3 Biotech (2024) 14:14
https://doi.org/10.1007/s13205-023-03857-9

ORIGINAL ARTICLE

Potential gene editing targets for developing haploid inducer stocks in rice and wheat with high haploid induction frequency

Lakshay Goyal[1] · Mehardeep Kaur[2] · Meghna Mandal[1] · Debasmita Panda[3] · Subhasis Karmakar[3] · Kutubuddin Ali Molla[3] · Dharminder Bhatia[1]

Received: 31 January 2023 / Accepted: 15 November 2023

Abstract
Doubled haploid (DH) breeding is a powerful technique to ensure global food security via accelerated crop improvement. DH can be produced *in planta* by employing haploid inducer stock (HIS). Widely used HIS in maize is known to be governed by *ZmPLA, ZmDMP*, *ZmPLD3*, and *ZmPOD65* genes. To develop such HIS in rice and wheat, we have identified putative orthologs of these genes using in silico approaches. The *OsPLD1*; *TaPLD1*, and *OsPOD6*; *TaPOD8* were identified as putative orthologs of *ZmPLD3* and *ZmPOD65* in rice and wheat, respectively. Despite being closely related to *ZmPLD3*, *OsPLD1* and *TaPLD1* have shown higher anther-specific expression. Similarly, *OsPOD6* and *TaPOD8* were found closely related to the *ZmPOD65* based on both phylogenetic and expression analysis. However, unlike *ZmPLD3* and *ZmPOD65,* two *ZmDMP* orthologs have been found for each crop. *OsDMP1* and *OsDMP2* in rice and *TaDMP3* and *TaDMP13* in wheat have shown similarity to *ZmDMP* in terms of both sequence and expression pattern. Furthermore, analogs to maize DMP proteins, these genes possess four transmembrane helices making them best suited to be regarded as *ZmDMP* orthologs. Modifying these predicted orthologous genes by CRISPR/Cas9-based genome editing can produce a highly efficient HIS in both rice and wheat. Besides revealing the genetic mechanism of haploid induction, the development of HIS would advance the genetic improvement of these crops.

Keywords Doubled haploid · Food security · Haploid inducer stock · Orthologs · Transmembrane helices

Introduction

Climate change is a growing concern for agriculture and has the potential to impact crop production significantly. Long-lasting droughts brought on by increased temperatures imperil global food security. Numerous simulation studies have forecasted crop losses in the coming years. It is assumed that rising temperatures will impact the yield of potatoes, barley, corn, and sorghum in Atlantic Canada (Kang et al. 2022), as well as the semi-arid northern High Plains of Texas (Chen et al. 2019). Moreover, erratic precipitation in Nigeria (Ani et al. 2022) and heat waves in South Asiatic regions have severely impacted crop yields. In North-West India in 2022, the sudden incidence of Southern Black Streak Rice Dwarf Virus outbreak in rice (Baranwal et al. 2022) caused chaos among the growers. Such events have prompted plant scientists to develop strategies to minimize the detrimental effects of climate change on agriculture. This demands plant scientists to come up every time with a new variety for a new unknown problem and that too in a short time. Plant Breeder's toolbox is continuously being supplemented with several approaches such as doubled haploid (DH) breeding, speed breeding, and gene editing. Among these, DH breeding is a useful technique that can help increase genetic gains in a short period of time, when used in combination with other selection approaches (Singh et al. 2019). Conventional plant breeding strategies often take five to six generations of continuous selfing for the fixation of desirable recombinants after hybridization (Kyum et al. 2021). Additionally, higher heterozygosity in earlier generations impedes the phenotypic selection of

✉ Dharminder Bhatia
d.bhatia@pau.edu

[1] Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab 141 004, India

[2] School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, Punjab 141 004, India

[3] ICAR-National Rice Research Institute, Cuttack, Odisha 753 006, India

Published online: 16 December 2023

quantitative traits. DH breeding can assist in resolving both of the issues by directly producing homozygous lines in F_2 generation. Complete homozygosity of these DHs lines facilitates the selection of both qualitative and quantitative traits. However, to reach its full potential, crop plants require an efficient haploid induction (HI) system.

Conventionally, haploids were produced by in vitro and in vivo methods. The in vitro methods of haploid production include androgenesis and gynogenesis that rely on culturing the gametophytic cells (n) in various modified tissue culture media. Although being known as a valuable tool, these tissue culture-based methods suffer from various limitations such as genotype dependency, technical complexities, and optimization which makes it a time-consuming, laborious, and costly approach (Kyum et al. 2021). The in vivo-based methods, on the other hand, include the use of haploid inducer stock (HIS), wide hybridization, and centromere-specific variant of histone H3 (*CENH3*)-mediated HI. HIS system is an *in planta* method to produce haploids that bypasses the need for complex tissue culture protocols. Haploid plants are generated by hybridization between HIS and any target line. The haploid seeds thus generated are formed as a result of preferential elimination of the uniparental chromosome set during the early zygotic development (Comai and Tan 2019). DH from these haploids can either be raised by spontaneous chromosome doubling or by using chemical agents such as colchicine. Wide hybridization has only been used in a few genera, such as *Hordeum* and *Triticum*, to produce haploids. However, its commercial application is limited due to labor-intensive, tedious pipelines, and fluctuating HI frequency. The *CENH3*-mediated HI is not well-characterized in crop plants except *Arabidopsis,* and a better understanding is still needed toward the consequences after the manipulation of *CENH3*-linked genes (Kalinowska et al. 2019). It has been seen that manipulation of *CENH3* may result in reduced plant height (Yuan et al. 2015), early embryonic lethality (Ravi and Chan 2010), and error-prone meiotic chromosome segregation (Lermontova et al. 2011). In addition, haploids' sterility is another barrier that makes it a less preferred choice in crop improvement programs. However, such issues can be overcome by adopting a haploid male fertility restorer system. It has been found that inducing mutations in the *parallel spindle 1* (*AtPS1*) gene can help restore the fertility of haploids. This makes the haploid production technology more efficient by avoiding the time-consuming procedure needed for artificial genome doubling (Aboobucker et al. 2023).

Rice and wheat are two major cereal crops feeding almost half of humanity. Many countries in Asia, including China, India, Bangladesh, Indonesia, Thailand, Burma, and Philippines primarily depend on these crops to meet their food demand. The regular and rapid development of varieties in these crops is a continuous demand, and devising tools for their rapid development is highly desirable. Till now, there is a lack of highly efficient HIS in both rice and wheat stalling their way to accelerated breeding. Developing HIS in rice and wheat can benefit our agri-food systems in unprecedented ways. It can supplement the speed breeding of both rice and wheat in direct and indirect ways. It can be employed for the rapid production of homozygous lines, easy genome editing in recalcitrant varieties, and releasing transgene-free genome-edited varieties that can skip regulatory approval easily. Harnessing the benefits from genome editing studies in rice and wheat is impeded by its recalcitrance to tissue culture and regulatory concerns. Low regeneration and transformation efficiency are the two major bottlenecks in genome editing studies (Meyers et al. 2010). Interestingly, HIS can be employed to solve both the issues by utilizing HI-IMGE (Wang et al. 2019a) and HI-Edit (Kelliher et al. 2019) approaches. These innovative techniques focus on combining CRISPR/Cas9 with HIS. In this case, HIS-containing CRISPR/Cas9 cassette of desired gene is crossed with the recalcitrant variety in which editing is supposed to be done. During the zygote formation, transient expression of CRISPR/Cas9 cassette followed by uniparental genome elimination will lead to formation of haploid embryos with desired edits. Due to the elimination of HIS chromosomes from the zygotes, these haploids are considered transgene-free edited haploids. Once developed, HIS can supplement genome editing studies in such types of recalcitrant genotypes to deliver edits in a transgene-free manner. Moreover, HIS has also been used to produce clonal seeds in rice via synthetic apomixis. It has been reported that manipulation of *OsSPO11-1*, *OsREC8*, *OsOSD1,* and *PAIR1* (Khanday et al. 2019; Xie et al. 2019) genes can replace meiosis to produce recombination-free diploid gametes. Such gametes ultimately possess genetic makeup comparable to their somatic cells. To avoid post-fertilization ploidy-related problems due to diploid gametes, a suitable HIS can be employed. Uniparental genome elimination from the zygote will give rise to an embryo that would be exactly similar to one of the parents in its genetic composition. Since such embryos are genetically same to one of the parents they are also referred to as clonal seeds (Wang et al. 2019b).

Apart from its use in DH breeding and gene editing, the generation of haploid plants always remained an interesting tool for plant breeders. They have been previously employed in the classical era for cytogenetic studies of complex polyploid crops such as wheat (Sears 1939) and potatoes (Hougas and Peloquin 1958). Moreover, many potential DHs have been directly released as varieties, such as Marglobe tomato (Cook 1936), Maris Haplona in *Brassica napus* (Thompson 1972), Mingo Barley (Ho and Jones 1980), Satyakrishna and Phalguni in rice (Anonymous 2008, 2010), and Florin in wheat (Buyser et al. 1987). Interestingly, DH has been also used as mapping population to study yield-related traits in

wheat (Pretini et al. 2021) and maize (Shi et al. 2017). DH population can also facilitate mapping in perennial crops. For example, fruit quality traits were studied in apples by using a DH population (Kunihisa et al. 2019). Furthermore, they have been also used for the recovery of recessive mutants during mutation breeding (Watts et al. 2018) and for precise estimation of additive effects (Snape and Simpson 1986).

Genome editing deals with the altering of DNA sequences to achieve the desired phenotype. It can insert, remove, or replace single or many nucleotides at a specific target site. Previously used genome editing tools such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are less suitable due to their obvious complexities (Basu et al. 2023). CRISPR/Cas9 is being utilized on a large scale to perform precise genome editing. Briefly, CRISPR/Cas9 can be used to induce knockout mutations by creating double-stranded breaks (DSBs) at targeted sites. These DSBs can undergo two pathways during their repair process. It can undergo either error-prone nonhomologous-end joining (NHEJ) or homology-directed repair (HDR) pathway (Mushtaq et al. 2021). The use of HDR is limited by its low frequency. NHEJ has been used at a large scale to induce random indels at target sites. Because of its precision and ease of design, CRISPR/Cas is the most favorable tool for genome editing. The use of CRISPR/Cas9 tool for crop improvement has increased unprecedentedly in the last decade as it possesses the immense potential to hasten crop improvement to ensure global food security.

In the last decade, extensive research has been done to uncover the underlying mechanism of HI in maize. Based on different studies, HI is found to be governed by numerous genes. So far, four genes, viz. *MTL/ZmPLA1/NLD* (Liu et al. 2017; Kelliher et al. 2017; Gilles et al. 2017), *ZmDMP* (Zhong et al. 2019), *ZmPLD3* (Li et al. 2021), and *ZmPOD65* (Jiang et al. 2022), have been reported in maize to govern HI ability. These genes have shown haploid induction rate (HIR) ranging from 0.1% to 7.7% when knocked out individually or in different combinations. These haploid-inducing genes are directly and indirectly engaged in processes related to pollen formation and fertilization, for example, *ZmMATL* encodes a patatin-like phospholipase A protein that is localized to the endo-plasma membrane of pollens.*ZmPLD3* is known to establish communication between male and female gametes via pollen tube expansion, and *ZmDMP* works by disrupting double fertilization events. *ZmPOD65* is a pollen-specific peroxidase that can induce haploids by an upsurge in reactive oxygen species causing chromosome fragmentation and HI. It is interesting to note that orthologs of *ZmMATL* have already been knocked out in rice (Yao et al. 2018) and wheat (Liu et al. 2020) to generate haploids. In this manuscript, we have identified potential orthologs of the other three genes in rice and wheat using various in silico approaches. These genes can be knocked out using CRISPR/Cas9-based genome editing for developing HIS in both wheat and rice.

Materials and methods

Sequence search and identification of putative orthologs

For the identification of putative orthologs, a two-pronged strategy was used. In the first case, the amino acid sequences of *ZmPLA, ZmDMP, ZmPLD3,* and *ZmPOD65* were retrieved from the maize genome database (maizegdb, https://www.maizegdb.org/). These sequences were then used as a query for performing BLAST analysis in Phytozome (https://phytozome-next.jgi.doe.gov/) for rice (*Oryza sativa* var. Kitaake) and ensemble plants (https://plants.ensembl.org/index.html) for wheat, respectively. Sequence similarity can be an important strategy to predict gene functions, and it has been shown that genes having sequence similarity of more than 40% may share the same function (Pearson 2013). So, loci showing similarity > 40% were selected for further analysis. In the second strategy, the genes selected on the basis of sequence similarity were subjected to domain architecture analysis. Domain prediction was done using the proteins family database (Pfam, http://pfam.xfam.org/), and their respective Pfam IDs were retrieved. Pfam is based on Hidden Markov model (HMMER3.0), and it is a widely used database to categorize protein sequences into families and domains (Mistry et al. 2021). Rice and wheat genes having domain architecture and Pfam IDs similar to *ZmPLA, ZmDMP, ZmPLD3,* and *ZmPOD65* were used for further analysis. Other sequences and genes having incomplete, redundant, or dissimilar sequences were filtered out.

Naming of PLA, PLD, DMP, and POD genes

Filtered genes were denoted in a standard manner. The first two letters depict the name of the species such as "Os" stands for *Oryza sativa* and "Ta" stands for *Triticum aestivum.* It is then followed by the respective family names of genes as shown in Table 1. In the case of wheat, genes were named based on their genomic and chromosomal locations. Moreover, owing to its genomic complexity, wheat genes were categorized into different homoeologs and were classified as dyads, triads, and tetrads using Ensembl plants. For example, *TaPLA8-4A* is a gene present on the fourth chromosome of the A genome, and it belongs to the Phospholipase A family. *TaPLA8-4A*, *TaPLA8-4B,* and *TaPLA8-4D* are the homoeologs of the gene *TaPLA8.*

Table 1 List of the genes used in this study along with their associated domains, motifs, and Pfam ID in rice and wheat

S. No	Genes	Family associated	Associated domains and motifs	Pfam ID
1	DMP	Domain membrane protein	Protein of unknown function (DUF679)	PF05078
2	PLA	Phospholipase A	Patatin-like phospholipase	PF01734
3	PLD	Phospholipase D	Phospholipase D C terminal	PF12357[a]
			Phospholipase D active site motif	PF00614
			C2 domain	PF00168
4	POD	Peroxidase	Heme peroxidase	PF00141

[a]The family is found in association with PF00168 and PF00614

Phylogenetic analysis

The ancestral relationships of rice and wheat genes with *ZmPLA*, *ZmDMP*, *ZmPLD3,* and *ZmPOD65* were unveiled using phylogenetic analysis. For this analysis, amino acid sequences of filtered genes of both rice and wheat were downloaded from Phytozome and Ensembl plants, respectively. Later on, these sequences were subjected to multiple sequence alignments with the help of CLUSTALW in MEGA11: Molecular Evolutionary Genetics Analysis version 11 (Tamura et al. 2021). CLUSTALW is a widely used method that is based on progressive alignment methods for multiple protein sequence alignments. Notably, the alignments for each of the four genes were carried out separately. Phylogenetic analysis was done using the maximum likelihood method with 1000 bootstraps and using the Poisson substitution model. Lastly, this phylogenetic tree was graphically visualized using the Interactive Tree of Life (https://itol.embl.de/).

In silico expression analysis and prediction of transmembrane (TM) domains

The filtered genes were subjected to in silico expression analysis. Expression data for different genes of rice and wheat were retrieved from the rice expression database (http://expression.ic4r.org/) and wheat expression browser (http://www.wheat-expression.com/), respectively. For rice, genes were analyzed for their expression at four different stages, viz. anther, panicle, leaf, and shoot using SRP047482 and SRP039045 accession ID, and the expression data for all these tissues were downloaded in the form of FPKM (fragments per kilobase of transcript per million mapped reads) values. Expression analysis of wheat genes was performed for the developmental time course of Chinese Spring and Azhurnaya. Akin to rice, the expression levels of wheat genes were analyzed for tissues derived from anther, leaf, shoot, and spike, and the data for these tissues were downloaded in the form of TPM (transcript per million) values. Later on, these data for both wheat and rice genes were comparatively analyzed by visualizing through a heat map which was made with the help of R-Studio using the pheatmap package (Kolde 2012). It was previously known that DMP is a transmembrane protein. Besides in silico expression analysis and protein sequence similarity, TM helixes can be a characteristic feature to identify HI genes based on already identified genes. A number of TM helices have been previously used to identify and validate DMP genes in *Arabidopsis* (Zhong et al. 2020). For further analysis of rice and wheat DMPs, their TM helices were analyzed using the Protter (Omasits et al. 2014), an open-source online tool (https://wlab.ethz.ch/protter/start/). It visualizes the TM helices predicted from the Phobius (Madeira et al. 2022) Web server (https://phobius.sbc.su.se/).

Results

Phospholipase A (PLA) family genes

ZmPLA1/*ZmMTL*/*ZmNLD* of maize belongs to phospholipase A family of phospholipases. This gene has been functionally validated for its role in HI in maize (Liu et al. 2017; Kelliher et al. 2017; Gilles et al. 2017) and its HIR ranges from 2 to 6.7%. The HI ability of *ZmPLA1* mutants is assumed to be caused by chromosome fragmentation and genomic instability, while the exact process is still unknown. For identifying *ZmPLA1* orthologs in rice and wheat, we retrieved rice and wheat candidate genes based on their protein domain and sequence similarity to maize *ZmPLA1*. We were able to get 12 and 18 PLA genes in rice and wheat, respectively. Phylogenetic analysis of these genes with *ZmPLA1* revealed that rice *OsPLA4* and three homoeologs of wheat *TaPLA8,* viz. *TaPLA8-4A*, *TaPLA8-4B,* and *TaPLA8-4D* share the same subclade, indicating their close ancestral relationship with *ZmPLA1* (Supplementary Fig. S1). Further, in silico expression analysis of the rice genes revealed that three genes, namely *OsPLA4*, *OsPLA7*, and *OsPLA9,* showed anther-specific expression with low or negligible expression in the shoot, leaf, and panicle (Supplementary Fig. S2). Similarly, *TaPLA3*, *TaPLA8,* and *TaPLA13-7B* of wheat also showed anther-specific expression with minimal or low expression in other comparative tissues (Supplementary Fig. S3). Interestingly, in both rice and wheat, the genes

that were closely related with *ZmPLA1,* i.e., *OsPLA4* and *TaPLA8,* showed the highest anther-specific expression. So, considering the close ancestral relationship of *TaPLA8* and *OsPLA4* with *ZmPLA* along with anther-specific expression, it can be inferred that they are the two closest orthologs of *ZmPLA* in wheat and rice. Although the role of these genes in HI has already been demonstrated in rice (Yao et al. 2018) and wheat (Liu et al. 2020) by CRISPR/Cas9-based genome editing, this indicates that identifying genes on the basis of phylogenetic and in silico expression analysis can be a sound approach for choosing target genes for genome editing. A similar approach to identifying genes prior to their editing has also been used by Zhong et al. (2022).

Phospholipase D (PLD) family genes

ZmPLD3 has been recently identified for its HI role in maize (Li et al. 2021). It belongs to the phospholipase D family of phospholipases. *ZmPLD3* protein is known to contain two HxKxxxD motifs and one C2 domain that are considered essential for PLD activity (Wang 2001). Its loss of function mutation has shown varied HIR (1–7%) in combination with *ZmDMP* and *ZmPLA1*. Interestingly, the knockout of *ZmPLD3* along with *ZmPLA1* increased its HIR by three times (Li et al. 2021). The orthologs in rice and wheat were identified based on sequence similarity to *ZmPLD3* (supplementary data 1) and functional domain analysis. Thirteen genes in rice and seventeen genes in wheat were filtered for further analysis. Phylogenetic analysis revealed that *OsPLD1* and *TaPLD1-1A, TaPLD1-1B,* and *TaPLD1-1D* were present in the same subclade with *ZmPLD3* (Fig. 1A) showing their close relationship with the *ZmPLD3*. Further, in silico expression analysis of these genes showed that *OsPLD1* and *OsPLD8* have anther-specific expression, whereas *OsPLD1* has the highest level of expression and *OsPLD8* follows closely behind (Fig. 1B) (Supplementary data 2). In the case of wheat, *TaPLD1, TaPLD9-5A, TaPLD9-5D, TaPLD14-4A*, and *TaPLD15* showed anther-specific expression and the highest being shown by *TaPLD1* (Fig. 1C). Notably, in case of *TaPLD9*, only a pair of the homoeologs, viz. *TaPLD9-5A* and *TaPLD9-5D,* showed anther-specific expression, whereas the third homoeolog *TaPLD9-5B* did not show any kind of specific expression. Furthermore, similar to phospholipase A (PLA) family genes, the genes that were closely associated with *ZmPLD3* phylogenetically, *i.e., OsPLD1* and *TaPLD1*, possess the highest anther-specific expression. Overall, these results indicate that *OsPLD1* and *TaPLD1-1A, TaPLD1-1B*, *and TaPLD1-1D* are the potential HIS development targets in both rice and wheat. Notably, the phospholipase D gene family contains a number of genes. These genes may have similarities with each other. So, care should be taken while designing the experiment to avoid any off-targets via careful domain selection and sgRNAs designing.

Domain membrane protein (DMP) family genes

The key characteristic of domain membrane proteins (DMP) is the presence of transmembrane (TM) helices. The *ZmDMP* gene belongs to this family and is involved in HI. Single *ZmDMP* mutant has extremely low HIR (0.1 to 0.3%), but when *ZmDMP* and *ZmPLA1* were simultaneously altered, it increased by fivefold to sixfold (Zhong et al. 2019). It is assumed that *ZmDMP* might induce haploids by impairing double fertilization and increasing single fertilization events. Studies have shown that one sperm cell of DMP mutant pollen preferably fertilizes the center cell, while the other sperm cell does not fuse and frequently remains attached to the egg cell giving rise to haploid embryos in *Arabidopsis* (Zhong et al. 2020). Later on, for identifying DMP family genes, we filtered five and thirty-three genes in rice and wheat, respectively (Supplementary data 1). Phylogenetic analysis showed that *OsDMP1* and *TaDMP13-7A, TaDMP13-7B,* and *TaDMP13-7D* were very close to *ZmDMP* (Fig. 2A). But when the in silico expression levels were observed for these genes, we found a different pattern. In contrast to PLAs and PLDs, the closest DMP gene (*OsDMP1*) has not shown anther-specific expression. However, the other DMP gene (*OsDMP2*) which has shown anther-specific expression did not share any close subclade with *ZmDMP* (Fig. 2B). Moreover, it was observed that *OsDMP2* has shown much higher anther-specific expression as compared to both PLA and PLD genes (Table 2). Furthermore, the in silico expression analysis of wheat DMPs also led to similar results. *In silico* expression analysis revealed that *TaDMP3*, *TaDMP8-6A*, and *TaDMP8-6B* were showing anther-specific expression (Fig. 2C). As previously mentioned, phylogenetically *TaDMP13* was the closest to *ZmDMP*, but it was not specifically expressing in anthers. It was observed that *TaDMP13-7A* and *TaDMP13-7D* were also expressed in the stem (Supplementary data 2). Similar to rice DMPs, wheat DMPs that were closest to *ZmDMP* did not have anther-specific expression and those having higher anther-specific expression did not share any clade with *ZmDMP*. This inherent similarity for DMP genes in both wheat and rice may be pointing out any important evolutionary process.

To remove the ambiguity regarding the selection of DMPs in both crops, we further analyzed the TM helices for the DMP genes. It is previously known that DMP proteins generally have TM helices (Zhu et al. 2021). The number of TM helices in rice ranged from 2 to 4, and we found that only *OsDMP1* and *OsDMP2* was having four TM helices (Supplementary data 3). However, on observing the predicted structure of TM helices, we found that among *OsDMP1* and *OsDMP2*, *OsDMP1* had

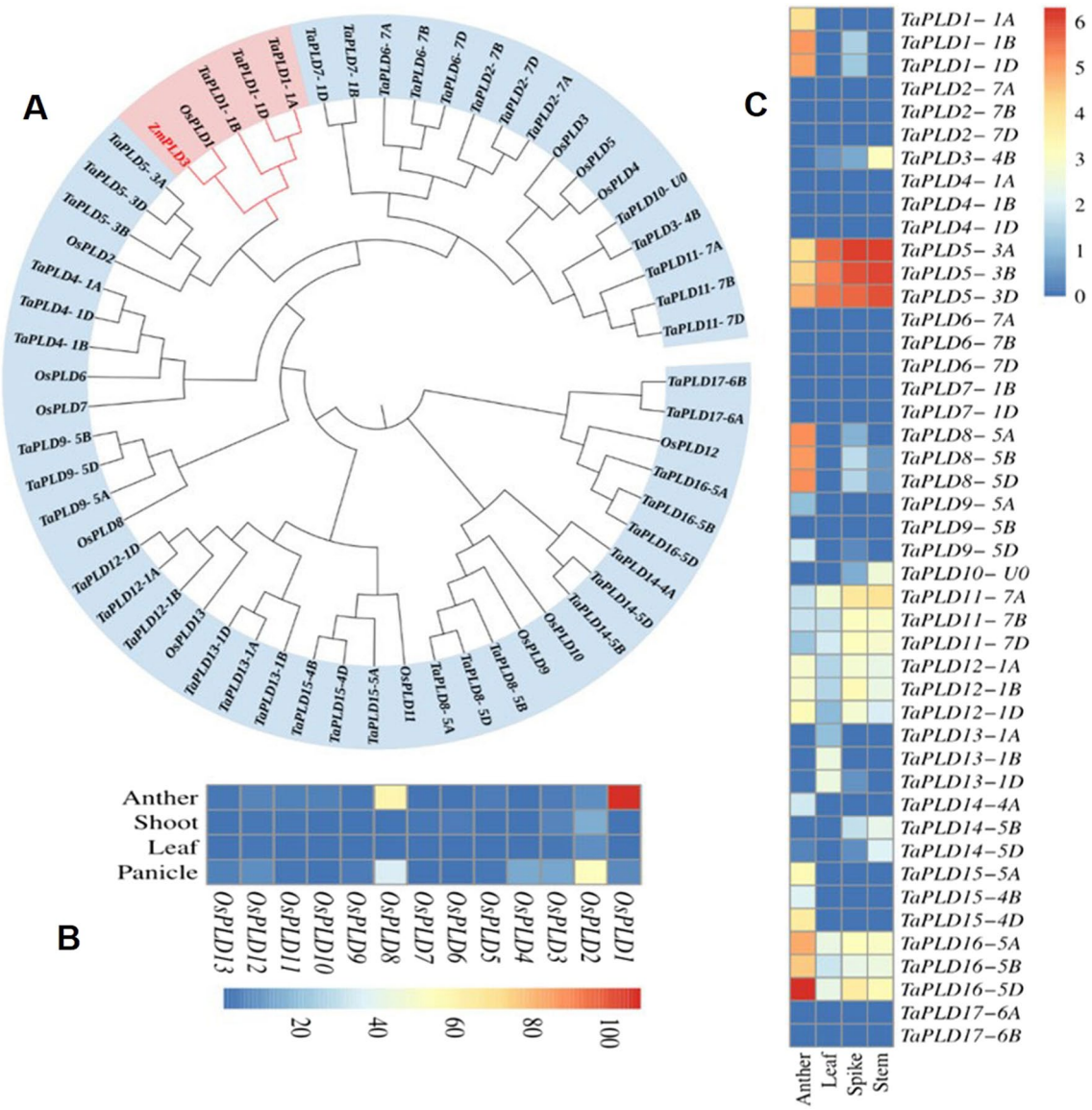

Fig. 1 **A** Phylogenetic tree of phospholipase D (PLD) genes in both rice and wheat and its comparison with maize *ZmPLD3*. *OsPLD1, TaPLD1-1A, TaPLD1-1B,* and *TaPLD1-1D* genes are present in the same subclade with *ZmPLD3* with the pink background showing that these genes are ancestrally more related to *ZmPLD3* gene. **B** Heat plot depicting in silico expression analysis of phospholipase D (PLD) genes in rice. *OsPLD1* and *OsPLD8* genes have shown anther-specific expression with little expression in panicle. **C** Heat plot depicting in silico expression analysis of phospholipase D (PLD) genes in wheat. *TaPLD1, TaPLD9-5A, TaPLD9-5D, TaPLD14-4A,* and *TaPLD15* have shown the anther-specific expression. *OsPLD1, TaPLD1-1A, TaPLD1-1B,* and *TaPLD1-1D* have shown the highest anther-specific expressions along with close ancestral relationship with *ZmPLD3*

a comparable structure to *ZmDMP*. It can be seen that the long intracellular loop present between the second and third TM helix in *ZmDMP* is similarly present in *OsDMP1* and is small in *OsDMP2* (Fig. 3A). This indicates that in terms of TM helices, *OsDMP1* protein has more structural similarity to *ZmDMP* as compared to *OsDMP2*, whereas in wheat, we observed that the number of TM helices ranged from 2 to 8. Akin to rice, the genes that had anther-specific expression or ancestral relationship with *ZmDMP* had four helices, but the orientation of TM helices was quite different among them. It was observed that similar to *OsDMP1*, *TaDMP13* had that particular long intracellular loop present between the second and third TM helix (Fig. 3B), whereas, in *TaDMP8* and *TaDMP3* (Fig. 3C, D), a comparatively smaller loop was present between the helices, pointing out more structural conservation of *TaDMP13* as compared to *TaDMP3* and *TaDMP8*. Interestingly, the TM helices of *OsDMP2* and *TaDMP3* showed similarity, and it can also be noted that both of them also show anther-specific expression and

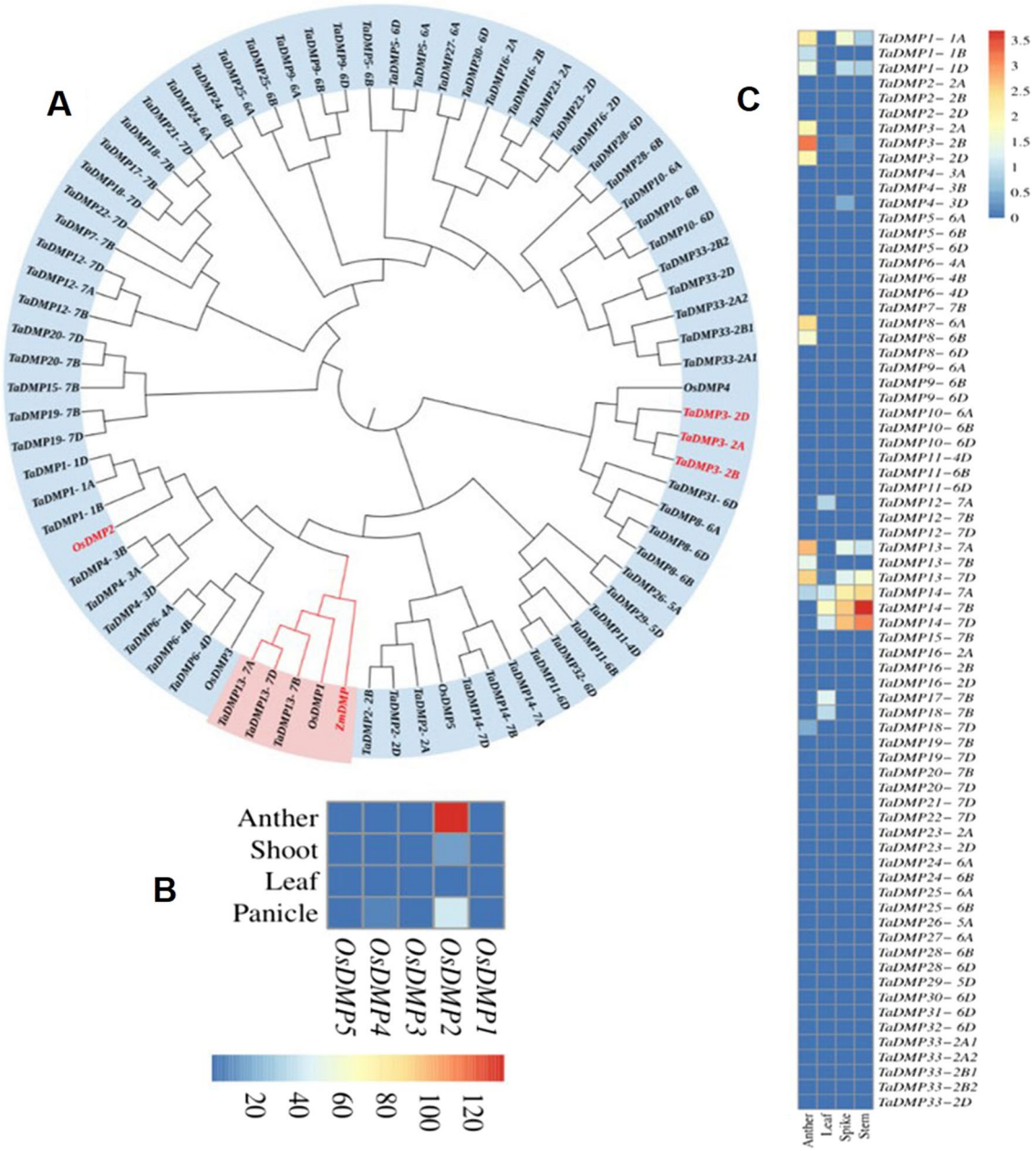

Fig. 2 **A** Phylogenetic tree of domain membrane protein (DMP) genes in both rice and wheat and its comparison with maize *ZmDMP*. *OsDMP1, TaDMP13-7A, TaDMP13-7B,* and *TaDMP13-7D* genes are present on same subclade with *ZmDMP* pointing out their close phylogenetic relationship with *ZmDMP*, whereas *OsDMP2, TaDMP3-2A, TaDMP3-2B,* and *TaDMP3-2D* are not situated close to *ZmDMP* depicting their ancestral diversity with *ZmDMP* gene. **B** Heat plot depicting in silico expression analysis of domain membrane protein (DMP) genes in rice. *OsDMP2* has shown the highest and specific expression in anther as compared to other genes. It has also shown higher expression in panicle as compared to other rice genes. **C** Heat plot depicting in silico expression analysis of domain membrane protein (DMP) genes in wheat. *TaDMP13-7A, TaDMP13-7B,* and *TaDMP13-7D* genes are expressing in all tissues except leaf, showing no anther-specific expression, whereas *TaDMP3-2A, TaDMP3-2B,* and *TaDMP3-2D* are specifically expressing in anthers

less ancestral relationship in their respective crops. So, it can be assumed that this specific expression can be attributed to the orientation of TM helices of these both genes. Though they are different from TM helices of *ZmDMP,* the inherent association between TM helix orientation and

Table 2 List of the genes which have been studied in rice and wheat along with their reference ID, chromosome number, in silico expression data, and number of transmembrane (TM) helices

S. No	Gene ID	IWGSC ref. no/MSU IDs	Chromosome	Expression values (FPKM/TPM)				No. of TM helices
				Anther	Shoot	Leaf	Panicle/spike	
1	*OsPLA4*	LOC_Os03g27610	3	23.14	0.06	0.02	3.29	NA
2	*OsPLA7*	LOC_Os08g37250	8	1.74	0.20	0.10	0.21	NA
3	*OsPLA9*	LOC_Os08g37210	8	4.41	0.02	0.00	0.00	NA
4	*TaPLA3-2A*	TraesCS2A02G159600	2A	1.69	0.00	0.00	0.00	NA
5	*TaPLA3-2B*	TraesCS2B02G185200	2B	1.83	0.00	0.00	0. 27	NA
6	*TaPLA3-2D*	TraesCS2D02G166800	2D	1.58	0.00	0.00	0.00	NA
7	*TaPLA8-4A*	TraesCS4A02G018100	4A	2.46	0.00	0.00	0.00	NA
8	*TaPLA8-4B*	TraesCS4B02G286000	4B	2.02	0.00	0.00	0.00	NA
9	*TaPLA8-4D*	TraesCS4D02G284700	4D	1.77	0.00	0.00	0.00	NA
10	*TaPLA13-7B*	TraesCS7B02G411300	7B	0.32	0.00	0.00	0.00	NA
11	*OsDMP1*	LOC_Os08g01530	8	0.31	0.87	0.08	0.95	4
12	*OsDMP2*	LOC_Os05g48840	5	134.15	14.48	0.36	40.43	4
13	*TaDMP3-2A*	TraesCS2A02G153900	2A	2.02	0.00	0.00	0.00	4
14	*TaDMP3-2B*	TraesCS2B02G179200	2B	3.19	0.00	0.00	0.22	4
15	*TaDMP3-2D*	TraesCS2D02G159400	2D	2.01	0.00	0.00	0.00	4
16	*TaDMP8-6A*	TraesCS6A02G123100	6A	2.44	0.00	0.00	0.00	4
17	*TaDMP8-6B*	TraesCS6B02G151400	6B	1.67	0.00	0.00	0.00	4
18	*OsPLD1*	LOC_Os05g07880	5	108.06	1.05	0.05	5.60	NA
19	*OsPLD8*	LOC_Os09g25390	9	59.38	0.17	0.00	35.07	NA
20	*TaPLD1-1A*	TraesCS1A02G115300	1A	4.08	0.00	0.00	0.00	NA
21	*TaPLD1-1B*	TraesCS1B02G135200	1B	5.08	0.00	0.00	1.40	NA
22	*TaPLD1-1D*	TraesCS1D02G116700	1D	4.95	0.00	0.00	1.30	NA
23	*TaPLD9-5A*	TraesCS5A02G223500	5A	1.09	0.00	0.00	0.00	NA
24	*TaPLD9-5D*	TraesCS5D02G231100	5D	1.92	0.00	0.00	0.37	NA
25	*TaPLD14-4A*	TraesCS4A02G347800	4A	1.93	0.00	0.00	0.00	NA
26	*TaPLD15-5A*	TraesCS5A02G520900	5A	3.30	0.00	0.00	0.00	NA
27	*TaPLD15-4B*	TraesCS4B02G352300	4B	2.18	0.00	0.00	0.00	NA
28	*TaPLD15-4D*	TraesCS4D02G346700	4D	3.73	0.00	0.00	0.00	NA
29	*OsPOD6*	LOC_Os02g50770	2	290.25	0.09	0.03	25.20	NA
30	*OsPOD8*	LOC_Os03g02920	3	40.57	6.38	0.63	50.21	NA
31	*OsPOD19*	LOC_Os06g13050	6	11.52	0.74	0.70	5.31	NA
32	*TaPOD8-6A*	TraesCS6A02G309900	6A	3.86	0.00	0.00	2.09	NA
33	*TaPOD8-6B*	TraesCS6B02G339700	6B	4.75	0.00	0.00	2.17	NA
34	*TaPOD8-6D*	TraesCS6D02G289300	6D	3.83	0.00	0.00	1.34	NA
35	*TaPOD19-U0*	TraesCSU02G096600	Unknown	1.00	0.00	0.00	0.00	NA

anther-specific expression of *OsDMP2* and *TaDMP3* cannot be avoided.

Peroxidase (POD) genes

Reactive oxygen species (ROS) are known to play a vital role during plant growth, metabolism, senescence, and biotic and abiotic stresses (Mhamdi and Breusegem 2018). It is reported that an increase in ROS in sperm cells could lead to sperm DNA fragmentation and HI. It has been recently demonstrated that genes encoding for peroxidases can also be manipulated to induce haploids in maize with an HIR of up to 7.7% (Jiang et al. 2022). Therefore, we identified putative orthologs of *ZmPOD65* in rice and wheat. Based on BLAST and domain architecture analysis, thirty-eight and twenty-two POD genes were identified in rice and wheat, respectively (Supplementary data 1). Phylogeny analysis with *ZmPOD65* revealed that *OsPOD6, OsPOD19, TaPOD8-6A*, *TaPOD8-6B*, and *TaPOD8-6D* share the same subclade with *ZmPOD65* (Fig. 4A).

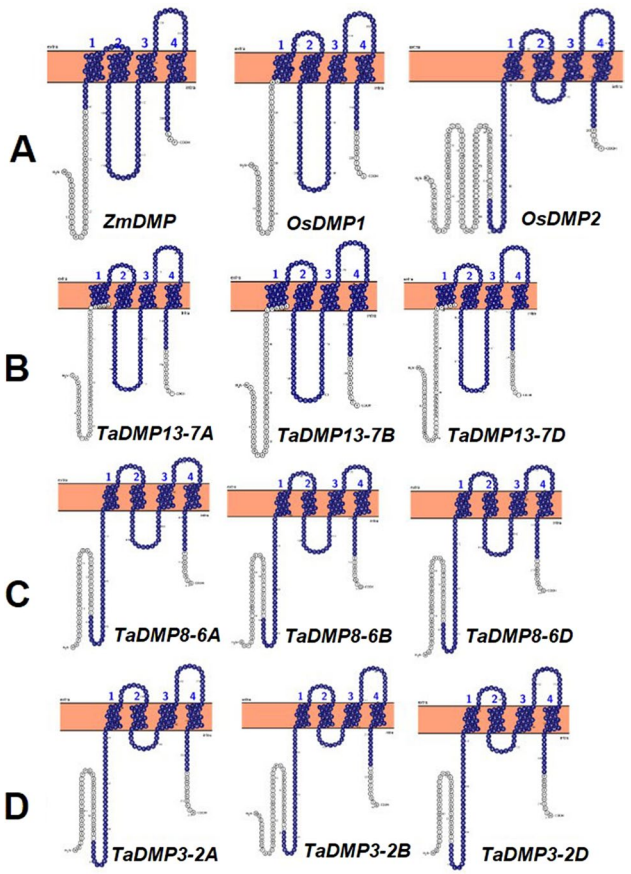

Fig. 3 Prediction of transmembrane (TM) helices in rice and wheat DMP genes and their comparison with maize *ZmDMP*. The numbers in the blue color depicts the number of transmembrane helices. The blue-circled amino acids represent the conserved amino acids of protein. **A** TM helices of maize and rice DMP genes. *OsDMP1* and *OsDMP2* genes have shown four TM helices analogs to maize. As compared to *OsDMP2* gene, orientation of *OsDMP1* TM helices is more similar to maize. **B–D** TM helices of wheat DMP genes. Among the showed genes, *TaDMP13-7A, TaDMP13-7B,* and *TaDMP13-7D* are more similar to *ZmDMP* gene in terms of TM helices. Homoeologs of *TaDMP13* have a long intracellular loop present between the second and third TM helix which is similar to the loop present between second and third TM helix of maize DMP gene

Further, in silico expression analysis in rice has shown that *OsPOD6, OsPOD8,* and *OsPOD19* possess higher anther-specific expression (Fig. 4B), whereas in wheat, *TaPOD8-6A*, *TaPOD8-6B, TaPOD8-6D,* and *TaPOD19-U0* were expressing specifically in anther (Fig. 4C) (Supplementary data 2). Among these genes, *OsPOD6* and *TaPOD8* showed the highest expression in anthers along with higher similarity to *ZmPOD65*. Based on the above analysis, these two genes (*OsPOD6* and *TaPOD8*) can be considered as the putative targets for inducing haploids via manipulating peroxidase genes.

Discussion

DH is an important tool with plant breeders for accelerating crop improvement programs. An ideal DH production system should be able to produce a multitude of DH without much effort. The development of HIS particularly in rice and wheat could be an important advance in this regard. Among the important cereal crops, maize is naturally endowed with a HIS. It was initially found by Coe (1959) and then it was refined later with intense breeding efforts including hybridization, mapping, and functional validation studies. The HIS in maize is giving a HI frequency of > 10% making it an ideal system among all cereal crops. Recent scientific efforts have identified four major genes controlling the HI ability of maize including *ZmPLA1*, *ZmDMP*, *ZmPLD3*, and *ZmPOD65*. Orthologs of *ZmPLA1* have already shown relatively higher HI frequency in rice (Yao et al. 2018) and wheat (Liu et al. 2020). Such investigations affirm the conservation of HI genes across the taxa. So, intending to expand its utility, we have identified several genes in both rice and wheat to develop HIS. These genes can be further functionally validated using genome editing to reveal the genetic mechanism underlying the HI trait.

It has been seen that *OsPLA4* and *TaPLA8* are the closest orthologs of *ZmPLA1* based on our analysis. As already discussed, they have been knocked out in both rice (Yao et al. 2018) and wheat (Liu et al. 2020) using CRISPR/Cas-based genome editing to produce haploids in both crops. Using a similar approach, we have identified potential orthologs of maize *ZmDMP*, *ZmPLD3*, and *ZmPOD65* in rice and wheat. Being highly conserved and having anther-specific expression, *OsPLD1* and *TaPLD1* (phospholipase D family) and *OsPOD6* and *TaPOD8* (peroxidase family) genes seem to be putative orthologs for HI. In some cases, only two homoeologs of wheat have shown specific expressions, while the third homoeologs exhibited no expression specificity. Such discrepancies are not new for wheat as it is also reported by lv et al. (2020) for the *CENH3* gene, which is widely studied for HI traits including maize (Wang et al. 2021) and wheat (Lv et al. 2020).

Interestingly, in the case of *DMP*s, the relationship between phylogeny and in silico expression was not in harmony. The conservation of *ZmDMP* across the monocots has been reported by several authors. Based on sequence similarity, *ZmDMP* is found to be highly conserved in several monocots such as *Musa acuminata* (66.51%), *Setaria italica* (73.53%), and *sorghum* (91.04%) (Zhong et al. 2019). But to the best of our knowledge, its conservation in rice and wheat has been not explored so far. Additionally, DMP has been confirmed for the HI trait in several

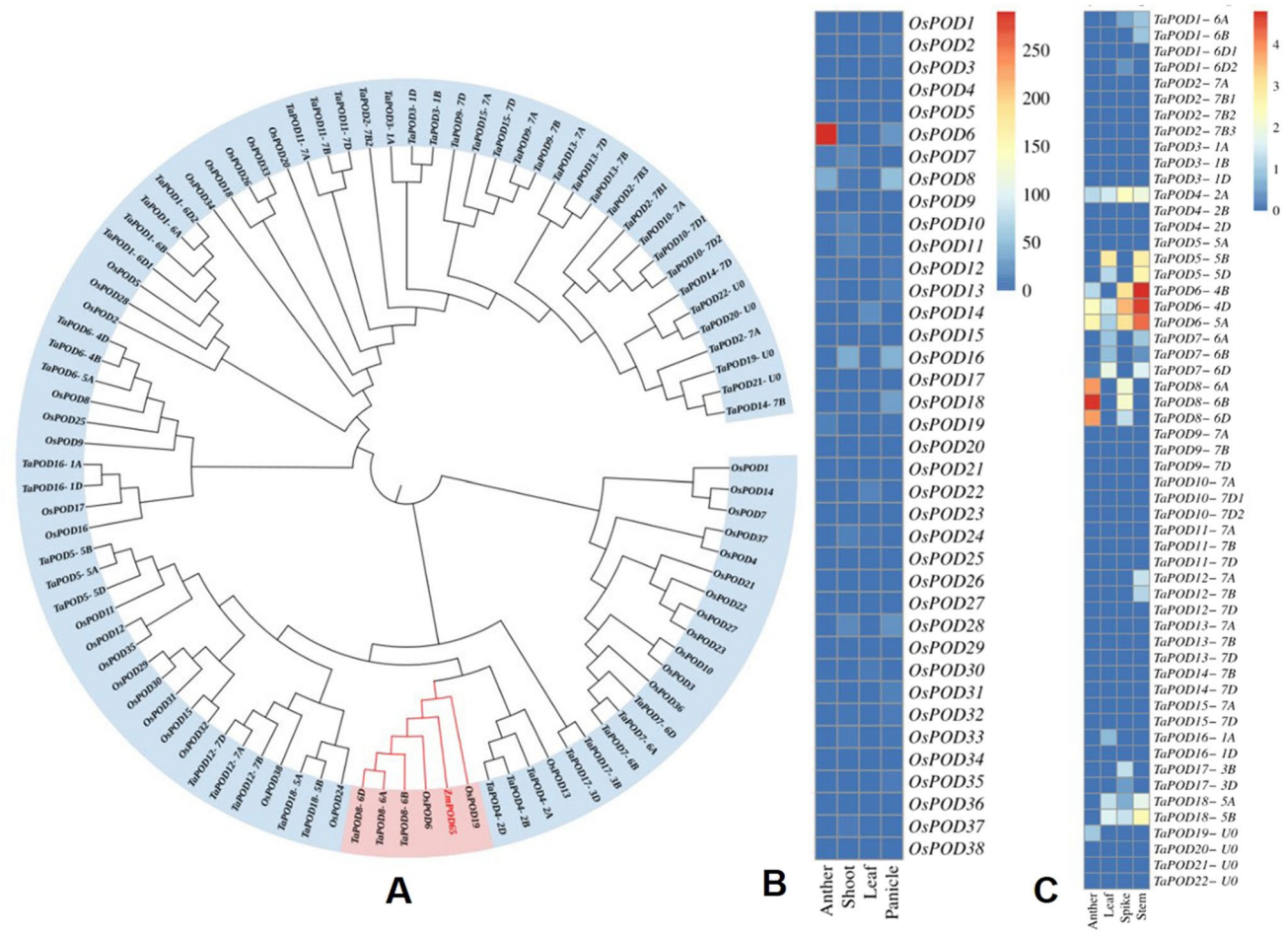

Fig. 4 **A** Phylogenetic tree of peroxidase genes (POD) in both rice and wheat and its comparison with maize *ZmPOD65*. *OsPOD6, OsPOD19, TaPOD8-6A, TaPOD8-6B, and TaPOD8-6D* are sharing same subclade with *ZmPOD65* depicting their close phylogenetic relationship with *ZmPOD65*. **B** Heat plot depicting in silico expression analysis of peroxidase genes (POD) genes in rice. *OsPOD6, OsPOD8,* and *OsPOD19* have shown anther-specific expression while *OsPOD6* showing the highest. **(C)** Heat plot depicting in silico expression analysis of peroxidase genes (POD) genes in wheat. *TaPOD8-6A*, *TaPOD8-6B, TaPOD8-6D,* and *TaPOD19-U0* are specifically expressing in anther. Homoeologs of *TaPOD8* gene have shown highest anther-specific expression

dicot crops; however, the number of genes examined varies considerably. It has been seen that one gene in tomato (Zhong et al. 2022), two genes in *Medicago* (Wang et al. 2022b) and *Arabidopsis* (Zhong et al. 2020), and three genes in Brassica and Tobacco (Li et al. 2022) have been already knocked out to generate haploids. Akin to this, it has been predicted that cotton can also have two copies of DMP genes (Zhu et al. 2021), which can be knocked out for inducing haploids. Conclusively, it can be inferred that there are higher chances that the rice and wheat may similarly possess more than one DMP gene for the induction of haploids. For example, *OsDMP1* and *OsDMP2* in rice and *TaDMP13* and *TaDMP3* in wheat can act as HIS-responsible genes. However, their validation through genome editing will unveil their mechanism appropriately.

A higher HIR is one of the desirable features of HIS. Several strategies can be employed for increasing the HIR via conventional and modern ways. One way of increasing HIR is by phenotypic selection for variable heading dates. It has been observed that a HIS panel with variable heading dates can provide an HIR of up to 12% in the case of rice (Wang et al. 2022a). However, HIR can also be increased via manipulation at the molecular level. For increasing HIR at molecular level, finding variations that behave synergistically is important. Synergistically acting HI genes can be effectively stacked to produce a line with significantly high HIR. As in the case of maize, a triple mutant of *zmpla1* + *zmdmp* + *zmpld3* exhibited significantly higher HIR as compared to all other combinations (Li et al. 2021). Pyramiding these three mutations in a superior background of rice and wheat could provide a more efficient HIS. However, these three mutations have not been investigated in conjunction with *ZmPOD65* yet. Hypothetically, editing of *ZmPOD65* along with these three alleles may provide an

even higher HIR. But increased HIR brings its limitations. High HIR comes at the expense of increased kernel abortion (10 to 47.6%) (Kelliher et al. 2017). This type of abortion is persistent in most of the studies of HI, except a few dicots. Moreover, it is even suggested that HI and kernel abortion phenomenon are linked and one can use kernel abortion as a marker for HI. Therefore, it is possible to pyramid all of these mutations, but doing so will increase the chance of embryo abortion.

After several functional validation studies, it has been firmly established that HIS is controlled by distinct genes. Moreover, different mutant alleles generated by genome editing possessed different HIR for *ZmPLA1*, *ZmPLD3*, *OsMATL*, and *TaMATL*. Though in some cases, the difference between the HIR of different alleles is non-significant. Interestingly, HIR can also be affected by the allelic state of HI genes. For example, when combined with *ZmPLA1,* the heterozygous *ZmDMP* was having low HIR as compared to homozygous *ZmDMP*. Similar reports for the different alleles of the same genes having different HIR were also reported for *ZmPOD65* (Jiang et al. 2022). As a result, producing an array of various mutations in predicted genes and screening them for their respective HIR could be a game changer in developing highly efficient HIS in both rice and wheat.

Prior to *MTL*, *CENH3* was widely investigated for producing haploid plants through genetic modifications. Following groundbreaking research in *Arabidopsis* (Ravi and Chan 2010), various changes have been made using transgenes, EMS-induced mutations (Kuppu et al. 2020), and genome editing (Wang and Ouyang 2022) to improve the HIR through *CENH3*-mediated approach. However, to the best of our knowledge, no study has reported an interaction between *ZmPLA1* and *CENH3* mutants. Studying the interaction between both of these genes can provide new insights into the mechanism of HI as the former is capable of inducing maternal haploids and the latter for paternal haploids. Moreover, stacking all three mutants (*zmpla1* + *zmdmp* + *zmpld3*) along with the *CENH3* gene might yield some exciting results. Genome editing appears to be the most ideal technique for exploring all these possibilities in rice and wheat.

Conclusion

The study has identified candidate orthologous genes for developing a HIS in rice and wheat. Using maize HI genes as a reference, several genes have been predicted by employing different in silico approaches. *OsPLD1, OsDMP1*, *OsDMP2,* and *OsPOD6* in rice and *TaPLD1, TaDMP3, TaDMP13,* and *TaPOD8* genes in wheat are predicted to possess similar function to the maize HI genes. The PLD and POD genes have shown direct association between there phylogenetic and expression analysis results, whereas the outcomes of DMP analysis suspect the presence of more than two DMP genes in both of the crops. Practical utilization of approaches like HI-mediated genome editing and synthetic apomixis can be expanded by high-frequency HIS. It can be enhanced by using an integrated approach of conventional and modern scientific techniques. Previous studies have established that genome editing can be used for the purpose of developing HIS in rice and wheat. So, CRISPR/Cas and its variants can be employed for targeting these genes. Additionally, different combinations of these genes can be targeted using multiplexed genome editing to understand the genetic mechanisms of HI in these crops.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-023-03857-9.

Authors contributions DB and LG conceived the idea, designed, and supervised the study; LG, MK, MM, DP, and SK carried out all the in silico analysis; LG wrote the first draft of the manuscript; DB, KAM, and LG edited and wrote the final version of manuscript. All authors read and approved the final manuscript.

Funding The financial support was provided by the Department of Biotechnology, India, under project number BT/GET/119/SP25803/2018 for carrying out this research.

Data availability There is no need of data availability statement since all the available data of publication is present in the main text and supplementary information.

Declarations

Conflict of interest Authors declare that there is no conflict of interest.

References

Aboobucker SI, Zhou L, Lübberstedt T (2023) Haploid male fertility is restored by parallel spindle genes in *Arabidopsis thaliana*. Nat Plants 9:1–5. https://doi.org/10.1038/s41477-022-01332-6

Ani KJ, Anyika VO, Mutambara E (2022) The impact of climate change on food and human security in Nigeria. Int J Clim Chang 14(2):148–167. https://doi.org/10.1108/IJCCSM-11-2020-0119

Anonymous (2008) Central Rice Research Institute (CRRI) Annual Report-2007–2008. National Rice Research Institute, India

Anonymous (2010) Central Rice Research Institute (CRRI) Annual Report-2009–2010. National Rice Research Institute, India

Baranwal V, Sharma SK, Ghosh A, Gupta N, Singh AK, Diksha D, Jangra TP, S (2022) Evidence for association of southern rice black-streaked dwarf virus with the recently emerged stunting disease of rice in North-West India. Indian J Genet Plant Breed 82(04):512–516

Basu U, Ahmed SR, Bhat BA, Anwar Z, Ali A, Ijaz A, Gulzar A, Bibi A, Tyagi A, Nebapure SM, Goud CA (2023) A CRISPR way for accelerating cereal crop improvement: Progress and challenges. Front Genet 13:866976. https://doi.org/10.3389/fgene.2022.866976

Chen Y, Marek GW, Marek TH, Moorhead JE, Heflin KR, Brauer DK, Gowda PH, Srinivasan R (2019) Simulating the impacts of climate

change on hydrology and crop production in the Northern High Plains of Texas using an improved SWAT model. Agric Water Manag 221:13–24. https://doi.org/10.1016/j.agwat.2019.04.021
Coe JEH (1959) A line of maize with high haploid frequency. Am Nat 93(873):381–382. https://doi.org/10.1086/282098
Comai L, Tan EH (2019) HI and genome instability. Trends Genet 35(11):791–803. https://doi.org/10.1016/j.tig.2019.07.005
Cook RC (1936) A haploid Marglobe tomato: practical application of a "short cut" for making pure lines. J Hered 27:433–435. https://doi.org/10.1093/oxfordjournals.jhered.a104153
De Buyser J, Henry Y, Lonnet P, Hertzog R, Hespel A (1987) Florin: a doubled haploid wheat variety developed by the anther culture method. Plant Breed 98(1):53–56. https://doi.org/10.1111/j.1439-0523.1987.tb01089.x
Gilles LM, Khaled A, Laffaire JB, Chaignon S, Gendrot G, Laplaige J, Bergès H, Beydon G, Bayle V, Barret P, Comadran J (2017) Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J 36(6):707–717. https://doi.org/10.15252/embj.201796603
Ho KM, Jones GE (1980) Mingo barley. Can J Plant Sci 60:279–280. https://doi.org/10.4141/cjps80-041
Hougas RW, Peloquin SJ (1958) The potential of potato haploids in breeding and genetic research. Am Potato J 35(10):701–707. https://doi.org/10.1007/BF02855564
Jiang C, Sun J, Li R, Yan S, Chen W, Guo L, Qin G, Wang P, Luo C, Huang W, Zhang Q (2022) A reactive oxygen species burst causes haploid induction in maize. Mol Plant 15(6):943–955. https://doi.org/10.1016/j.molp.2022.04.001
Kalinowska K, Chamas S, Unkel K, Demidov D, Lermontova I, Dresselhaus T, Kumlehn J, Dunemann F, Houben A (2019) State-of-the-art and novel developments of in vivo haploid technologies. Theor Appl Genet 132:593–605. https://doi.org/10.1007/s00122-018-3261-9
Kang X, Qi J, Li S, Meng FR (2022) A watershed-scale assessment of climate change impacts on crop yields in Atlantic Canada. Agric Water Manag 269:107680. https://doi.org/10.1016/j.agwat.2022.107680
Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, Green J, Chen Z, McCuiston J, Wang W, Liebler T (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542(7639):105–109. https://doi.org/10.1038/nature20827
Kelliher T, Starr D, Su X, Tang G, Chen Z, Carter J, Wittich PE, Dong S, Green J, Burch E, McCuiston J (2019) One-step genome editing of elite crop germplasm during haploid induction. Nat Biotechnol 37(3):287–292. https://doi.org/10.1038/s41587-019-0038-x
Khanday I, Skinner D, Yang B, Mercier R, Sundaresan V (2019) A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. Nature 565(7737):91–95. https://doi.org/10.1038/s41586-018-0785-8
Kolde R (2012) Pheatmap: pretty heatmaps. R Package Version 1.0.10. https://CRAN.R-project.org/package=pheatmap
Kunihisa M, Takita Y, Yamaguchi N, Okada H, Sato M, Komori S, Nishitani C, Terakami S, Yamamoto T (2019) The use of a fertile doubled haploid apple line for QTL analysis of fruit traits. Breed Sci 69(3):410–419. https://doi.org/10.1270/jsbbs.18197
Kuppu S, Ron M, Marimuthu MP, Li G, Huddleson A, Siddeek MH, Terry J, Buchner R, Shabek N, Comai L, Britt AB (2020) A variety of changes, including CRISPR/Cas9-mediated deletions, in *CENH3* lead to haploid induction on outcrossing. Plant Biotechnol J 18(10):2068–2080. https://doi.org/10.1111/pbi.13365
Kyum M, Kaur H, Kamboj A, Goyal L, Bhatia D (2021) Strategies and prospects of haploid induction in rice (*Oryza sativa*). Plant Breed 141(1):1–11
Lermontova I, Koroleva O, Rutten T, Fuchs J, Schubert V, Moraes I, Koszegi D, Schubert I (2011) Knockdown of *CENH3* in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. Plant J 68(1):40–50. https://doi.org/10.1111/j.1365-313X.2011.04664.x
Li Y, Lin Z, Yue Y, Zhao H, Fei X, Liu C, Chen S, Lai J, Song W (2021) Loss-of-function alleles of *ZmPLD3* cause haploid induction in maize. Nat Plants 7(12):1579–1588. https://doi.org/10.1038/s41477-021-01037-2
Li Y, Li D, Xiao Q, Wang H, Wen J, Tu J, Shen J, Fu T, Yi B (2022) An in planta haploid induction system in *Brassica napus*. J Integr Plant Biol 64(6):1140–1144. https://doi.org/10.1111/jipb.13270
Liu C, Li X, Meng D, Zhong Y, Chen C, Dong X, Xu X, Chen B, Li W, Li L, Tian X (2017) A 4-bp insertion at *ZmPLA1* encoding a putative phospholipase A generates haploid induction in maize. Mol Plant 10(3):520–522. https://doi.org/10.1016/j.molp.2017.01.011
Liu C, Zhong Y, Qi X, Chen M, Liu Z, Chen C, Tian X, Li J, Jiao Y, Wang D, Wang Y (2020) Extension of the in vivo haploid induction system from diploid maize to hexaploid wheat. Plant Biotechnol J 18(2):316. https://doi.org/10.1111/pbi.13218
Lv J, Yu K, Wei J, Gui H, Liu C, Liang D, Wang Y, Zhou H, Carlin R, Rich R, Lu T (2020) Generation of paternal haploids in wheat by genome editing of the centromeric histone *CENH3*. Nat Biotechnol 38(12):1397–1401. https://doi.org/10.1038/s41587-020-0728-4
Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50(W1):W276–W279
Meyers B, Zaltsman A, Lacroix B, Kozlovsky SV, Krichevsky A (2010) Nuclear and plastid genetic engineering of plants: comparison of opportunities and challenges. Biotechnol Adv 28(6):747–756. https://doi.org/10.1016/j.biotechadv.2010.05.022
Mhamdi A, Van Breusegem F (2018) Reactive oxygen species in plant development. Development 145(15): dev164376. https://doi.org/10.1242/dev.164376
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, Tosatto SC, Paladin L, Raj S, Finn RLJ, RD (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49(D1):D412–D419. https://doi.org/10.1093/nar/gkaa913
Mushtaq M, Ahmad Dar A, Skalicky M, Tyagi A, Bhagat N, Basu U, Bhat BA, Zaid A, Ali S, Dar TUH, Rai GK (2021) CRISPR-based genome editing tools: Insights into technological breakthroughs and future challenges. Genes 12(6):797. https://doi.org/10.3390/genes12060797
Omasits U, Ahrens CH, Müller S, Wollscheid B (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30(6):884–886. https://doi.org/10.1093/bioinformatics/btt607
Pearson WR (2013) An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinform 42(1):3–1. https://doi.org/10.1002/0471250953.bi0301s42
Pretini N, Vanzetti LS, Terrile II, Donaire G, González FG (2021) Mapping QTL for spike fertility and related traits in two doubled haploid wheat (*Triticum aestivum* L.) populations. BMC Plant Biol 21:1–8. https://doi.org/10.1186/s12870-021-03061-y
Ravi M, Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464(7288):615–618. https://doi.org/10.1038/nature08842
Sears ER (1939) Cytogenetic studies with polyploid species of wheat. I. Chromosomal aberrations in the progeny of a haploid of *Triticum vulgare*. Genetics 24(4):509. https://doi.org/10.1093/genetics/24.4.509
Shi Z, Song W, Xing J, Duan M, Wang F, Tian H, Xu L, Wang S, Su A, Li C, Zhang R (2017) Molecular mapping of quantitative trait loci for three kernel-related traits in maize using a double haploid population. Mol Breed 37:1–10. https://doi.org/10.1007/s11032-017-0706-9

Singh AK, Krishnan SG, Vinod KK, Ellur RK, Bollinedi H, Bhowmick PK, Nagarajan M (2019) Precision breeding with genomic tools: a decade long journey of molecular breeding in rice. Indian J Genet Plant Breed 79(Sup-01):181–191.

Snape JW, Simpson E (1986) The utilisation of doubled haploid lines in quantitative genetics. Bulletin de la Société Botanique de France. Actualités Botaniques 133(4):59–66. https://doi.org/10.1080/01811789.1986.10826799

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Bio Evo 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Thompson KF (1972) Oil-seed rape. Reports of the plant breeding institute. Cambridge University Press, Cambridge, pp 94–96

Wang X (2001) Plant phospholipases. Annu Rev Plant Biol 52(1):211–231. https://doi.org/10.1146/annurev.arplant.52.1.211

Wang S, Ouyang K (2022) Rapid creation of *CENH3*-mediated haploid induction lines using a cytosine base editor (CBE). Plant Biol 25(1):226–230. https://doi.org/10.1111/plb.13482

Wang B, Zhu L, Zhao B, Zhao Y, Xie Y, Zheng Z, Li Y, Sun J, Wang H (2019a) Development of a haploid-inducer mediated genome editing system for accelerating maize breeding. Mol Plant 12(4):597–602. https://doi.org/10.1016/j.molp.2019.03.006

Wang C, Liu Q, Shen Y, Hua Y, Wang J, Lin J, Wu M, Sun T, Cheng Z, Mercier R, Wang K (2019b) Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. Nat Biotechnol 37(3):283–286. https://doi.org/10.1038/s41587-018-0003-0

Wang J, Cao Y, Wang K, Liu C (2022a) Development of multiple-heading-date *mtl* haploid inducer lines in rice. Agriculture 12(6):806. https://doi.org/10.3390/agriculture12060806

Wang N, Xia X, Jiang T, Li L, Zhang P, Niu L, Cheng H, Wang K, Lin H (2022b) In planta haploid induction by genome editing of DMP in the model legume *Medicago truncatula*. Plant Biotechnol J 20(1):22–24. https://doi.org/10.1111/pbi.13740

Wang N, Gent JI, Dawe RK (2021) Haploid induction by a maize *CENH3* null mutant. Sci Adv 7(4): eabe2299. https://doi.org/10.1126/sciadv.abe2299

Watts A, Kumar V, Raipuria RK, Bhattacharya RC (2018) In vivo haploid production in crop plants: methods and challenges. Plant Mol Biol Rep 36(5):685–694. https://doi.org/10.1007/s11105-018-1132-9

Xie E, Li Y, Tang D, Lv Y, Shen Y, Cheng Z (2019) A strategy for generating rice apomixis by gene editing. J Integr Plant Biol 61(8):911–916. https://doi.org/10.1111/jipb.12785

Yao L, Zhang Y, Liu C, Liu Y, Wang Y, Liang D, Liu J, Sahoo G, Kelliher T (2018) *OsMATL* mutation induces haploid seed formation in indica rice. Nat Plants 4(8):530–533. https://doi.org/10.1038/s41477-018-0193-y

Yuan J, Guo X, Hu J, Lv Z, Han F (2015) Characterization of two *CENH3* genes and their roles in wheat evolution. New Phytol 206(2):839–851. https://doi.org/10.1111/nph.13235

Zhong Y, Liu C, Qi X, Jiao Y, Wang D, Wang Y, Chen S (2019) Mutation of *ZmDMP* enhances haploid induction in maize. Nat Plants 5(6):575–580. https://doi.org/10.1038/s41477-019-0443-7

Zhong Y, Chen B, Li M, Wang D, Jiao Y, Qi X, Wang M, Liu Z, Chen C, Wang Y, Chen M (2020) A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous *Arabidopsis*. Nat Plants 6(5):466–472. https://doi.org/10.1038/s41477-020-0658-7

Zhong Y, Chen B, Wang D, Zhu X, Li M, Zhang J, Chen M, Wang M, Riksen T, Liu J, Qi X (2022) In vivo maternal haploid induction in tomato. Plant Biotechnol J 20(2):250–252. https://doi.org/10.1111/pbi.13755

Zhu S, Wang X, Chen W, Yao J, Li Y, Fang S, Lv Y, Li X, Pan J, Liu C, Li Q (2021) Cotton DMP gene family: characterization, evolution, and expression profiles during development and stress. Int J Biol Macromol 183:1257–1269. https://doi.org/10.1016/j.ijbiomac.2021.05.023