteny between the maize and sorghum genomic regions in which the *Pho1;2* sequences are present.

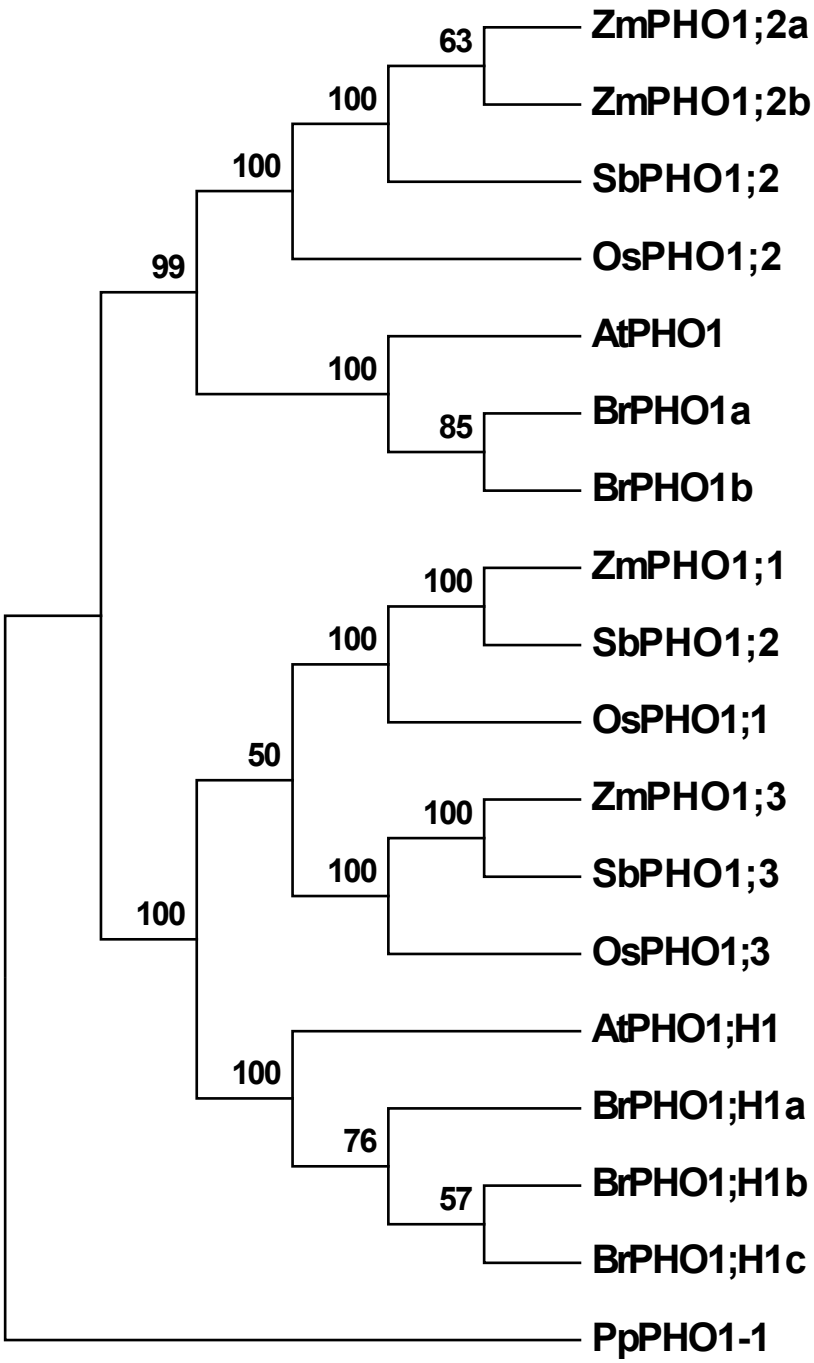

Figure 1. Homologues for Arabidopsis PHO1 and Rice Pho1 family were identified in the B73 genome sequence. The tree was made using the Maximum Likelihood method based on the JTT matrix-based model with AtPHO1, AtPHO1;H1, the *Brassica rapa*, rice, sorghum and maize PHO1 family. Evolutionary analyses were conducted in MEGA6.

ZmPho1;2a and *ZmPho1;2b* are located on chromosomes (Chr) 4 and 5, respectively (Table 3), and the homologous gene *SbPho1;2* on sorghum Chr4. The regions carrying the two maize *Pho1;2* genes have been assigned previously to distinct pre-tetraploid ancestral genomes (Chr4:168,085,162..179,711,318 to sub-genome 2; Chr5:208,925,180..217,680,842 to sub-genome 1; [17]). Furthermore, both maize regions of maize exhibit synteny with the region of sorghum Chr4 that carries *SbPho1;2*. The genomic region surrounding the *Pho1;2* genes exhibits a high frequency of candidate retained paralog pairs (Fig. 2A), providing ample evidence of micro-synteny between the regions containing *ZmPho1;2a* and *ZmPho1;2b*. In all cases, sorghum and maize *Pho1;2* genes are adjacent to a putative WD40 protein encoding gene, present on the opposite strand and partially overlapping the annotated 3' UTR region of the *Pho1;2* sequence (Fig. 2B), a feature not observed in the other *Pho1* paralogs.

Transcripts encoded by *ZmPho1;2a* and *ZmPho1;2b* exhibit divergent patterns of expression

To determine if both *ZmPho1;2a* and *ZmPho1;2b* are potentially functional, we used RT-PCR to amplify gene-specific fragments of the annotated 3' UTR regions of the two genes from cDNA prepared from roots or leaves of 10-day-old seedlings (B73) grown in sand, watered with either complete Hoagland solution (+P) or a modified phosphate-free Hoagland solution (-P) (Fig. 3A). Transcripts of *ZmPho1;2a* were detected in roots but not leaves, with an indication of greater accumulation under -P. In contrast, transcripts of *ZmPho1;2b* showed constitutive accumulation in in both roots and leaves, indicating that the two maize *Pho1;2* genes, while potentially both functional, have diverged at the level of transcript accumulation. We examined also the accumulation of *SbPho1;2* transcripts in sorghum (Tx623) seedlings under the same growth conditions. Transcripts of *SbPho1;2* accumulated in a pattern similar to that observed for *ZmPho1;2b*, suggesting that constitutive expression was the ancestral state. Subsequently, we designed additional gene-specific PCR primers to the 5' UTR regions of *ZmPho1;2a* and *ZmPho1;2b* and used these to amplify the complete coding-sequence from cDNA and to confirm the gene-model structure.

ZmPho1;2a and *ZmPho1;2b* are associated with phosphate-regulated *cis*-natural anti-sense transcripts

Although we observed differential transcript accumulation between maize *ZmPho1;2a* paralogs, it has been reported that *OsPho1;2* is largely regulated at the post-transcriptional level by an associated with a cis-natural anti-sense transcript (*cis*-NAT$_{OsPho1;2}$), the accumulation of which responds more strongly to P availability than the corresponding sense transcript [5]; [9]. The (*cis*-NAT$_{OsPho1;2}$) transcript has been shown to act as a translational enhancer, and has been proposed to act by direct interaction with the sense transcript [9]. The rice *cis*-NAT$_{OsPho1;2}$ is initiates in Intron 4 of *OsPho1;2* and extends into the 5' UTR region [9]. A putative anti-sense sequence has been annotated in the maize reference genome to initiate in a homologous position in *ZmPho1;2a* (maizeGDB.com; Jabnoune2013), although, on the basis of cDNA evidence, the transcript is considerably shorter than *cis*-NAT$_{OsPho1;2}$ and extends only to Intron 2 of *ZmPho1;2a* (Fig. 2B). No paralogous sequence has been annotated associated with *ZmPho1;2b*.

To confirm the presence of *cis*-NAT transcripts associated with *ZmPho1;2a* and explore the possibility that a paralogous transcript might be generated from *ZmPho1;2b*, we designed gene-specific primers to the introns flanking the homologous Exons 4 and 3 of *ZmPho1;2a* and *ZmPho1;2b*, respectively, and attempted to amplify products from

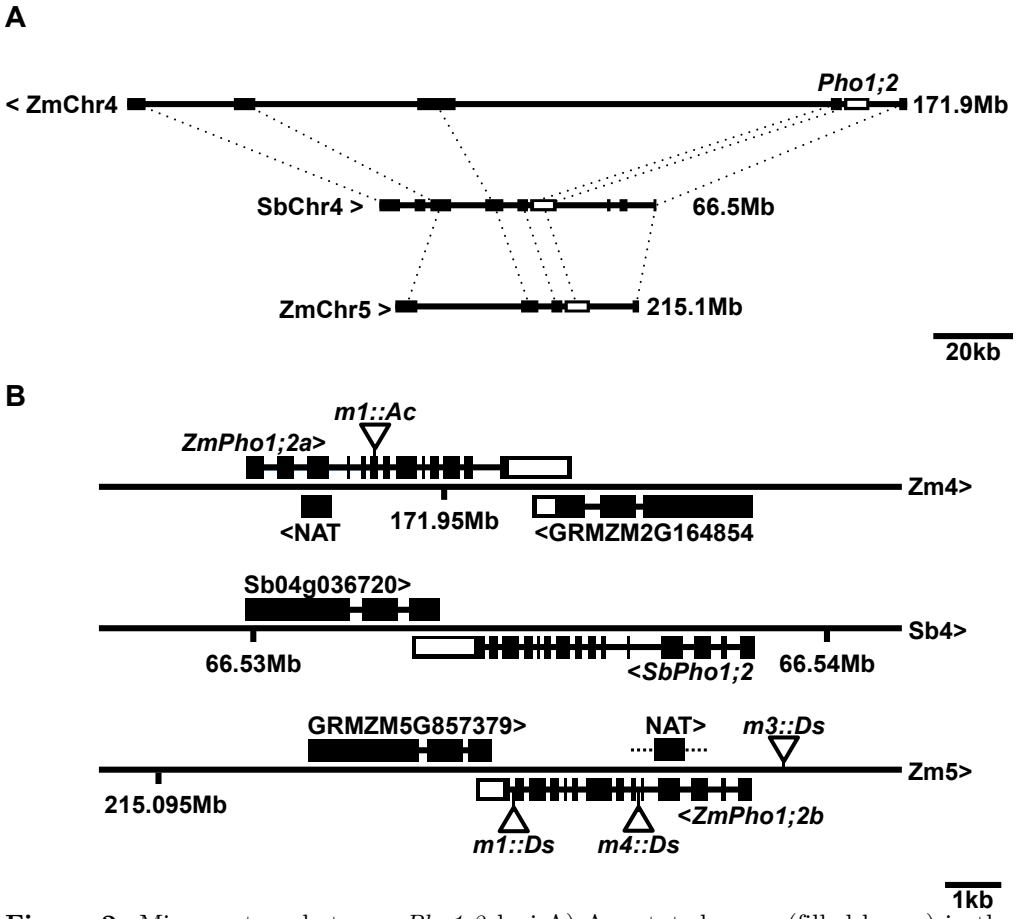

Figure 2. Microsynteny between *Pho1;2* loci A) Annotated genes (filled boxes) in the region of *SbPho1;2* on Sorghum chromosome 4 (SbChr4) and their candidate orthologs on maize (B73) chromosomes 4 (ZmChr4) and 5 (ZmChr5). Orthologous sequences are connected by dashed lines. *Pho1;2* genes on the three chromosomes shown in unfilled boxes. Regions shown to scale, the right hand position indicated. SbChr4 and ZmChr5 run left to right, ZmChr4 right to left. B) *Pho1;2* gene models on ZmChr4, SbChr4 and ZmChr5. Exons are shown as boxes, coding regions filled, UTR unfilled. Anti-sense transcripts associated with *ZmPho1;2a* and *ZmPho1;2b* shown as filled boxes (NAT). Angle brackets indicate the direction of transcription. The maize syntenic paralog pair GRMZM2G164854/GRMZM5G853379 and their sorghum ortholog are also shown. Triangles indicate the position of *Activator* and *Dissociation* insertion, respectively. Refer to Table 4 for full description of insertional alleles. Shown to scale.

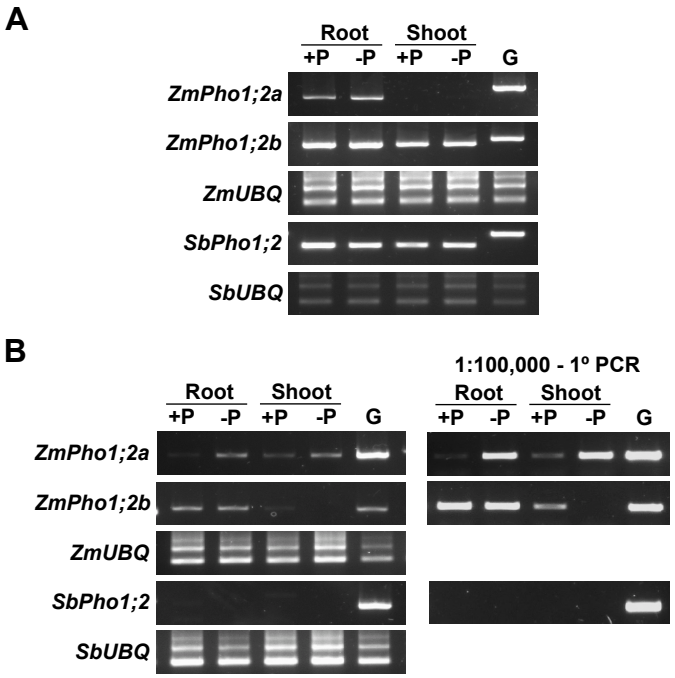

Figure 3. Accumulation of *Pho1;2* sense and antisense transcripts under contrasting phosphate conditions. A. Accumulation of sense transcripts encoded by *ZmPho1;2a*,*ZmPho1;2b* and *SbPho1;2*expression in roots and shoot of 10-day-old B73 seedlings fertilized with Hoagland solution adjusted to 1mM (+P) and 0mM (-P) PO_4^{-3}. B. Accumulation of putative anti sense transcripts encoded by *ZmPho1;2a*,*ZmPho1;2b* in seedlings as Panel A. Primary PCR (left) and nested PCR (right) performed using 1:100,000 dilution of primary reaction as template. Amplification of transcripts encoded by *ZmUBQ*, *SbUBQ* and amplification from 50ng genomic DNA were template (G) were used as controls.

cDNA prepared from seedling root and leaves as described above. Products of the predicted size were amplified using both *ZmPho1;2a* and *ZmPho1;2b* primer sets, consistent with the accumulation of *cis*-NATs (Fig. 3B). No products were amplified from no-RT control samples (data not shown). Putative *cis*-NAT products were sequenced and confirmed to originate from the *ZmPho1;2a* and *ZmPho1;2b* genes. We had not observed evidence of accumulation of alternatively or partially spliced transcripts in our amplification of the full length coding sequence.

The accumulation of *cis*-NAT$_{ZmPho1;2a}$ was observed to be induced under low P in both roots and leaves. The transcript *cis*-NAT$_{ZmPho1;2b}$ was observed to accumulate in roots but not leaves, with no response to P availability, providing further evidence of functional divergence between the paralogs. Interestingly, using the approach we employed in maize, we found no evidence of an equivalent *cis*-NAT associated with *SbPho1;2* (Fig. 3B), although additional experiments will be required to rule out the possibility that antisense transcripts are produced from other regions of the sorghum gene.

Transposon mutagenesis of *ZmPho1;2a* and *ZmPho1;2b*

To investigate functional divergence between ZmPho1;2a and *ZmPho1;2b*, we initiated a program to mutagenize both loci using the endogenous *Activator/Dissociation* (*Ac/Ds*) transposon system. Autonomous *Ac* elements and their non-autonomous *Dissociation* (*Ds*) derivatives are members of the hAT (*hobo- Activator- Tam3*) family of DNA transposons, moving via a cut-and-paste mechanism. Both *Ac* and *Ds* show a strong preference for reinsertion into locations genetically linked to the donor site. This enrichment for local transposition can be exploited for mutagenesis by identification and remobilization of a donor element closely linked to a gene of interest. Once established, it becomes possible to generate multiple alleles from a single test-cross population.

To mutagenize *ZmPho1;2a*, we recovered 1082 novel transposition events from the element *Ac* (*bti31094::Ac*) located 650.8kb upstream the target (Fig. 4). A PCR-based strategy was designed to screen for reinsertion of *Ac* into ZmPho1;2a in which the gene was divided into three overlapping fragments. Allowing for both possible orientations of *Ac* insertion, we performed a total of 12 reactions to cover the gene space, screening first pools of 18 seedlings, and subsequently the individuals constituting positive pools. Putative insertions were re-amplified using DNA extracted from a second seedling leaf to discount somatic events. Using this strategy, we recovered a novel germinal *Ac* insertion in Exon 6 of *ZmPho1;2a* (*zmpho1;2a-m1::Ac*). Left and right flanking border fragments were amplified and sequenced, confirming the exact location of the element and identifying an 8bp target site duplication (AGCCCAGG) consistent with *Ac* insertion.

To mutagenize *ZmPho1;2b*, we remobilized *I.S06.1616::Ds*), a *Ds* element identified to be inserted in intron 13 of the target gene, and that we designated *Zmpho1;2b-s1::Ds*. Plants homozygous for the *Zmpho1;2b-s1::Ds* allele did not present any observable phenotype and RT-PCR analysis of transcript accumulation indicated such plants to accumulate correctly-spliced transcript to normal levels (data not shown). To derive novel functionally significant insertions, individuals homozygous for *Zmpho1;2b-s1::Ds* carrying the unlinked stable transposase source *Ac-immobilized* were crossed as males to T43 (Fig. 5). Test-cross progeny were screened using a strategy similar to that employed in the mutagensis of *ZmPho1;2a*. Two novel *Ds* insertions were identified, one in the promoter region, 591bp upstream of the ATG (*zmpho1;2b-s3::Ds*) and the second in intron 5 (*zmpho1;2b-s4::Ds*). Sequencing of the region flanking each novel insertions identified the expected 8bp target site duplication.

Derivation of stable derivatives of *ZmPho1;2a*

To generate stable "footprint" alleles by *Ac* excision from *ZmPho1;2a-m1::Ac*, a *ZmPho1;2a-m1::Ac* homozygous individual was crossed as male to T43, and the resulting progeny screened to identify rare colorless kernels, indicating loss of *Ac* function (Figure 7). Colorless kernels were germinated, DNA extracted from seedlings, and PCR used to amplify the genomic region spanning the *ZmPho1;2a-m1::Ac* insertion site. Products of a size consistent with *Ac* excision were cloned and sequenced. Two footprint alleles were identified, one with an 8bp insertion (GCCCAGCT) (*ZmPho1;2a'm1.1*) and the second with a 5bp insertion (GCCCA) (*ZmPho1;2a'm1.2*). For confirmation, and to facilitate future genetic experiments, the region spanning *ZmPho1;2a'm1.1* was re-amplified and digested with the enzyme *Bse*YI that specifically recognized a target site generated by the 8bp duplication 8. As a result of non-triplet duplication, both *ZmPho1;2a'm1.1* and *ZmPho1;2a'm1.2* alleles disrupt the DNA reading frame and our predicted to result in a premature termination of translation.

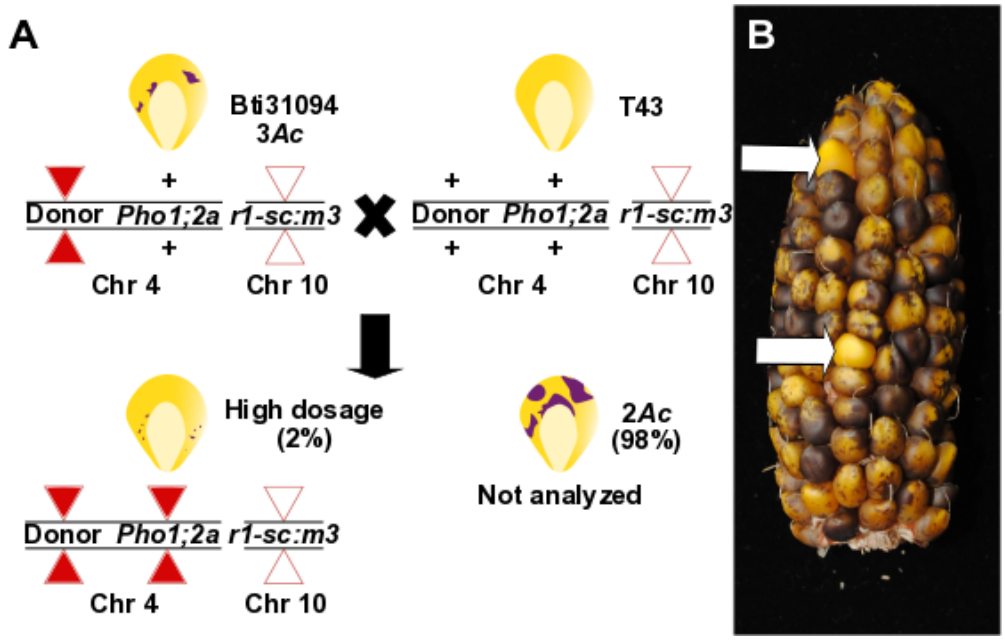

Figure 4. *Ac* mutagenesis of *ZmPho1;2a* gene. (A) Diagram of testcross to mobilized the *Ac* element into the *ZmPho1;2a* gene from the line Bti31094 (B) Ear carrying kernels with two copies of *Ac* in the triploid aleurone, from population of *ZmPho1;2a* tagging. An example of high dose kernels used for novel alleles identification are pointed by white arrows.

Discussion

Maize is the most widely grown cereal in the world. Much of this cultivated area is P limited. And yet, the molecular basis of P uptake and translocation in maize remains poorly characterized (reviewed in [11]). In this study, we have described the maize *Pho1* gene family and generated novel insertional mutant alleles of *pho1* using the endogenous maize *Ac/Ds* transposon system. The genetic material described will open the way to functional analysis of P homeostasis in maize.

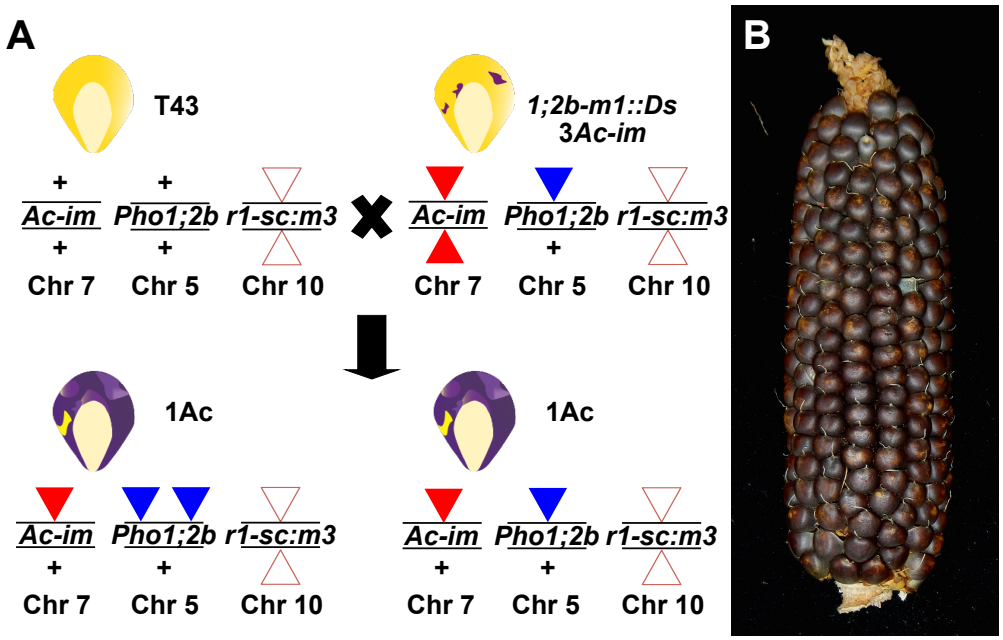

Figure 5. *Ds* mutagenesis of *ZmPho1;2b* gene. (A) Diagram of testcross to remobilized the *Ds* element from the intron 13th on *ZmPho1;2b* gene in the line I.S06.1616 (B) Ear carrying kernels with one copy of *Ac-im* in the chr7 expressed in the triploid aleurone, from population of *ZmPho1;2b* reinsertion.

Table 4. Transposon insertion alleles of maize *Pho1;2* genes

Alleles	Position to ATG (bp)	Description
pho1;2a-m1::Ac	2261	*Ac* insertion on 6th exon, with AGCCCAGG as TSD sequence.
pho1;2a'-m1.1	2254	Footprint generated from *pho1;2a-m1::Ac*, with 8bp insertion (GCCCAGCT) on
pho1;2a'-m1.2	2254	Footprint generated from *pho1;2a-m1::Ac*, with 5bp insertion (GCCCA) on 6th
pho1;2b-m1::Ds	4104	*Ds* insertion on 13th intron, with GGTGGGAG as TSD sequence.
pho1;2b-m3::Ds	-583	*Ds* insertion on Promoter, with ATCACTAT as TSD sequence.
pho1;2b-m4::Ds	1943	*Ds* insertion on 5th intron, with CACACGCT as TSD sequence.

B73 genomic sequence as reference.

The maize *Pho1* family consists of four genes, corresponding to the three gene (*PHO1;1*, *PHO1;2*, *PHO1;3*) structure reported previously in rice (*Oryza sativa*) [5], with the elaboration of the duplication of *PHO1;2*. The sorghum *Pho1* family was also found to consist of three genes. The restricted *PHO1* family present in these cereals is in contrast to larger 11 members family of Arabidopsis [10]. Specifically, the cereals lack a large clade of *PHO1* related sequences present in Arabidopsis that has been implicated in a range of biological functions [10, 18, 19]. Indeed, in studies aimed at complementing the *Atpho1* by expression of other Arabidopsis *PHO1* family members it was only *AtPHO1;H1* that could rescue the mutant [18]. Similar experiments expressing rice *PHO1* genes in the null *Atpho1* background led to plants that despite the low shoot Pi as indication of low Pi transfer activity, plants maintain normal growth and suppression of the Pi-deficiency response, this phenotype resulted similar to PHO1 underexpressing lines, suggesting the PHO1 role in the way the available Pi is used and partitioned [6]. Recently the AtPHO1 role by the EXS domain has been address to Pi export a [4].

Although phylogenetic analysis and experimental data from Arabidopsis and rice

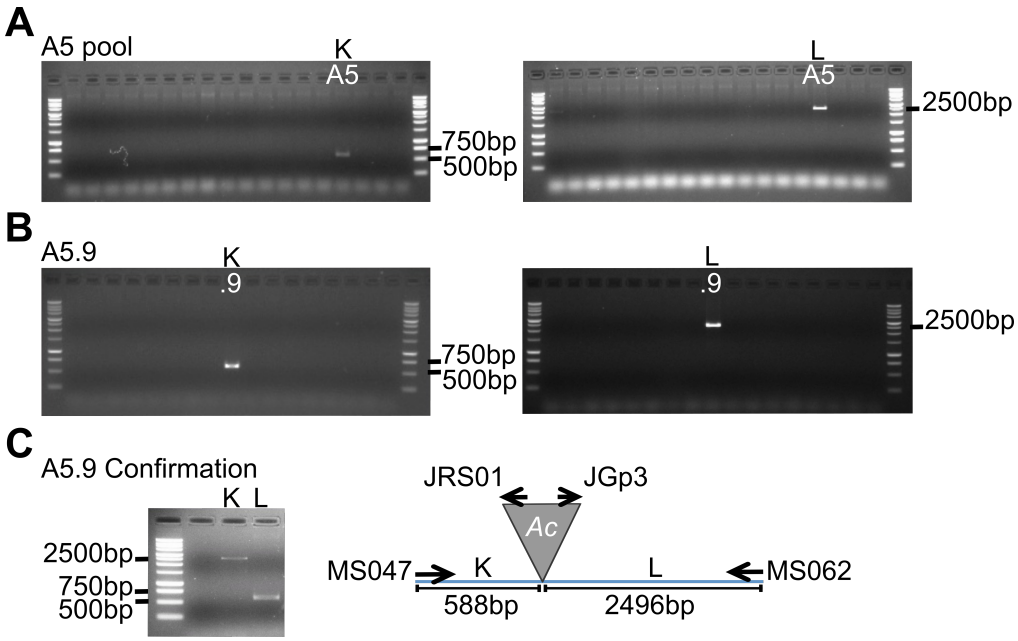

Figure 6. PCR screening for A5.9 allele detection on ZmPho1-2a gene. (A) PCR reactions done using DNA pool. Pool A5 shows amplification in complementary reactions K and L. (B) PCR reactions using individual DNA. Detection of strong amplification from the individual A5.9 in both complementary reactions. (C) PCR for A5.9 allele confirmation. Was obtained amplification on A5.9 allele in a different tissue. (D)Diagram of expected fragment size on A5.9 allele. Reaction K (588bp) and reaction L (2496bp) are indicated

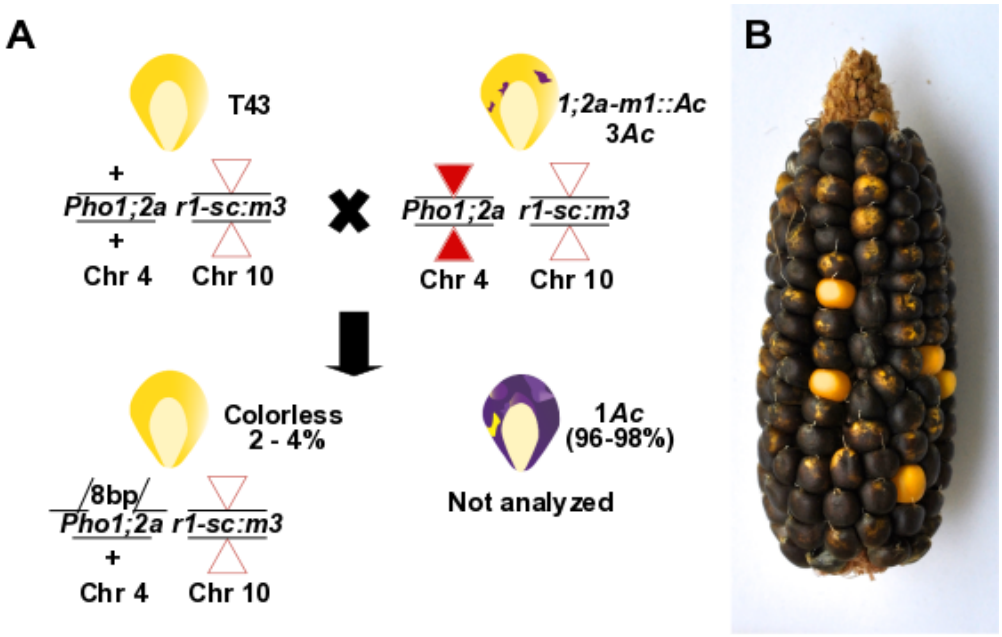

Figure 7. Footprint generation by *Ac* excision event. (A) Diagram of testcross to generate *Ac* excision from *ZmPho1;2a-m1::Ac* homozygous line. (B) The ear shown mostly heterozygous *ZmPho1;2a-m1::Ac* kernels (spotted kernels) but with also isolated events of *Ac* excision (colorless kernels).

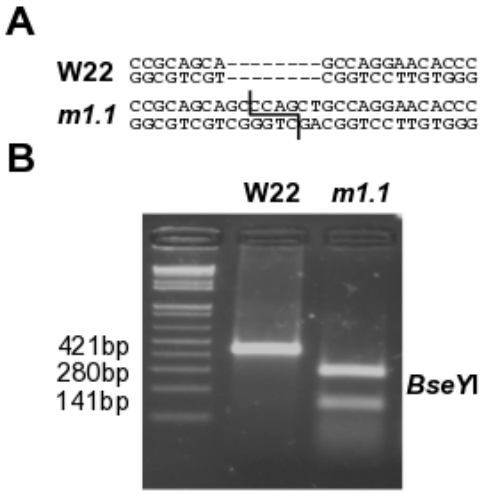

Figure 8. Footprint genotyping by CAPS assay. (A) W22 wild type background shown absence of *Bse*YI cutting site present in *pho1;2a'm1.1* allele as part of footprint sequence (8bp insertion). (B) Gel image representation of *pho1;2a'm1.1* genotyping by CAPS assay, using *Bse*YI enzyme. PCR amplification using specific primers for 421bp product flanking footprint sequence, was digested with *Bse*YI enzyme, when the *pho1;2a'm1.1* allele is present are obtained 280bp and 141bp fragments, and W22 wild type product is not digested. The 1Kb plus DNA ladder (Invitrogen) is on first lane.

suggest all four maize PHO1 to be directly involved in P homeostasis, further work in heterologous systems, and ultimately the analysis of the mutants described here, will be required to determine functional equivalence across species and the biological role in maize.

The lineage leading to maize experienced a tetraploidy event sometime after it split with the sorghum lineage, 5-12 million years ago [17]. Considering contemporary sorghum as an unduplicated outgroup, and taking the structure of the rice *PHO1* family into account, suggests that immediately following this tetraploid event, maize would have carried six *Pho1* genes, represented by three pairs of syntenic paralogs (homeologs). Subsequently, the maize genome has returned to a diploid state through a process of reorganization that has been coupled with extensive fractionation - the loss of one of a pair of syntenic paralogs [20]. Large scale gene loss following whole genome duplication appears to be a general trend, observed across taxa and across timescales [21]. Gene loss is presumed to be buffered by the presence of a functionally equivalent paralogs. Where both paralogs of a syntenic pair are retained, it may indicate either functional constraint or simply incomplete fractionation. In the former case, such constraint would imply either functional divergence or a selective advantage of increased dosage. In maize, 3228 pairs of syntenic paralogs are retained, representing ˜20% of the total complement of ˜32,000 total genes, or closer to ˜10% of the unduplicated gene set [17] [22]. While gene loss is the more likely outcome following genome duplication, it is difficult to determine the balance of selective gene-by-gene reduction and the largely random loss of larger sections of DNA. Similarly, where a pair of syntenic paralogs are retained, as is the case with *Pho1;2*, it may indicate selection directly on the gene pair or a genomic context that insulates the gene pair from larger scale DNA loss events. It is noticeable that a number of syntenic paralog pairs have been retained close to the *Pho1;2* locus, potentially "hitchhiking" on direct selection to maintain one or more of the adjacent pairs. In the case of the pair GRMZM2G164854/GRMZM5G853379, the two paralogs overlap *Pho1;2* sequence on the opposite DNA strand. Consequently, selection to maintain either the *Pho1;2* or GRMZM2G164854/GRMZM5G853379 paralog pair might protect also the adjacent genes from silencing or deletion.

Analysis of *ZmPho1;2a* and *ZmPho1;2b* transcript accumulation clearly demonstrated regulatory divergence between the two paralogs, with *ZmPho1;2b* transcripts presenting a pattern similar to that of *SbPho1;2*, the presumed unduplicated state. Characterization of the leaf transcriptome has estimated 13% of retained syntenic paralogs to undergo regulatory neofunctionalization [22], placing the *Pho1;2* pair among just 2-3% of the total maize gene set. Characterization of putative *cis*-NAT transcripts offered further evidence of regulatory divergence between maize *Pho1;2* paralogs. Accumulation of *cis*-NAT$_{ZmPho1;2a}$ was induced by P limitation, in a manner similar to that observed for *cis*-NAT$_{OsPho1;2}$, while *cis*-NAT$_{ZmPho1;2b}$ accumulation mirrored that of the *ZmPho1;2b* sense transcript. Given that we failed to detect *Pho1;2* associated anti-sense transcripts in sorghum using the techniques applied, we might infer the unduplicated state from the more distantly related rice. Interestingly, on such a basis, our data are consistent with regulatory neofunctionalization, acting on the one hand on *ZmPho1;2a* sense, and on the other on *ZmPho1;2b* anti-sense, transcript accumulation. Although characterized *cis*-NATs act on the activity of adjacent protein coding genes, the translational enhancer function postulated for *PHO1;2* NATs may allow for *trans* action between *ZmPho1;2* paralogs given the degree of sequence similarity.Indeed, one intriguing hypothesis, suggested by our transcript accumulation data, is that subfunctionalization at the maize *PHO1;2* loci has resulted in the primary production of sense transcripts from *Pho1;2b*. Characterization of the insertional alleles described here will be central to determine the functional role of *ZmPho1;2a* and *ZmPho1;2b* and associated sense and anti-sense transcripts. We are continuing to mobilize *Ac* and *Ds*

elements at the maize *Pho1;2* loci, taking full advantage of the capacity of the system to generate allelic series, impacting variously sense and anti-sense transcripts. Such material will be invaluable in the fine-scale evaluation of regulatory crosstalk and functional redundancy between between *ZmPho1;2* paralogs and, ultimately, the biological role of PHO1 proteins in maize.

Acknowledgments

This work was funded by Mexican National Council of Science and Technology (CONACYT) grant CB2012-151947.

References

1. Veneklaas EJ, Lambers H, Bragg J, Finnegan PM, Lovelock CE, Plaxton WC, et al. Opportunities for improving phosphorus-use efficiency in crop plants. New Phytologist. 2012;195:306–320.

2. Vance CP. Update on the State of Nitrogen and Phosphorus Nutrition Symbiotic Nitrogen Fixation and Phosphorus Acquisition . Plant Nutrition in a World of Declining Renewable Resources. 2014;.

3. Poirier Y, Thoma S, Somerville C, Schiefelbein J. Mutant of Arabidopsis deficient in xylem loading of phosphate. Plant physiology. 1991;97:1087–1093.

4. Wege S, Khan GA, Jung JY, Vogiatzaki E, Pradervand S, Aller I, et al. The EXS domain of PHO1 participates in the response of shoots to phosphate deficiency via a root-to-shoot signal. Plant physiology. 2015 Dec;Available from: http://www.ncbi.nlm.nih.gov/pubmed/26546667.

5. Secco D, Baumann A, Yves P. Characterization of the rice PHO1 gene family reveals a key role for OsPHO1; 2 in phosphate homeostasis and the evolution of a distinct clade in dicotyledons. Plant physiology. 2010;152(March):1693–1704. Available from: http://www.plantphysiol.org/content/152/3/1693.short.

6. Rouached H, Secco D, Arpat B, Poirier Y. The transcription factor PHR1 plays a key role in the regulation of sulfate shoot-to-root flux upon phosphate starvation in Arabidopsis. BMC plant biology. 2011 jan;11(1):19. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3036608{&}tool=pmcentrez{&}rendertype=abstract.

7. Secco D, Wang C, Arpat Ba, Wang Z, Poirier Y, Tyerman SD, et al. The emerging importance of the SPX domain-containing proteins in phosphate homeostasis. New Phytologist. 2012 mar;193(4):842–851. Available from: http://doi.wiley.com/10.1111/j.1469-8137.2011.04002.x.

8. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. Trends in Biochemical Sciences. 2004;29(2):88–94.

9. Jabnoune M, Secco D, Lecampion C, Robaglia C, Shu Q, Poirier Y. A Rice cis-Natural Antisense RNA Acts as a Translational Enhancer for Its Cognate mRNA and Contributes to Phosphate Homeostasis and Plant Fitness. The Plant Cell. 2013;25:4166–4182. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.113.116251.

10. Wang Y, Ribot C, Rezzonico E, Poirier Y. Structure and expression profile of the Arabidopsis PHO1 gene family indicates a broad role in inorganic phosphate homeostasis. Plant physiology. 2004;135(May):400–11. Available from: http://www.plantphysiol.org/content/135/1/400.full.

11. Calderon-Vazquez C, Sawers RJ, Herrera-Estrella L. Phosphate deprivation in maize: genetics and genomics. Plant Physiol. 2011;156(3):1067–1077. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21617030.

12. Bai L, Singh M, Pitt L, Sweeney M, Brutnell TP. Generating novel allelic variation through Activator insertional mutagenesis in maize. Genetics. 2007 Mar;175(3):981–92. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1840078&tool=pmcentrez&rendertype=abstract.

13. Dooner HK, Belachew A. Transposition Pattern of the Maize Element Ac from the Bz-M2(ac) Allele. Genetics. 1989;122(2):447–457. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17246501.

14. Brutnell TP, Conrad LJ. Transposon tagging using Activator (Ac) in maize. Methods in molecular biology (Clifton, NJ). 2003 Jan;236:157–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14501064.

15. Kolkman JM, Conrad LJ, Farmer PR, Hardeman K, Ahern KR, Lewis PE, et al. Distribution of Activator (Ac) throughout the maize genome for use in regional mutagenesis. Genetics. 2005;169(2):981–995. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15520264.

16. Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, et al. Genome-wide distribution of transposed Dissociation elements in maize. Plant Cell. 2010;22(6):1667–1685. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20581308.

17. Schnable JC, Springer NM, Freeling M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A. 2011 Mar;108(10):4069–4074. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3053962&tool=pmcentrez&rendertype=abstracthttp://www.ncbi.nlm.nih.gov/pubmed/21368132.

18. Stefanovic A, Ribot C, Rouached H, Wang Y, Chong J, Belbahri L, et al. Members of the PHO1 gene family show limited functional redundancy in phosphate transfer to the shoot, and are regulated by phosphate deficiency via distinct pathways. The Plant journal : for cell and molecular biology. 2007 jun;50(6):982–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17461783.

19. Ribot C, Zimmerli C, Farmer EE, Reymond P, Poirier Y. Induction of the Arabidopsis PHO1;H10 gene by 12-oxo-phytodienoic acid but not jasmonic acid via a CORONATINE INSENSITIVE1-dependent pathway. Plant physiology. 2008;147(June):696–706.

20. Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, et al. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS biology. 2010

Jan;8(6):e1000409. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2893956&tool=pmcentrez&rendertype=abstract.

21. Sankoff D, Zheng C, Zhu Q. The collapse of gene complement following whole genome duplication. BMC genomics. 2010 Jan;11:313. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896955&tool=pmcentrez&rendertype=abstract.

22. Hughes TE, Langdale JA, Kelly S. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. Genome research. 2014 Aug;24(8):1348–55. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4120087&tool=pmcentrez&rendertype=abstract.