

MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction

Timothy Kelliher¹, Dakota Starr¹, Lee Richbourg¹, Satya Chintamanani², Brent Delzer³, Michael L. Nuccio¹, Julie Green¹, Zhongying Chen¹, Jamie McCuiston¹, Wenling Wang¹, Tara Liebler¹, Paul Bullock² & Barry Martin^{1†}

Sexual reproduction in flowering plants involves double fertilization, the union of two sperm from pollen with two sex cells in the female embryo sac. Modern plant breeders increasingly seek to circumvent this process to produce doubled haploid individuals, which derive from the chromosome-doubled cells of the haploid gametophyte. Doubled haploid production fixes recombinant haploid genomes in inbred lines, shaving years off the breeding process¹. Costly, genotype-dependent tissue culture methods are used in many crops², while seed-based *in vivo* doubled haploid systems are rare in nature³ and difficult to manage in breeding programmes⁴. The multi-billion-dollar maize hybrid seed business, however, is supported by industrial doubled haploid pipelines using intraspecific crosses to *in vivo* haploid inducer males derived from Stock 6, first reported in 1959 (ref. 5), followed by colchicine treatment. Despite decades of use, the mode of action remains controversial^{6–10}. Here we establish, through fine mapping, genome sequencing, genetic complementation, and gene editing, that haploid induction in maize (*Zea mays*) is triggered by a frame-shift mutation in *MATRILINEAL* (*MTL*), a pollen-specific phospholipase, and that novel edits in *MTL* lead to a 6.7% haploid induction rate (the percentage of haploid progeny versus total progeny). Wild-type *MTL* protein localizes exclusively to sperm cytoplasm, and pollen RNA-sequence profiling identifies a suite of pollen-specific genes overexpressed during haploid induction, some of which may mediate the formation of haploid seed^{11–15}. These findings highlight the importance of male gamete cytoplasmic components to reproductive success and male genome transmittance. Given the conservation of *MTL* in the cereals, this discovery may enable development of *in vivo* haploid induction systems to accelerate breeding in crop plants.

Plants have distinct haploid and diploid generations. In flowering plants, pollen grains are the haploid male gametophyte, and consist of a large vegetative cell and two sperm cells. During pollination, the grain germinates a tube that exhibits rapid tip-growth as it navigates down the style, guided by chemo-attractants secreted by ovular tissues at the base of the embryo sac¹⁶. During transmittance down the tube, sperm cell nuclei are connected to the vegetative nucleus by a common cytoplasm called the male germ unit. Shortly after contact with one of the two synergid cells at the micropylar pole of the embryo sac, the pollen tube bursts¹⁷ and the two sperm cell nuclei are propelled across the dying synergid cell's cytoplasm to fuse with the egg and central cell. This produces the embryo and endosperm, the main components of the seed. If double fertilization fails, a second pollen tube can fertilize the ovule via interaction with the persistent synergid.

Haploid gametophytes can be induced to develop as doubled haploid sporophytes through *in vitro* tissue culture² or haploid seed induction^{3,4,18}. Doubled haploid lines enable recombinant populations to be screened rapidly in homozygous states¹. In short, doubled haploid breeding is the most efficient way to improve crops and integrate new

genes to make crop production more sustainable in a changing world. In maize, haploids are routinely produced using crosses to Stock 6, reported in 1959 as producing 2–3% maternal haploid seed during self-pollination or when outcrossed as a male⁵. The presence of a functional endosperm enables the haploid seed to be viable and implies at least partial fertilization has occurred⁶. Academic and commercial routine users of Stock 6 have bred derivatives for elevated (7–15%) haploid induction rates (HIR), at the expense of increasing rates of kernel abortion, leading to the hypothesis that these phenomena are inextricably linked^{7,8,10}.

To clarify the developmental genetics underlying haploid induction, the Stock 6 derivative RWK (approximately 13% HIR) was obtained from the University of Hohenheim in 2006, crossed to the inbreds NP2460 and NP2391, and subsequently backcrossed to RWK to generate mapping populations (Extended Data Fig. 1). Elevated HIR in both inbred populations co-segregated with the marker SM020SDQ in bin 1.04, consistent with recent reports that a quantitative trait locus at this location called *qhir1* is critical for haploid induction¹⁰. Several rounds of fine mapping narrowed this quantitative trait locus to an approximately 0.57 Mb region containing seven genes (Fig. 1a) consistent with *qhir11*, a sub-region of *qhir1* (refs 19, 20). Using the Illumina HiSeq2000 platform, we sequenced RWK, Stock 6, and a BC3F5 non-inducer 'RWK-NIL' that is near-isogenic to RWK but has NP2391 haplotypes in *qhir11*. The sequences for the seven genes were nearly identical in B73 and RWK-NIL, but RWK and Stock 6 lacked GRMZM2G062320, a PHOSPHOGLYCERATE MUTASE (PGM), and had a 4 bp insertion in GRMZM2G471240, a patatin-like phospholipase (Fig. 1b and Supplementary Table 1).

A screen of 19 Stock 6 derivatives found that all 19 were homozygous for the inducer haplotypes of these two candidates, including NP2222-Haplod Inducer (NP2222-HI), a BC3 introgression of RWK into Syngenta's standard transformable inbred line (Fig. 1c and Extended Data Table 1). Meanwhile, all nine non-inducer lines were homozygous for the genes' B73/RWK-NIL haplotypes. Heterologous complementation of NP2222-HI (10.2% HIR) with wild-type copies of those genes virtually eliminated haploid induction and kernel abortion for GRMZM2G471240 only (Table 1, Fig. 1d and Extended Data Table 2). Attempts to knockdown the gene candidates' transcripts using RNA interference (RNAi) led to elevated rates of haploid formation for GRMZM2G471240^{RNAi} only (Extended Data Table 3). Together these data suggest that the frame-shift in GRMZM2G471240 is a loss-of-function mutation responsible for haploid induction and the associated ear phenotypes. Because GRMZM2G062320 did not complement (Table 1), it probably has no role in haploid induction. We renamed GRMZM2G471240 *MATRILINEAL* (*MTL*) and the mutant allele *matrilineal* (*mtl*).

Several *mtl*-like alleles were generated in the non-inducer NP2222 by introducing small deletions near the native 4 bp insertion site, using transcription-activator-like effector nuclease (TALEN) technology²¹.

¹Seeds Research, Syngenta Crop Protection, 9 Davis Drive, Research Triangle Park, North Carolina 27709, USA. ²Syngenta Seeds, 2369 330th Street, Slater, Iowa 50244, USA. ³Syngenta Seeds, 4133 East County Road O, Janesville, Wisconsin 53546, USA. [†]Present address: CiBO Technologies, 155 2nd Street, Cambridge, Massachusetts 02141, USA.

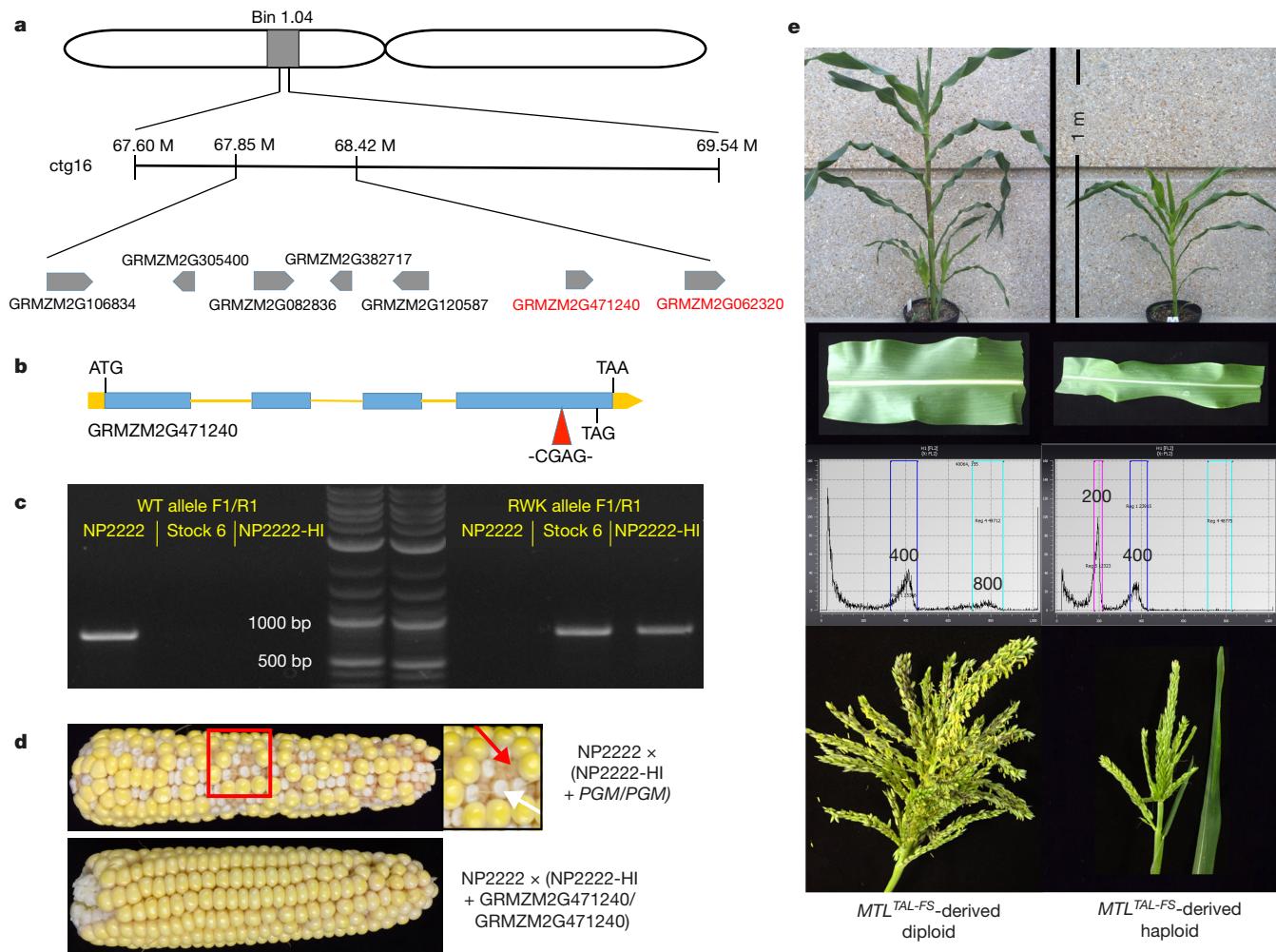


Figure 1 | A mutation in MTL is responsible for haploid induction in maize. **a**, A bin 1.04 marker co-segregated with haploid induction. Fine mapping identified a 0.57 Mb region containing seven genes (key genes are in red). **b**, GRMZM2G471240 structure with insertion and premature stop codon (TAG) shown. **c**, PCR products of GRMZM2G471240 RWK and non-inducer alleles with marker sizes

Several mutant events were self-pollinated and T1 plants lacking the TALEN transfer DNA (tDNA) insert but biallelic for target site frame-shift mutations in *MTL* were identified. These T1 plants and their derivatives, hereafter called *MTL*^{TAL-FS} lines, exhibited a HIR of 4.0–12.5% (average 6.65%) when outcrossed onto NP2222 (Table 1 and Extended Data Table 4). Ploidy status of putative haploids was confirmed by flow cytometry and phenotypic evaluations (Fig. 1e). This proves that the frame-shift in *MTL* is sufficient to induce high rates of haploid induction in maize. Other contributors to the phenotype are defined, such as *qhir12* (ref. 20), a region of *qhir1* just downstream of

Table 1 | Reproductive characteristics of haploid inducer, complementation and edited lines

Cross	Events	Crosses	KA (%)	Embryos	Haploids	HIR (%)
NP2222 × NP2222	NA	8	0.7	1432	1	<0.1
NP2222 × NP2222-HI	NA	4	47.6	531	54	10.2
NP2222 × NP2222-HI + 471240/471240*	3	17	0.7	4321	11	0.3
NP2222 × NP2222-HI + PGM/PGM†	3	10	47.1	1508	149	9.9
NP2222 × <i>MTL</i> ^{TAL-FS}	4	15	41.6	1774	118	6.7

*471240, GRMZM2G471240; †PGM, GRMZM2G062320; KA, kernel abortion frequency.

indicated. **d**, NP2222 ears crossed by NP2222-HI complemented by wild-type copies of PGM or GRMZM2G471240. The PGM ear exhibits kernel abortion (inset, white arrow) and fertilization failure (inset, red arrow) typical of inducer crosses. **e**, Progeny of NP2222 crossed by *MTL*^{TAL-FS}. Diploid (left) and haploid (right) plants, V7 leaves, flow cytometry results (signal intensity values indicated), and tassels.

qhir11. It is reasonable to infer that haploid induction and kernel abortion rates are set through *mtl* by paternal²⁰ and maternal²² genotype-specific interactions.

Haplod seed formation in maize is a post-zygotic character triggered by a defective male gametophyte, a fact reflected in the tissue specificity of the *MTL* gene. Public RNA sequence (RNA-seq) profiles indicated that wild-type *Mtl* is specific to pre-dehiscent (R1-staged) anthers²³, in agreement with a recent anther developmental profile that found *Mtl* exclusively in anthers during pollen shed²⁴. Wild-type pollen had nine times more *Mtl* than post-anthesis anther sacs (Fig. 2a), indicating *Mtl* is male gametophyte-specific. Compared with NP2222, *Mtl* was elevated in NP2222-HI but not *MTL*^{TAL-FS}, while the abundance of the two annotated splice variants was consistent (Extended Data Fig. 2).

The frame-shift in *mtl* occurs at amino acid 380, which is followed by 20 altered residues and a premature stop codon that truncates the protein by 29 amino acids (Extended Data Fig. 3a). Full-length functional reporter lines were used to characterize wild-type and mutant *MTL* localization. No signal was seen in NP2222 pollen or fertilized embryo sacs (Fig. 2b, e, h and Supplementary Video 1), nor in *mtl*-GFP (*mtl* fused to a fluorophore of green fluorescent protein) pollen or embryo sacs fertilized by *mtl*-GFP pollen (Fig. 2c, f, i and Supplementary Video 2). By contrast, a fluorescent signal was found in sperm cell cytoplasm in

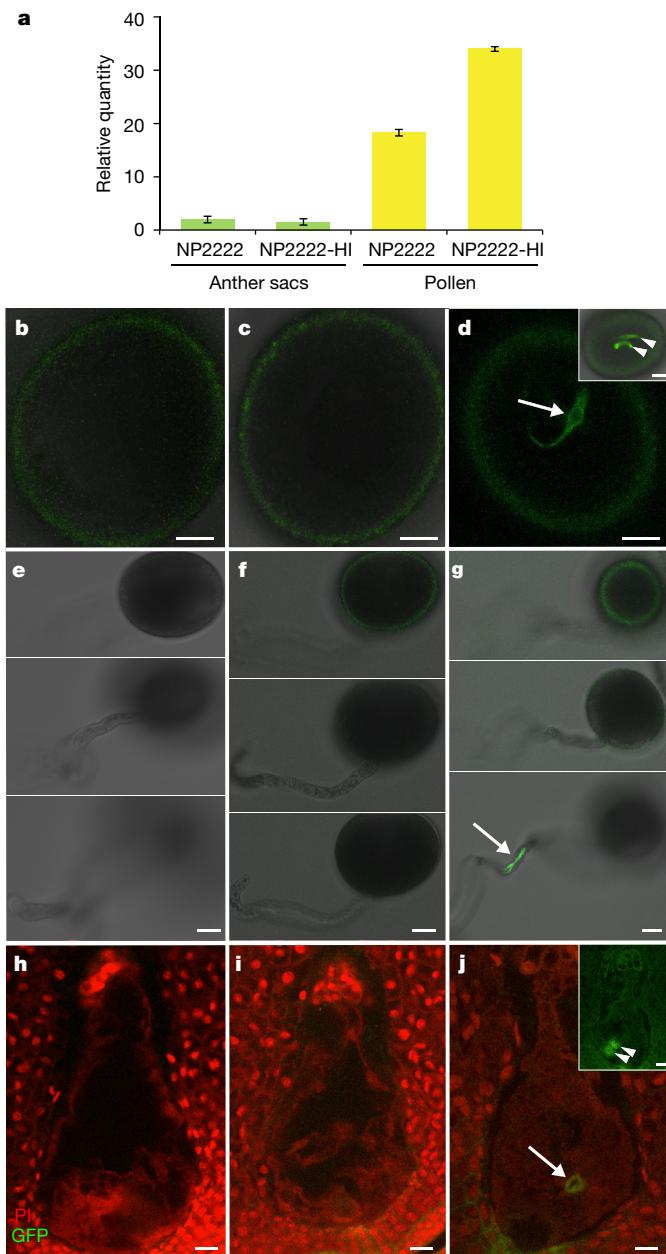


Figure 2 | MTL is specifically found in the cytoplasm of the male gametes in pollen grains. **a**, *Mtl* is pollen-specific. Relative quantity = $\frac{1}{2} - \Delta C_t$. Error bars, s.d. **b**, **c**, NP2222 (**b**) and *mtl*-GFP (**c**) pollen with background exine fluorescence. **d**, *MTL*-GFP signal in sperm cell cytoplasm (arrow); inset, signal in both sperm cells in one plane (arrowheads). **e**, **f**, Optical sections of germinated NP2222 (**e**) and *mtl*-GFP (**f**) pollen grains and tubes showing no signal. **g**, Germinated *MTL*-GFP pollen with localization to the sperm cells of the male germ unit (arrow). **h**, **i**, NP2222 embryo sacs crossed by NP2222 (**h**) or *mtl*-GFP (**i**) 18 h after pollination showing no signal. PI, propidium iodide counterstain. **j**, NP2222 embryo sac crossed by *MTL*-GFP 18 h after pollination showing fluorescence in a sperm cell (arrow); inset, embryo sac showing fluorescence in both sperm cells in one plane (arrowheads) without propidium iodide counterstain. Scale bar, 15 μ m.

MTL-GFP pollen (Fig. 2d and Supplementary Videos 3 and 4) and germinated pollen tubes (Fig. 2g and Supplementary Video 5). NP2222 embryo sacs fixed 18 h after pollination with MTL-GFP pollen had signal in the area of the degenerating synergid consistent with that of sperm cells delivered during fertilization (Fig. 2j and Supplementary Videos 6 and 7). MTL-GFP but not *mtl*-GFP eliminated haploid

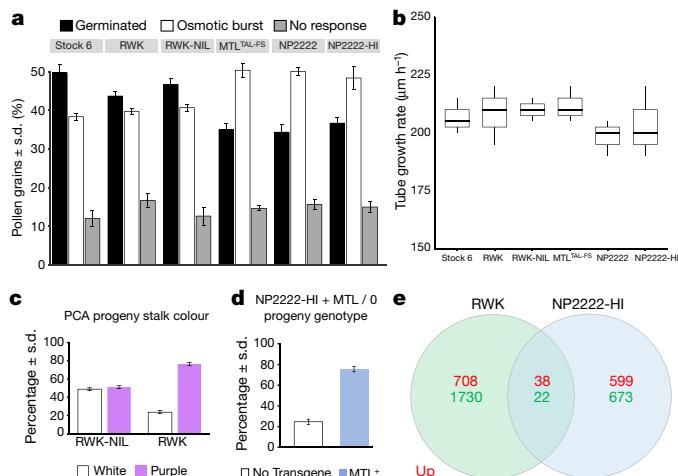


Figure 3 | The *mtl* allele is responsible for pleiotropic phenotypes associated with haploid induction. **a**, Pollen tube germination rate was similar in inducers and non-inducers ($n = 200$). **b**, Initial pollen tube elongation was also similar ($n = 25$). **c**, RWK but not RWK-NIL is subject to segregation distortion based on low (25%) trait transmission in germinated progeny ($n = 300$). **d**, *MTL*/0 complementation lines also exhibit segregation distortion against *mtl* in germinated progeny ($n = 400$). **e**, Venn diagram showing RNA-seq profiling results of two haploid inducer-near isogenic pairs (left, RWK versus RWK-NIL; right, NP2222-HI versus NP2222; red text, upregulated; green text, downregulated). Only 60 genes were found significantly changed in the same direction.

induction in NP2222-HI (Extended Data Table 2). Both mutant and wild-type recombinant MTL proteins exhibited phospholipase²⁵ activity *in vitro* in PLA2 fluorescent liposome cleavage assays (Extended Data Fig. 3b). MTL was found by liquid chromatography/tandem mass spectrometry (LC-MS/MS) in NP2222 pollen, but was missing in all three NP2222-HI replicates (Extended Data Table 5). Collectively these data indicate that MTL is a phospholipase specific to the sperm cell cytoplasm, and that the frame-shift in *mtl* compromises MTL localization or stability in haploid inducer pollen.

The identification of *MTL* as the causative gene in maize haploid induction permitted dissection of the pleiotropic phenotypes historically associated with the trait. Phospholipase mutations are associated with delayed pollen germination and tube growth¹¹, but these were normal in RWK, Stock 6, and *MTL*^{TAL-FS} lines (Fig. 3a, b). Ears pollinated with NP2222-HI and *MTL*^{TAL-FS} pollen exhibit ~10–25% fertilization failure, and a pollen competition assay showed that RWK is subject to segregation distortion (Fig. 3c), consistent with previous work⁷. Crosses with hemizygous NP2222-HI + *MTL*/0 pollen produced a proportional bias towards *MTL*⁺ progeny (Fig. 3d), indicating that inducer segregation distortion is attributable to *mtl*. Embryo abortion, a persistent by-product of haploid induction linked to endosperm proliferation failure⁹, occurred at similar rates in native and *MTL*^{TAL-FS} inducers (Table 1). Collectively these data implicate *mtl* in every reproductive defect associated with haploid induction. The two mechanisms typically proffered to explain haploid formation are single fertilization⁶ and post-zygotic genome elimination^{8,9}. In the former, haploids result from fertilization of the central cell but not the egg, which subsequently develops via parthenogenesis. In the latter, double fertilization precedes male chromosome elimination^{8,9}. Clarifying the precise mechanism will require careful embryology after *MTL*^{TAL-FS} pollinations, along with quantitative data tracking the rare persistence of male DNA in maize haploids^{8,9}.

To address the molecular impact of *mtl*, RNA-seq profiling was performed on two pairs of near-isogenic lines (RWK versus RWK-NIL, and NP2222-HI versus NP2222) (Supplementary Tables 2–5).

Table 2 | Genes upregulated in both RWK and NP2222-HI lines

Gramene identifier	Annotation	log ₂ value	
		RWK vs RWK-NIL	NP2222-HI vs NP2222
GRMZM2G118919*	Ca ²⁺ -transporting ATPase	0.77	1.73
GRMZM2G476000*	Ca ²⁺ -transporting ATPase	0.52	0.90
GRMZM2G022180	Calnexin, cnx2	0.71	0.73
GRMZM2G134668	Calnexin, cnx1	0.65	1.49
GRMZM2G314292*	Pollen calcium binding protein, EF hand protein	0.41	0.31
GRMZM2G067511	Calmodulin	0.66	0.76
GRMZM2G422045*	Inositol phosphatase (specificity for IP1,4 5)	1.10	0.64
GRMZM2G072089*	RHO GDP-dissociation inhibitor	1.15	0.62
GRMZM2G075719	RAS-related GTP-binding, RAB6	0.72	0.52
GRMZM2G095441*	Syntaxin-type t-SNARE	0.76	1.45
GRMZM2G485898*	Dynamin GTP-ase	0.58	0.67
GRMZM2G416733*	Syntaxin-type t-SNARE	1.68	2.13
GRMZM2G435294*	XIE, myosin family protein	0.56	1.27
GRMZM2G017170	Amino-acid transporter	0.51	0.77
GRMZM2G064603*	ABC transporter	0.45	0.87
GRMZM2G018148*	Oligopeptide transporter, ysl7	0.96	0.81
GRMZM2G130831*	Cation exchange CHX15, Na ⁺ /H ⁺ antiporter	0.58	1.33
GRMZM2G363557	Cation exchange CHX15, Na ⁺ /H ⁺ antiporter	0.52	0.65
GRMZM2G366851*	Cation exchange CHX15, Na ⁺ /H ⁺ antiporter	0.50	0.66
GRMZM2G430755*	Cation exchange CHX2, Na ⁺ /H ⁺ antiporter	0.63	1.41
GRMZM2G438378*	Phospholipase D alpha 1	0.70	1.53
GRMZM2G170760	ENTH/ANTH/VHS superfamily protein	0.50	0.41
GRMZM2G162954	RNA binding protein	0.46	0.92
GRMZM2G168163	RNA binding protein	0.65	1.28
GRMZM2G062476	Glycine-enriched protein	0.50	0.64
GRMZM5G864319	Acylic-coenzyme A oxidase 1.2 peroxisomal	0.71	0.93
GRMZM2G390013	Unannotated	1.00	0.63
GRMZM5G894862	Unknown function	1.33	1.31

*Pollen-specific genes.

This permitted identification of 60 differentially expressed genes significantly changed in the same direction at a 1% false discovery rate (Fig. 3e, Table 2 and Extended Data Table 6). Strikingly, 15 of the 28 differentially expressed genes upregulated in inducer pollen were co-expressed with *MTL* (that is, pollen-specific), including GRMZM2G435294, which codes for a myosin and maps to the minor quantitative trait locus *qhir8* (ref. 26). A Fisher's exact test²⁷ indicated that 15/28 represents a highly significant enrichment of pollen-specific genes ($P < 2.2 \times 10^{-16}$). Six of the upregulated differentially expressed genes act in Ca²⁺ signalling, which is involved in pollen tube guidance, sperm capacitation, and fertilization^{12–15,28}, while others are annotated as being involved in endomembrane transport and signalling (Table 2 and Extended Data Table 6). Given the intensive communication between the two gametophytes and the dynamic membrane reshuffling involved in fertilization²⁸, we speculate that some of these differentially expressed genes may mediate one or more of the pleiotropic phenotypes triggered by *mtl*. It would be worthwhile investigating whether *mtl* sperm cells have altered membrane composition or lipid profiles, disrupted intracellular signalling cascades, or altered chromatin conformation.

Haploid induction was recently engineered in *Arabidopsis* via manipulation of *CENTROMERIC HISTONE3*, which causes uniparental genome elimination¹⁸ through post-zygotic centromere imbalance between hybridized genomes. An attempt to replicate this in maize was successful at a low rate²⁹, but this is the first instance of a haploid inducer system triggered by a cytoplasmic protein.

MTL is highly conserved in cereals (Extended Data Fig. 4). The rice orthologue is pollen-specific³⁰, suggesting functional conservation, but the closest *Arabidopsis* homologue, *AtPLP2*, is only expressed in vegetative tissues²⁵. These findings highlight the importance of non-nuclear sperm components in reproductive success and faithful male genome transmittance, and may lead to the development of intraspecific *in vivo* haploid inducer lines in important crop plants.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 April; accepted 23 November 2016.

Published online 23 January 2017.

- Chang, M.-T. & Coe, E. in *Molecular Genetics Approaches to Maize Improvement* (eds Kriz, A. L. & Larkins, B. A.) 127–142 (Springer, 2009).
- Seguí-Simarro, J. M. & Nuez, F. Embryogenesis induction, callogenesis, and plant regeneration by *in vitro* culture of tomato isolated microspores and whole anthers. *J. Exp. Bot.* **58**, 1119–1132 (2007).
- Kasha, K. J. & Kao, K. N. High frequency haploid production in barley (*Hordeum vulgare* L.). *Nature* **225**, 874–876 (1970).
- Chase, S. S. Monoploids and monoploid-derivatives in maize (*Zea mays* L.). *Bot. Rev.* **35**, 117–167 (1969).
- Coe, E. H. A line of maize with high haploid frequency. *Am. Nat.* **93**, 381–382 (1959).
- Sarkar, K. R. & Coe, E. H. A genetic analysis of the origin of maternal haploids in maize. *Genetics* **54**, 453–464 (1966).
- Xu, X. et al. Gametophytic and zygotic selection leads to segregation distortion through *in vivo* induction of a maternal haploid in maize. *J. Exp. Bot.* **64**, 1083–1096 (2013).
- Zhao, X., Xu, X., Xie, H., Chen, S. & Jin, W. Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. *Plant Physiol.* **163**, 721–731 (2013).
- Qiu, F. et al. Morphological, cellular and molecular evidences of chromosome random elimination *in vivo* upon haploid induction in maize. *Curr. Plant Biol.* **1**, 83–90 (2014).
- Prigge, V. et al. New insights into the genetics of *in vivo* induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics* **190**, 781–793 (2012).
- Kim, H. J. et al. Endoplasmic reticulum- and Golgi-localized phospholipase A2 plays critical roles in *Arabidopsis* pollen development and germination. *Plant Cell* **23**, 94–110 (2011).
- Denninger, P. et al. Male-female communication triggers calcium signatures during fertilization in *Arabidopsis*. *Nature Commun.* **5**, 4645 (2014).
- Hamamura, Y. et al. Live imaging of calcium spikes during double fertilization in *Arabidopsis*. *Nature Commun.* **5**, 4722 (2014).
- Schiött, M. et al. A plant plasma membrane Ca²⁺ pump is required for normal pollen tube growth and fertilization. *Proc. Natl Acad. Sci. USA* **101**, 9502–9507 (2004).
- Escoffier, J. et al. Group X phospholipase A2 is released during sperm acrosome reaction and controls fertility outcome in mice. *J. Clin. Invest.* **120**, 1415–1428 (2010).
- Okuda, S. et al. Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**, 357–361 (2009).
- Leydon, A. R. et al. Pollen tube discharge completes the process of synergid degeneration that is initiated by pollen tube-synergid interaction in *Arabidopsis*. *Plant Physiol.* **169**, 485–496 (2015).
- Ravi, M. & Chan, S. W. L. Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618 (2010).
- Dong, X. et al. Fine mapping of *qhir1* influencing *in vivo* haploid induction in maize. *Theor. Appl. Genet.* **126**, 1713–1720 (2013).
- Hu, H. et al. The genetic basis of haploid induction in maize identified with a novel genome-wide association method. *Genetics* **202**, 1267–1276 (2016).
- Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
- Wu, P., Li, H., JiaoJiao, R. & Chen, S. Mapping of maternal QTLs for *in vivo* haploid induction rate in maize (*Zea mays* L.). *Euphytica* **196**, 413–421 (2014).
- Sekhon, R. S. et al. Genome-wide atlas of transcription during maize development. *Plant J.* **66**, 553–563 (2011).
- Zhai, J. et al. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc. Natl Acad. Sci. USA* **112**, 3146–3151 (2015).
- La Camera, S. et al. A pathogen-inducible patatin-like lipid acyl hydrolase facilitates fungal and bacterial host colonization in *Arabidopsis*. *Plant J.* **44**, 810–825 (2005).
- Liu, C. et al. Fine mapping of *qhir8* affecting *in vivo* haploid induction in maize. *Theor. Appl. Genet.* **128**, 2507–2515 (2015).
- Schug, J. et al. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).

28. Dresselhaus, T., Sprunck, S. & Wessel, G. M. Fertilization mechanisms in flowering plants. *Curr. Biol.* **26**, R125–R139 (2016).
29. Kelliher, T. et al. Maternal haploids are preferentially induced by *CENH3-tailswap* transgenic complementation in maize. *Front. Plant Sci.* **7**, 414 (2016).
30. Singh, A. et al. Rice phospholipase A superfamily: organization, phylogenetic and expression analysis during abiotic stresses and development. *PLoS ONE* **7**, e30947 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Data2Bio for the RNA-sequencing and data processing efforts. Thanks to M. Evans of the Carnegie Institute of Plant Biology for embryo sac dissection advice. We thank M. McLaughlin and T. Strebe for plant care, C. Fan for Taqman design, and K. Cox for sampling. We thank S. Dong, S. Nalapalli, H. Zhong, and A. Prairie for construct design, transformation and event production. We thank D. Skibbe and I. Jepson for advice and suggestions, and R. Sessler for proteomics efforts. We thank B. Dietrich for sequence assembly support, Q. Que and L. Shi for laboratory space and personnel, and R. Riley and E. Dunder for project support. Thanks to C. Leming, M. Dunn, S. Miles, D. Skalla, and B. Houghteling for licensing and guidance on intellectual property.

Author Contributions T.K.: experimental design, management, crossing, embryo extraction, imaging, analysis, writing. D.S.: plant sampling, PCR, crossing, embryo extraction, manuscript editing. L.R.: embryo extraction for HIR determination. S.C.: experimental design, broad and fine mapping, manuscript preparation and editing. B.D.: experimental design, line acquisition, broad and fine mapping. M.N.: project initiation, experimental design, candidate gene analysis, phospholipase assays. J.G.: RNA-seq data analysis, manuscript preparation and editing. Z.C.: TALEN technology, construct design, experimental design, manuscript editing. J.M.: Taqman trial design and marker analysis for HIR determination. W.W.: Taqman trial design and marker analysis for HIR determination. T.L.: trial planning and management, plant care and pollinations. P.B.: experimental design, line acquisition, mapping, guidance and theory. B.M.: experimental design, project sponsorship, manuscript editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.K. (tim.kelliher@syngenta.com).

Reviewer Information *Nature* thanks S. Scholten and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mapping. As shown in Extended Data Fig. 1, the Stock 6 derivative RWK was obtained from the University of Hohenheim in 2006 and subsequently crossed to inbreds NP2460 and NP2391 and then backcrossed to RWK to generate mapping populations. HIR testing and marker analysis was performed on both sets of mapping populations and co-segregated with the marker SM020SDQ on the long arm of chromosome 1. Several rounds of fine mapping in RWK-NP2391 BC3F3-BC3F5 populations narrowed the quantitative trait locus to a ~0.57 Mb interval between SM2589 and SM2608. A BC3F5 non-inducer with NP2391 haplotypes in the interval but RWK haplotypes for the rest of the genome was named RWK-NIL (for RWK near-isogenic line).

Genome sequencing. Whole-genome sequencing of RWK-NIL, RWK, and Stock 6 was performed on the Illumina HiSeq2000. Short insert (200–500 bp) paired-end libraries were sequenced to a depth of 50–80 times the genome size using 100 cycle paired-end reads. Long insert paired-end libraries (inserts 4–6 kb) were sequenced using 50 cycle paired-end runs to a depth of 8–20 times genome coverage. Local assemblies of reads (and their pair) aligning to region chr1:67,500,000–71,500,000 of the maize genome assembly B73 RefGen_v2 were generated using the ABySS genome assembler. The final ABySS scaffolds were aligned to the chr1:67,500,000–71,500,000 reference region using the BLASTN program and the best single hits >500 bp per query region were retained.

Vector construction. Three *MTL* transgenes were synthesized under control of the endogenous *MTL* promoter, including *MTL-GFP*, *mtl-GFP*, and *MTL* (the last for complementation). A fourth transgene was synthesized containing the wild-type *PGM* under control of the *PGM* native promoter. For the *MTL-GFP* and *mtl-GFP* transgenes, the Clontech (<http://www.clontech.com>) fluorophore AcGFP1 (which we call *GFP*) was fused to the carboxy (C) terminus of wild-type *MTL* and *mtl* alleles respectively. These lines were used to establish *MTL* protein localization during pollen development and fertilization. Two additional constructs were made containing RNAi hairpins designed to knockdown native *Mtl* and *Pgm* transcripts and driven by a ubiquitin promoter. Finally, to generate targeted mutations in the *MTL* gene, a binary vector was constructed to express a pair of TALENs targeting the fourth exon of *MTL* near the 4 bp insertion site in inducer lines 5'-TCCAGGGTC
AACGTGGAGACAGGGAGGTACGAACCGGTGACTGGCGAAGGAAGCA-3' with the underlined portions corresponding to the forward and reverse recognition sequences of individual TALEN molecules. The selectable marker in this construct was *PMI* (phosphomannose isomerase) driven by the maize ubiquitin promoter for mannose selection³¹. tDNA in this construct was transformed into the standard transformable inbred line NP2222 using *Agrobacterium tumefaciens*-mediated transformation of immature embryos. T0 plants carrying single-copy were sent to the greenhouse, grown to maturity, and self-pollinated.

Confocal microscopy. Homozygous NP2222 T1 plants carrying two copies of the *MTL-GFP* and *mtl-GFP* transgenes were sampled and imaged to determine the tissue- and subcellular localization of the mutant and wild-type versions of *MTL*. In two individuals from two events for both transgenes, mature pollen samples were either imaged fresh, germinated on liquid pollen germination medium³² (10% sucrose, 0.01% boric acid, 0.1% yeast extract, 10 mM CaCl₂, 50 μM KH₂PO₄, 15% PEG 4000), or imaged on silks less than 1 h after pollination. Embryo sacs pollinated by homozygous T1 plants carrying two copies of *MTL-GFP* or *mtl-GFP* were fixed in FAA (formalin acetic acid: 3.7% formaldehyde, 5% acetic acid, 50% ethanol) between 16 and 21 h after pollination from greenhouse-grown ears. Depending on the position of the ovule on the ear and environmental factors including temperature and humidity, fertilization should occur between 12 and 16 h after pollination. Dissected ovules were fixed using vacuum infiltration on ice, transferred to PBS and de-stained in ClearSee³³ for 1–3 weeks, stained with propidium iodide using established techniques³⁴, and imaged. Imaging was performed on a Zeiss laser scanning microscope (LSM 710). The *GFP* fluorophore was visualized at 488 nm and the emission spectra were collected at 520–550 nm; propidium iodide was excited at 560 nm and the emission spectra were collected at 600–650 nm. Images were captured using an AxioCam MRC camera and processed in ZEN 2011 analysis software.

Ploidy evaluation. Haplod induction assays were performed by outcrossing putative haploid inducer lines onto NP2222. Fourteen to seventeen days after pollination, ears were harvested, disinfected with 25% bleach, and then viable and aborted kernels were counted. In most cases, well over 100 embryos were extracted and placed on solid media and allowed to germinate a rudimentary shoot during 3- to 7-day incubations in the dark. The root and base of the scutellum and embryo proper were then sampled and tested for ploidy via PCR assay, while the remaining tissues (scutellum, embryo, and shoot) were stored at -80°C for

ploidy confirmation via flow cytometry (ploidy analysis) as described previously²⁹. Putative haploids generated by native inducer lines such as RWK and NP2222-HI were identified by Taqman zygosity assays that detected the wild-type *MTL* and mutant *mtl* alleles (Fig. 1d and Extended Data Table 1). Diploids were identified as having 1 copy of the *MTL* allele and one copy of the *mtl* allele (the paternal copy), while putative haploids were identified as having zero copies of the *mtl* allele and 'two' copies of the maternal *MTL* allele. Putative haploid embryos generated by test crosses to homozygous complementation lines NP2222-HI + *PGM/PGM* and NP2222-HI + *MTL/MTL* were identified by their having zero copies of the tDNA using tDNA-specific Taqman assays. Diploids were identified by having one copy of the tDNA insertion. All putative haploids in these events were confirmed as such by flow cytometry. For the RNAi events, homozygous T1 transgenic lines were backcrossed as males to NP2222 and progeny were extracted and germinated as described above. Diploids were identified by having one copy of the transgene, while putative haploids were identified by having zero copies of the transgene.

Flow cytometry. Because aneuploidy or pollen contamination could easily lead to false scoring using these methods, all putative haploids were tested by flow cytometry on a CySpace Ploidy Analyzer with an ultraviolet lamp to detect 4',6-diamidino-2-phenylindole (DAPI stained) nuclei (<http://www.sysmex-partec.com>). Samples were sliced in 400 μl extraction buffer for 30 s with a razor blade. Nuclei were passed through a 30 μM filter and mixed with 1.6 ml of DAPI staining buffer. The gain was set at 482, the speed at 3 μl/s, and the sample concentration was established to be ~1,500 cells per millilitre. Ploidy status was called as follows: haploid, first peak 190–210, second peak 360–420; diploid, first peak 390–410, second peak 760–820. The few putative haploids that were not confirmed by flow cytometry predominantly had peaks that warranted a diploid call. The contradiction in results was attributed to assay or sampling error, although it remains a possibility that partial chromosome elimination resulting in chimaeric plants is the true cause.

TALEN mutant analysis. T0 and T1 TALEN events were analysed using Taqman assays that detected the TALEN tDNA insertion and the *MTL* target site (using the probe 5'-AGGGAGGTACGAACC-3', which overlaps with the native 4 bp insertion site and the TALEN target site). Transgenic events with reduced copy number of the *MTL* target site-specific assay were identified as putative mutants and subsequently sequenced using PCR with the primers GRMZM2G471240_TAL-F and _TAL-R. PCR products were sub-cloned and sequenced and alignments were performed using Sequencher 5.2.3 (<http://www.genecodes.com>). Several deletion alleles were found. Events carrying monoallelic or biallelic frame-shift mutations were self-pollinated, and Taqman assays were used to select T1 plants that lacked the tDNA but biallelic (homozygous) for the frame-shift mutations. The mutations present were further confirmed by PCR, subcloning, and sequencing, as described for the T0. At least six colonies per PCR product were sequenced. Of the multiple events screened in the T1 generation, four had several T1 progeny that lacked the TALEN tDNA insert but contained unique frame-shift mutations in both copies of *MTL*, including deletions at the target site of 5 bp, 8 bp, 13 bp, 11 bp, and 28 bp. We refer to such biallelic T1 plants and their derivatives as *MTL*^{TAL-FS} plants. Several *MTL*^{TAL-FS} T1 plants were self-pollinated to make T2 plants that were used in subsequent analyses. We refer to these self-pollinated lines (belonging to the T1 generation and beyond) as *MTL*^{TAL-FS} lines. In the haploid induction tests, these were outcrossed onto several NP2222 ears and the resulting progeny were extracted and evaluated for ploidy by the site-specific assay (see Supplementary Table 7 and the below paragraph for an example and explanation of HIR data generation and analysis). All putative haploid embryos were tested by flow cytometry for ploidy status.

Supplementary Table 7 provides an example of how we determined putative HIR of *MTL*^{TAL-FS} lines via Taqman assays. The data in that table are from a population of embryos dissected from an NP2222 ear that was pollinated with *MTL*^{TAL-FS} pollen. The 'TALEN cut site' assay detects the wild-type *MTL* haplotype at the target site for mutagenesis. Only alleles with an identical sequence to wild-type will amplify; the mutant (edited) alleles will not. The probe used is 5'-AGGGAGGTACGAACC-3', which overlaps with the 4 bp insertion site. As a zygosity assay, it measures the PCR amplification of the target against an endogenous control, enabling copy number analysis. This assay is a good method to evaluate HIRs because the scores can be used to distinguish putative haploids from diploids. A score of '1' indicates one wild-type copy, which in this case represents a diploid embryo with one *MTL* allele from the maternal NP2222 parent and nothing from the male parent, because the male parent has one (novel, edited) *MTL*^{TAL-FS} allele. This one copy is compared with the endogenous control, which will always have a score of 2 for diploids. The ratio of one copy for the wild-type *MTL* assay to two copies for the control assay leads to a 'score' of 1 ('single copy'). By contrast, a score of '2' indicates a putative haploid. This is because in the haploid embryos, the male DNA is missing, so there is only one copy of every gene in the genome. The one wild-type maternal copy is amplified, and in the zygosity analysis it is

compared with the one endogenous control copy, and the ratio of 1:1 ends up with a 'score' of 2 ('two copy'). Using this test, anything that is called a '1 or 2' or a '2' is a putative haploid, while all plants scoring as '1' are putative diploids. We then test each putative haploid by flow cytometry and either confirm or reject the hypothesis of haploidy. Only the confirmed haploids were used in the total HIR calculation. Using similar logic but different assays, we were able to consistently use Taqman to assay HIRs in progeny of the native, complementation, and RNAi lines.

LC–MS/MS. Label-free quantitative proteomic data were generated on a variety of pollen samples to evaluate the levels of GRMZM2G471240. Fresh pollen samples (~200 mg) were obtained from NP2222 and NP2222-HI anthesis-staged tassels, frozen on liquid nitrogen, and ground at –80 °C using a ceramic mortar and pestle. A small spatula tip of each sample was put in 4% SDS/200 mM TRIS (pH 8)/100 mM DTT and heated for 10 min at 95 °C. Insoluble material was pelleted, and supernatants moved to new tubes and precipitated with methanol/chloroform. Precipitate samples were resuspended in 100 µl 8 M urea/4% CHAPS and measured for protein content by Bradford assay. A small amount (25 µg) of each sample was processed and digested with the Expedeon FASP trypsin digestion kit (<http://www.expedeon.com>). Final digests were cleaned via C18 tips, and peptide concentrations were measured by a Nanodrop spectrophotometer. Samples were diluted to 100 ng/µl final peptide concentration and 500 ng (5 µl) of each sample was run three times on 240 min LC–MS/MS gradients (Eksigent nano LC and Q Exactive+ MS). Data from each run were searched using MaxQuant version 1.5.0.25 and Perseus version 1.5.0.15 against maize peptide and common contaminants databases. Samples were run in triplicate in MaxQuant. label-free quantification values (minimum ratio count = 1) and raw intensity values were calculated. Results were imported into the Perseus Maxquant module. Reverse hits, contaminants, and identifications by site were removed. Abundances were log₂-transformed.

qRT–PCR. Pollen and post-dehiscence anthers (with pollen removed) were sampled in quadruplicate from NP2222, NP2222-HI, and *MTL*^{TAL-FS} to query *Mtl* transcript abundance by quantitative reverse transcriptase polymerase chain reaction (qRT–PCR). To quantify *Mtl*, primers specific to *Mtl* but agnostic with regard to the 4 bp mutation and the splice site were used ('generic qRT F1' and 'generic qRT R1') (see Supplementary Table 6). Transcript abundance was compared with an endogenous control to standardize for starting cDNA amounts, and the relative quantity was plotted ($\frac{1}{2}^{ΔC_t}$). The endogenous control gene had slightly higher expression in pollen than post-dehiscence anthers but this fact did not affect the data. The primers exhibited normal-looking amplification curves.

Recombinant MTL phospholipase activity assays. Recombinant wild-type and mutant MTL protein was produced and isolated from *Escherichia coli* by GenScript (<http://www.genscript.com>), and resuspended in a specially formulated buffer (50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 1 mM CaCl₂, 0.5% CTAB, pH 8.0). The proteins were then tested for phospholipase activity using an EnzChek Phospholipase A₂ Assay Kit (Invitrogen catalogue number E10217). Because CTAB is known to inhibit the fluorescent readings of this kit, a standard curve was generated using an assay buffer that included the same final concentration of CTAB (0.005%) present in the recombinant MTL assays. This standard curve was then used to calculate the units per milligram per minute of phospholipase (sn-2 type) activity of wild-type and mutant MTL. Four technical replicates were performed for wild-type and mutant MTL, and the entire assay, including the standard curve, was tested three times yielding very similar results, although we only report one set of those results in Extended Data Fig. 3b.

Pollen germination and tube growth assays. Pollen germination assays were conducted on liquid pollen germination medium (10% sucrose, 0.01% boric acid, 0.1% yeast extract, 10 mM CaCl₂, 50 µM KH₂PO₄, 15% PEG 4000)³². The medium was prepared freshly and kept cold and in the dark until use. Twenty-five millilitres of the medium was added to 10 cm Petri dishes, which were then taken to a laboratory room next to the greenhouse. Fresh pollen samples were then taken from three biological replicates (separate plants) in the greenhouse of RWK, Stock 6, RWK-NIL, NP2222, NP2222-HI, and *MTL*^{TAL-FS}. For each sample, approximately 50 mg of pollen was weighed using an Eppendorf tube and added to a single Petri dish containing the 25 ml of medium. Each sample of pollen was added to the medium within 3 min of collection from the tassel. Samples were collected between 08:00 and 10:00 on the same day and the time the pollen was added to each Petri dish was marked on the dish. Sixty minutes after the start time, the rate of pollen germination was evaluated by counting 200 pollen grains and scoring as germinated (a normal-looking pollen tube was growing), as experiencing osmotic shock (no tube is present, but evidence is seen of pollen grain bursting, defined as expulsion of material out the annulus without germination), or as non-germinated (no response seen). Exactly 3 h after germination, the germinated pollen samples were centrifuged at 500 r.p.m. for 10 min at 4 °C and resuspended in cold FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol). The samples were fixed in FAA on ice with vacuum infiltration for at least 5 min, followed by at least 4 h

of incubation at 4 °C. Fixed pollen grains and tubes were then gently centrifuged, resuspended in cold PBS, and pipetted onto slides and imaged on the Zeiss laser scanning microscope (LSM 710). At least 25 tubes per sample were evaluated for tube length by capturing 20× bright-field images and performing segment length analysis on ZEN 2011 software.

Pollen competition assays. In the RWK pollen competition assay, inducer and non-inducer pollen were applied to NP2222 silks, and 4-week-old progeny were evaluated for the genotype of the male parent on the basis of a colour assay. For each cross, ~0.1 ml of RWK or RWK-NIL pollen and ~0.1 ml of non-inducer pollen from a dominant purple-stemmed line (ZAN21PR) were collected at the same time, mixed, and applied to three NP2222 ears each. To ensure equal volume meant equal pollen grain count, RWK, RWK-NIL, and ZAN21PR grain size was checked with KI stain, bright-field imaging, and ZEN 2011 software. All three genotypes were found to have pollen grains that were between 80 and 95 µm in diameter (data not shown). Six pollinations were made, and nearly 300 progeny from the six ears were evaluated for the purple stem trait. The seed set on all the resulting ears was >80%, and the embryo abortion rate was around 20% (see Extended Data Table 2).

For the *MTL* pollen competition assay, three T1 NP2222-HI plants that were complemented with one copy of the *MTL* transgene were crossed onto one NP2222 ear each. Over 400 embryos were evaluated for the presence of the *MTL* transgene by Taqman assay to detect the tDNA. Because the male parent was hemizygous for *MTL* but homozygous for *mtl* at the native locus, 50% of the pollen grains were effectively *mtl* (or, haploid inducer) and 50% were effectively *MTL* (assuming normal expression and functioning of the transgene). The seed set on all the resulting ears was >80%, the embryo abortion rate was around 20%, and the HIR was around 2–3%.

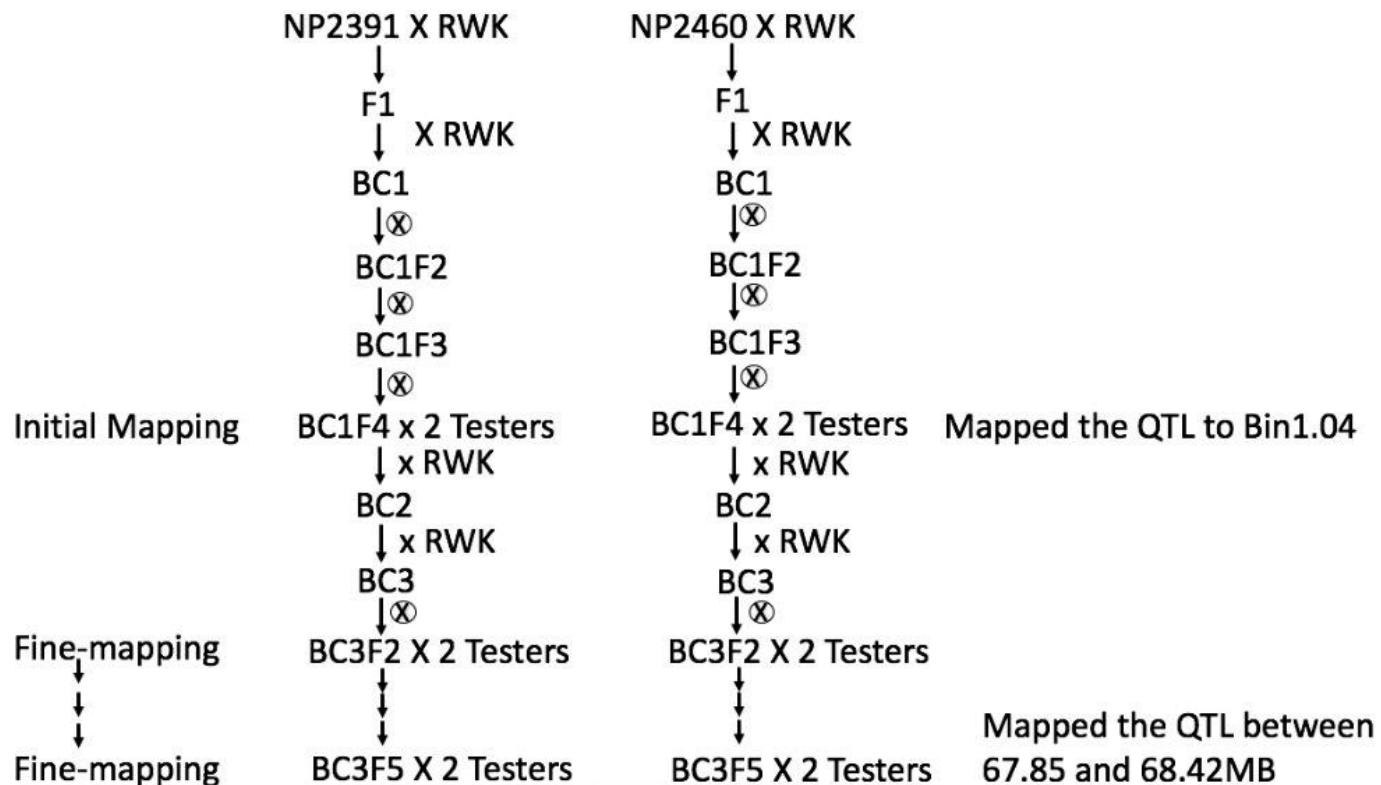
RNA-seq profiling and analysis. About 200 mg of fresh pollen was collected between 09:00 and 09:30 from four biological replicates of RWK, RWK-NIL, NP2222, and NP2222-HI lines, frozen on liquid nitrogen, and shipped to Data2Bio for RNA-seq analysis. Pollen samples were ground in liquid nitrogen using an RNase-free mortar and pestle. An RNeasy Plant Mini Kit from Qiagen was used for RNA extraction. Lysis buffer was added to ground pollen followed by additional grinding at room temperature. Once fully ground, samples were processed using the RNeasy extraction protocol, and RNA was quality checked using an Agilent RNA Nano kit. One library was constructed for each sample (that is, 16 libraries in total) and sequenced in two Illumina Hi-Seq paired-end lanes (2 × 100 bp). Raw reads were subjected to quality checks and trimming. About 1% of raw reads were dropped and 95.0–95.7% of base pairs remained after trimming in each pair. Trimmed reads were aligned to the public B73 RefGen_v2 (5b60) maize reference genome using GSnap³⁵. Only uniquely aligned reads were used to obtain read counts per gene. Gene counts were generated for public B73 RefGen_v2 nuclear gene models, which were then mapped to RefGen version 3 genes using information obtained from MaizeGDB (<http://www.maizegdb.org/assembly/#v2v3change>). The R package DESeq2 (<http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>) was used to test the null hypothesis that expression of a given gene was not different between the two genotypes. *P* values of all the statistical tests were converted to adjusted *P* values (*q* values)³⁶. A false discovery rate of 1% (*q* value) was used to account for multiple testing. Only those genes that were found to be expressed in at least four samples and having a minimum average read count of five were classified as 'informative' and used for differential expression analysis.

Tissue specificity analysis of upregulated set of differentially expressed genes in haploid inducer lines. Public maize B73 gene atlas RNASeq data³⁷ were mapped to public V3 genes using information from MaizeGDB (<http://www.maizegdb.org/assembly/#v2v3change>). Tissue specificity was calculated according to established techniques²⁷. Entropy values were binned for all genes and those with the lowest 5% entropy were cross-referenced with those having the lowest categorical tissue specificity/highest expression in R1-staged anthers (anthesis-staged anthers, when mature pollen is about to be shed). The genes were classified as R1-anther-specific genes. For the 28 genes upregulated in both haploid induction lines outside the *qh1r* region, 15 of 28 (54%) were specific for R1 anthers. A Fisher's exact test comparing this figure with the percentage of anther-specific genes in the maize genome (698/39,360, or 1.77%) indicated that the genes upregulated in both inducer lines were significantly enriched for anther-specific genes ($P < 2.2 \times 10^{-16}$).

Data availability. The RNA-seq data that support the findings of this study have been deposited in the NCBI SRA BioProject database under accession number PRJNA348129. All other data are available from the corresponding author upon reasonable request.

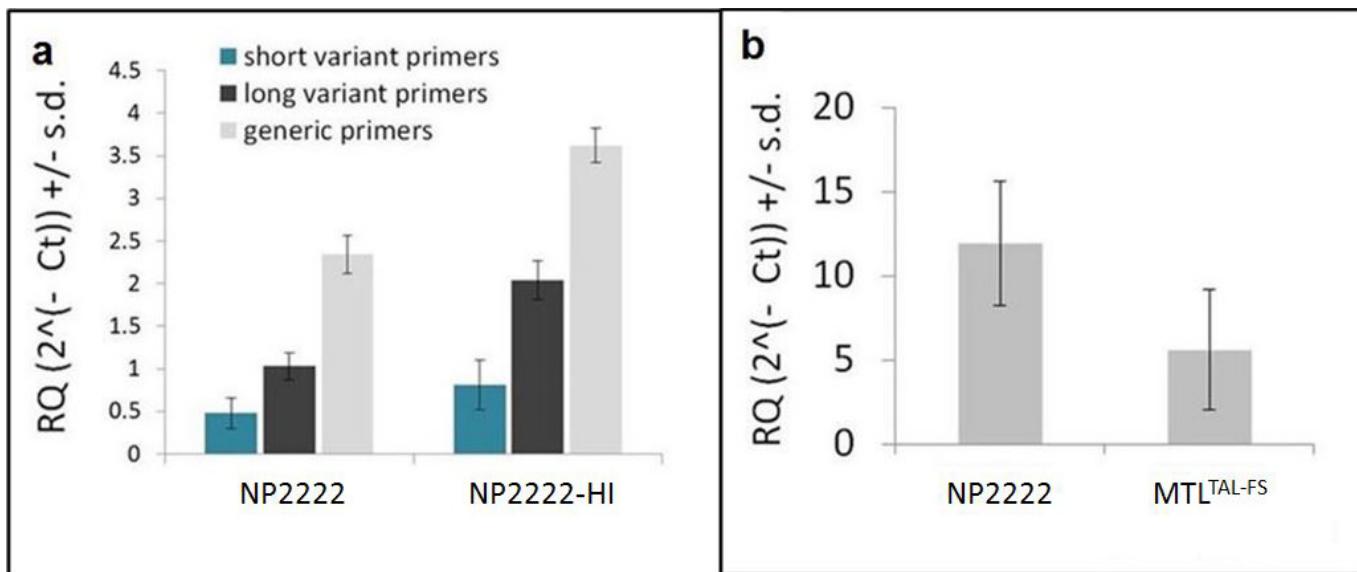
31. Negrotto, D. et al. The use of phosphomannose isomerase as a selectable marker to recover transgenic maize plants (*Zea mays* L.) via Agrobacterium transformation. *Plant Cell Rep.* **19**, 798–803 (2000).

32. Mascarenhas, J. P. Pollen tube growth and RNA synthesis by vegetative and generative nuclei of *Tradescantia*. *Am. J. Bot.* **53**, 563–569 (1966).
33. Kurihara, D., Mizuta, Y., Sato, Y. & Higashiyama, T. ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. *Development* **142**, 4168–4179 (2015).
34. Kelliher, T. & Walbot, V. Emergence and patterning of the five cell types of the *Zea mays* anther locule. *Dev. Biol.* **350**, 32–49 (2011).
35. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
36. Benjamini, Y. & Hochberg, Y. Controlling false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**, 289–300 (1995).
37. Stelpflug, S. C. *et al.* An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* **9**, 1–15 (2016).



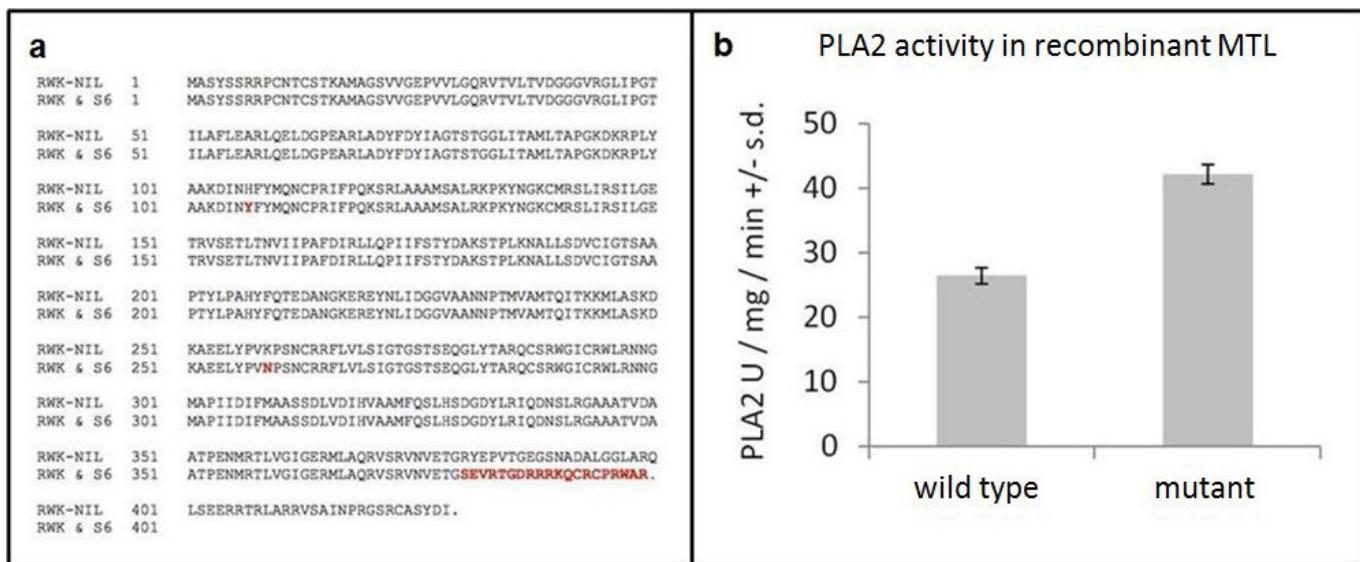
Extended Data Figure 1 | The Stock 6 derivative line RWK was obtained from the University of Hohenheim in 2006 and subsequently crossed to inbreds NP2460 and NP2391 and backcrossed to RWK to generate **mapping populations**. HIR testing and marker analysis were performed on both sets of mapping populations. HIRs in both

populations co-segregated with the marker SM020SDQ on the long arm of chromosome 1. Several rounds of fine mapping in RWK-GP6664 BC3F3–BC3F5 populations narrowed the quantitative trait locus to a ~0.57 kb interval between SM2589 and SM2608.



Extended Data Figure 2 | Quantification of *Mtl* RNA in haploid inducer and non-inducer pollen. **a**, Splice-specific qRT-PCR results for *Mtl* transcripts. Three biological replicates of R1-staged anthers were tested in technical triplicate, and the average C_t and s.d. was calculated for each reaction. The relative quantity of each transcript type was compared with the endogenous control using a \log_2 regression of ΔC_t . Two sets of primers were used to assess the relative abundance of each of the two annotated splice variants compared with a primer set that was agnostic with respect to the splice variants. The shorter transcript variant had relatively low

abundance compared with the long transcript in both NP2222 (wild-type) and NP2222-HI (haploid inducer) genotypes. Expression of the mutant copies of the gene in NP2222-HI was significantly higher for all three primer pairs tested. **b**, Five biological replicates of fresh pollen from NP2222 and MTL^{TAL-FS} lines were tested in technical triplicate on the generic primer, and the average C_t and s.d. were calculated for each reaction. The relative quantity of each transcript type was compared with the endogenous control using a \log_2 regression of ΔC_t . MTL^{TAL-FS} pollen has lower transcript abundance than NP2222 (wild-type) pollen.



Extended Data Figure 3 | MTL protein sequence and activity. **a**, An amino-acid alignment of the B73 predicted protein sequence of the long splice variant of the GRMZM2G471240 gene in B73 and RWK-NIL, with the predicted sequence of the *mtl* allele found in RWK and Stock 6 (S6). Amino acids that differ are in red; amino acids that match are indicated in normal grey text; stop codons are indicated with a full stop. Two point mutations result in amino-acid substitutions: a histidine (H) to a tyrosine (Y), and a lysine (K) to an arginine (N). These changes are not

conservative; it is possible that one or both of these modifies the haploid induction phenotype—suggesting that an allelic series could be uncovered with further investigation of variants. **b**, Wild-type MTL and mutant (truncated) MTL encoded by the *mtl* allele have *in vitro* phospholipase activity. PLA2 phospholipase activity as measured by fluorescent liposome assay on recombinant, purified protein produced using the *MTL* and *mtl* cDNAs. Error bars, s.e.m. based on the average of four replicates.

Zea mays	-----MAS-YSSRRPCNTCSTKAMAGSVVGEPV-VLGQRVTVLTVDGGGVRGLIPGTILAFLEARLQELDGPEARLADYFDYIAGTST
Sorghum bicolor	-----MATYSSRRPCNCACRTKAMAGSVVGEPV-VLGQRVTVLTVDGGGIRGLIPGTILAFLEARLQELDGPEVRLADYFDYIAGTST
Setaria italica	-----MAS-YSSRRPCNCACRTKAMAGSVVGEPV-VPGQRVTVLTI DGGGIRGLIPGTILAFLEARLQELDGPEARLADYFDIAGTST
Hordeum vulgare	-----MAS-YCRRPCCESCTRAGSVVGEPV-APGQRVTVLTI DGGGIRGLIPGTILAFLEARLQELDGPDARLADYFDIAGTST
Brachypodium distachyon	-----MAAS-YSCRRTCEACSTRAGCVVGEPASAPGQRVTLLAIDGGGIRGLIPGTILAFLEARLQELDGPDARLADYFDIAGTST
Oryza sativa v. japonica	-----MAGCVVGEPASAPGQRVTLLAIDGGGIRGLIPGTILAFLEARLQELDGPDARLADYFDIAGTST
Oryza sativa v. indica	-----
Triticum aestivum	-----MASDQTTPPAT-----ADAPISTPPP-SFGKRVTVLCIDGGGVRGLIPATIAFLEALQELDGPFAARIADYFDVIAGTST
Musa acuminata	-----MGS-----NESPVANPPP-CKGKVVTILSIDGGGVRGIIPGTILEFLEAKLQELDGPDARLADYFDIIAGTST
Elaeis guineensis	-----MQ-----MDS-----PKSPHQ-PP-TYGNLVTLISIDGGGIRGLIPAVILGFLESELQELDGEEARLADYFDVIAGTST
Arabidopsis thaliana	-----
Zea mays	-----GGLITAML TAPGKDKRPL YAAKDIHNFYMNCPRIFPQK-----SRLAAAMSALRKPKYNKGCMRSLIRSILGETRVSETLTNVII PAFD
Sorghum bicolor	-----GGLITAML TAPGKDKRPL YAAKDIQFYMCNCPRIFFPQK-----SSRLAAAMSALRKPRYNKGCLRNLI MSLGETRVSDLTNVIIPTFD
Setaria italica	-----GGLITAMIT PGEKDRPLFAARDINRFYFDNCPRIFPQS-----RSSLAAAMSALRKPRYNKGYLRLSTIRSMGETRVSDALTNVVIPTFD
Hordeum vulgare	-----GGLITAMIT PGEKDRPLFAARDINRFYFDNCPRIFPQS-----RCALAAVITASLRPRYNGSKYLHGKIRSMGETRLCDLTDVVIPTFD
Brachypodium distachyon	-----GGLITAMIT PGEKDRPLFAAADINRFYLDNGPQIFPQK-----RSSLMSVLA SLTRPRYNGKFLHGKIRSMGETRVCDLTDVVIPTFD
Oryza sativa v. indica	-----GGLITAMIT PGEKDRPLFAAADINRFYLDNGPQIFPQK-----RCGMAAAA MALT RPRYNGKYLQGKIRKMLGETRVDTLTNVVIPTFD
Oryza sativa v. japonica	-----GGLITAMIT PGEKDRPLFAAADINRFYLDNGPQIFPQK-----MLGETRLSDALT DVVIPTFD
Triticum aestivum	-----GGLITAMIT PGEKDRPLFAAADINRFYLDNGPQIFPQK-----RCGMAAAA MALT RPRYNGKYLQGKIRKMLGETRVDTLTNVVIPTFD
Musa acuminata	-----GGLITAMIT PGEKDRPLFAAADINRFYLDNGPQIFPQK-----GGLVTTML TAPNKDNRPLFSAKDIIQFYLENCPKIFPQR-TGLLAGALNLF GAVSGPKYDGKFLHSKVKE LGD TKLHQTLTNIVI PAFD
Elaeis guineensis	-----GGLVTTML TAPNKDNRPLFSAKDIIQFYLENCPKIFPQR-HFFPSAAKKLVSLTGPKYDGKYLHQLIHAKLGDTKLSQLTLNVVIPTFD
Arabidopsis thaliana	-----GGLVTTML TAPNKDNRPLFSAKDIIQFYLENCPKIFPQR-----
Zea mays	-----IKLQPIIFSTYDAKSTPLKNALLSDVCIGTSAAPTYLPAHYFQTEDA-NGKEREYNLIDGGVAANNPTMVAMTQITKKMLASKDAEEL
Sorghum bicolor	-----VRLQPIIFSTYDAKSMPLKNALLSDVCIGTSAAPTYLPAHYFQTKDAGSGKEREYNLIDGGVAANNPTMVAMTQITKKMLASKEAEL
Setaria italica	-----IKLQPIIFSTYDVKNMP LKNALLSDVCISTSAAPTYLPAHYFOIQDA-GKKTREYNLIDGGVAANNPTMVAMTQITKKMLAKDK--EEL
Hordeum vulgare	-----VRLQPIIFSTYDARNMPLKNARLADICIGTSAAPTYLPAHFHTQDD-NGKEREYNLIDGGVAANNPTMVAMTQITKKMMVKDR--EEL
Brachypodium distachyon	-----VRLQPIIFSTYDAKSMPLKNALLSDVCISTSAAPTYLPAHFQTEDD-NGKVR EYNLIDGGVAANNPTMVAMTQITKKIMVKDK--EEL
Oryza sativa v. indica	-----VRLQPIIFSTYDAKSMPLKNALLSDCICISTSAAPTYLPAHCFCQTTDDATGKVREFDLIDGGVAANNPTMVAMTQITKKIMVKDK--EEL
Oryza sativa v. japonica	-----VRLQPIIFSTYDAKSMPLKNALLSDCICISTSAAPTYLPAHCFCQTTDDATGKVREFDLIDGGVAANNPTMVAMTQITKKIMVKDK--EEL
Triticum aestivum	-----VRLQPIIFSTYDAKSMPLKNARLADVCIGTSAAPTYLPAHFHTHDG-NGKEREYNLIDGGVAANNPTMVAMTQITKKMMVKDR--EEL
Musa acuminata	-----IKFLQPTIFSTYQT KSTPLKD ALLSDCICISTSAAPTYLPAH FETKDD-NGNKRFSNLVDGGVAANNPTL TAMEVSKEILSN--PDF
Elaeis guineensis	-----IKLLQPVIFSTFETKTDPSKD ALLSDCICISTSAAPTYLPAH FETKDS-QGKTRSFNLVDGGVAANNPMIATQSITKQI FWN N--EDE
Arabidopsis thaliana	-----IKHLQPTIFSSYEVKNHPLKDA ALADIAISTSAAPTYLPAHFFKVEDL-NGNAKEYNLIDGGVAANNPALLAIGEV TNEISGG S--SDF
Zea mays	-----YPVKPSNCRRFLVLSIGTGSTSEQGLYTARQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HVAAMFQSLHSDGD-YLRIQDNSLRGA
Sorghum bicolor	-----YPVKPWNCRKFLVLSIGTGSTSEQGLYTARQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HVAAMFQSLHSDGD-YLRIQDNSLRGA
Setaria italica	-----YPVKPEDCRKFLVLSIGTGSTSEQGLYTARQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-----HSLRG A
Hordeum vulgare	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Brachypodium distachyon	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Oryza sativa v. indica	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Oryza sativa v. japonica	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Triticum aestivum	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Musa acuminata	-----YPVKPSDCGKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Elaeis guineensis	-----FSYQPVEYDRFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Arabidopsis thaliana	-----SKFKPTDFAKFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Zea mays	-----FPIRPNDYGRFLVLSIGTGSTSDQGLYTAKQCSRWGICRWLNNGM APIIIDIFMAASSDLVDI HAAVLFQSLHSDGD-YLRIQDNSLRGA
Sorghum bicolor	-----AATVDAATPENMRTL VLGIGERMIAQRVS RVNVETGRYEPVTGE GSNA DALGGLARQLSEERRTRLARRVSAINPRGS---RCAS--YDI
Setaria italica	-----AATVDAATPENMRTL VLGIGERMIAQRVS RVNVETGRYEPVTGE GSNA DALAGIARQLSEERRTRLARRTS AIVSSGGASRRTCA SKVSN
Hordeum vulgare	-----AATVDAATPENMRTL VLGIGERMIAQRVS RVNVETGRYEPVTGE GSNA DALVALARQLSDERRARIARRAACAGGS---RCSP-VKT
Brachypodium distachyon	-----AATVDAATPENMAELLRIGERMIAQRVS RVNVETGRYEEIRGGSNA DALAGFAKQLSDERRTRLGRRRVGAGR LKS--RR-----
Oryza sativa v. indica	-----AATVDTATPDNMRELVRIGERMIAQRVS KVNVETGRYEEQM QAGTNADALAGFARQLSDERRARFG PRDGAPANGS---RC-----
Oryza sativa v. japonica	-----AATVDAATRDNMRALVGIGERMIAQRVS KVNVETGRYEEVPGAGSNADALRG FARQLSEERRARLGRRNACGGGGEG---EPSGVACKR
Triticum aestivum	-----AATVDAATRDNMRALVGIGERMIAQRVS KVNVETGRYEEVPGAGSNADALRG FARQLSEERRARLGRRNACGGGGEG---EPSGVACKR
Musa acuminata	-----AATVDAATPENMAELLRIGERMIAQRVS RVNVETGRYEEVKGAGNNADALAGFARQLSDERRTRLGRRNACGGGGEG-SR-----
Elaeis guineensis	-----TSSDVSTKKNLQDLV DIGNSLKKPVS RVNIE GTGHSE A DGE GTNE A ALTGFAKKL SDE RRRQS KQLTSSDATQH-----
Arabidopsis thaliana	-----TASDVSTSENLRKLVQVQGDLLKKPVS RVNLETGVSEACDVEGTNE DALIRFAKMLSNERKS RNAKMSAA-----
	-----AASV DIATVENL DILAKT GDELLKPVARVNLD SGCG NENA-YETTNE HALIKLAGL ISKEKKI DIRSPHAKAPIR I-----

Extended Data Figure 4 | Amino-acid alignment of the publicly available MTL orthologues in eight grasses, two non-grass monocots, and thale cress. This alignment includes maize (*Z. mays*), sorghum (*Sorghum bicolor*, 92% sequence identity to MTL), foxtail millet (*Setaria italica*, 85% identity), barley (*Hordeum vulgare*, 78% identity), *Brachypodium distachyon* (78% identity), Indica and Japonica variety rice (*Oryza sativa* var. *indica* and *japonica*, Os3g27610, 78 and 79% identity, respectively), bread wheat (*Triticum aestivum*, 55% identity), banana (*Musa acuminata*, 57% identity), oil palm (*Elaeis guineensis*), and *Arabidopsis thaliana* (52% identity).

It is clear that this gene is highly conserved in the cereals, but less conserved in more distant monocots and dicots. The N terminus is missing in wheat but the meaning of this is not clear. Expression data also indicate that the rice orthologue, called PLAII β , is specifically expressed in pollen³⁰, while the closest *Arabidopsis* homologue by amino-acid conservation, also known as *PLP2*, is expressed in leaves and thus is not functionally conserved. It is worth noting that the lysine (K) residue that is changed to an arginine (N) in haploid inducer lines is a conserved amino acid in the grasses.

Extended Data Table 1 | Average HIRs, origin details, and haplotype data for 19 inducer and 9 non-inducers queried for GRMZM2G471240 and GRMZM2G062320 (PGM)

			Gene	GRMZM2G062320	GRMZM2G471240	GRMZM2G471240
Inducers	HIR	Description	Primer names	nil.F1 and nil.R1	rwk.F1 and rwk.R1	nil.F1 and nil.R1
			Sequences used to design	B73 / RWK-NIL	RWK / STOCK6	B73 / RWK-NIL
ZMS	7.0%	Zarodishev Marker Saratov		-	+	-
Z22	7.0%	Inbred selfed from ZMS		-	+	-
Z21	7.0%	87.5% Z22 + 12.5% NP2276		-	+	-
Z19-PR	7.0%	87.5% Z22 + 12.5% NP2276 with B1PI1		-	+	-
ZI29	7.0%	87.5% Z22 + 12.5% NP2052		-	+	-
ZR75	8.0%	50% Z21 + 50% RWS		-	+	-
K13	9.0%	Selfed from synthetic from Krasnodar Institute		-	+	-
ZA19-PR	9.5%	75% Z19 + 25% NPH8431 with B1PI1		-	+	-
RWS	10.0%	From University of Hohenheim		-	+	-
RWK	10.0%	From University of Hohenheim		-	+	-
NP2222-HI	10.2%	BC3 NIL to NP2222, RWK haplotypes in qhir1		-	+	-
ZR86	12.0%	75% RWS + 25% Z21		-	+	-
ZR53	12.0%	50% RWS + 50% Z21		-	+	-
ZR75	13.0%	50% RWS + 50% Z21		-	+	-
Poor Inducers						
Stock 6	2.5%	Seed from Ed Coe		-	+	-
Stock6 R1-nj	2.5%	Stock6 with R1-nj		-	+	-
ZR33	3.0%	75% RWS + 25% Z21		-	+	-
KZ2	3.0%	75% K13 + 25% ZMS with B1PI1		-	+	-
ZMS-PR	3.0%	ZMS with B1PI1		-	+	-
Non-inducers						
NP2460	<1%	NP2460 - non-inducer		+	-	+
NP2391	<1%	NP2461 - non-inducer		+	-	+
NP2276	<1%	NP2276 - non-inducer		+	-	+
NP2052	<1%	NP2052 - non-inducer		+	-	+
NPH8431PR	<1%	NPH8431 with B1PI1 - non-inducer		+	-	+
NP2222	<1%	Syngenta's standard transformable inbred		+	-	+
ZAN21PR	<1%	75% Z21 + 25% NPH8431PR - non-inducer		+	-	+
Stock6 R1-nj E	<1%	Stock6 with R1-nj B1 PI1		+	-	+
RWK-NIL	<1%	RWK NIL with NP2391 haplotypes in qhir1		+	-	+

A + indicates a positive PCR result (a band in the gel); - indicates a negative PCR result (no band)

Extended Data Table 2 | Reproductive phenotypes in complementation and pollen competition assays

Complementation Assays			Kernel characteristics				Embryos tested			
Male parent	Ind ID	event	ears	viable	aborted	% aborted	embryos	haploids	diploids	HIR
NP2222-HI		N/A	4	548	498	47.61%	531	54	477	10.17%
NP2222-HI + 471240 / 471240*	8399	006	4	1389	12	0.86%	1369	2	1367	0.15%
NP2222-HI + 471240 / 471240	8401	006	3	866	6	0.69%	854	3	851	0.35%
NP2222-HI + 471240 / 471240	3583	024	3	842	5	0.59%	824	3	821	0.36%
NP2222-HI + 471240 / 471240	3534	024	3	945	2	0.21%	924	2	922	0.22%
NP2222-HI + 471240 / 471240	3667	005	4	361	4	1.10%	350	1	349	0.29%
Totals			17	4403	29	0.65%	4321	11	4310	0.25%
NP2222-HI + PGM / PGM†	4198	002	4	510	415	44.86%	495	47	448	9.49%
NP2222-HI + PGM / PGM	4022	025	3	436	420	49.07%	415	44	371	10.60%
NP2222-HI + PGM / PGM	4179	041	3	635	571	47.35%	598	58	540	9.70%
Totals			10	1581	1406	47.07%	1508	149	1359	9.88%
NP2222-HI + mtl-GFP / mtl-GFP	5722	006A	1	152	109	41.76%	148	16	132	10.81%
NP2222-HI + mtl-GFP / mtl-GFP	5800	006A	1	76	67	46.85%	75	5	70	6.67%
NP2222-HI + mtl-GFP / mtl-GFP	5801	006A	1	143	122	46.04%	137	13	124	9.49%
Totals			3	371	298	44.54%	360	34	326	9.44%
NP2222-HI + MTL-GFP / MTL-GFP	5668	004A	1	253	12	4.53%	244	2	242	0.82%
NP2222-HI + MTL-GFP / MTL-GFP	5673	004A	1	452	37	7.57%	287	2	285	0.70%
NP2222-HI + MTL-GFP / MTL-GFP	5754	004A	1	314	3	0.95%	305	1	304	0.33%
Totals			3	1019	52	4.86%	836	5	831	0.60%
Pollen Competition Assays			Kernel characteristics				Embryos tested			
Male parent	Ind ID	event	ears	viable	aborted	% aborted	embryos	haploids	diploids	HIR
NP2222-HI + 471240 / 0 (hemizygous)	3598	024	1	324	38	10.50%	316	6	310	1.90%
NP2222-HI + 471240 / 0 (hemizygous)	3532	024	1	263	62	19.08%	249	7	242	2.81%
NP2222-HI + 471240 / 0 (hemizygous)	8404	006	1	180	45	20.00%	175	5	170	2.86%
50 % RWK and 50 % ZAN21PR	n.a.	n.a.	1	315	42	11.76%	n.d.	n.d.	n.d.	n.d.
50 % RWK and 50 % ZAN21PR	n.a.	n.a.	1	228	67	22.71%	n.d.	n.d.	n.d.	n.d.
50 % RWK and 50 % ZAN21PR	n.a.	n.a.	1	288	52	15.29%	n.d.	n.d.	n.d.	n.d.
50 % RWK-NIL and 50 % ZAN21PR	n.a.	n.a.	1	423	6	1.40%	n.d.	n.d.	n.d.	n.d.
50 % RWK-NIL and 50 % ZAN21PR	n.a.	n.a.	1	412	5	1.20%	n.d.	n.d.	n.d.	n.d.
50 % RWK-NIL and 50 % ZAN21PR	n.a.	n.a.	1	386	5	1.28%	n.d.	n.d.	n.d.	n.d.

*471240, GRMZM2G471240; †PGM, GRMZM2G062320

Top: haploid induction tests for T1 NP2222-HI plants homozygous for wild-type GRMZM2G471240 or GRMZM2G062320 (PGM) transgenes under the expression control of their native promoters, as well as for T1 NP2222-HI plants homozygous for MTL-GFP or mtl-GFP transgenes under the control of the MTL native promoter. Homozygous MTL-GFP and GRMZM2G471240 (MTL) restored normal reproduction in NP2222-HI, but GRMZM2G062320 (PGM) and mtl-GFP did not. Bottom: embryo abortion and haploid induction results from the pollen competition assays.

Extended Data Table 3 | Haploid induction test crosses of the GRMZM2G471240^{RNAi} and GRMZM2G062320^{RNAi} lines

GRMZM2G471240 RNAi		Kernel Characteristics				Embryos tested for ploidy			
Individual ID	Event ID	ears	viable	aborted	% aborted	embryos	haploids	diploids	HIR
5148	001	2	701	43	5.78%	369	3	366	0.81%
5149	001	2	186	22	10.58%	166	1	165	0.60%
5153	001	2	625	61	8.89%	323	7	316	2.17%
5161	001	3	1116	87	7.23%	485	4	481	0.82%
5170	028	2	629	23	3.53%	324	1	323	0.31%
5173	028	2	551	33	5.65%	322	0	322	0.00%
5187	028	3	379	27	6.65%	333	9	324	2.70%
3731	014	2	894	23	2.51%	263	4	259	1.52%
3732	014	2	648	49	7.03%	351	0	351	0.00%
3736	007	1	277	21	7.05%	277	0	277	0.00%
3737	007	1	223	49	18.01%	175	3	172	1.71%
3751	005	1	133	6	4.32%	118	0	118	0.00%
TOTALS	5 events	23	6362	444	6.52%	3506	32	3474	0.91%
GRMZM2G062320 RNAi		Kernel Characteristics				Embryos tested for ploidy			
Individual ID	Event ID	ears	viable	aborted	% aborted	embryos	haploids	diploids	HIR
5195	929	2	316	17	5.11%	114	0	114	0.00%
5197	929	2	612	10	1.61%	176	0	176	0.00%
5203	929	2	748	20	2.60%	176	1	175	0.57%
5207	929	3	252	14	5.26%	248	1	247	0.40%
5210	929	3	1235	69	5.29%	258	0	258	0.00%
5213	929	2	347	22	5.96%	262	0	262	0.00%
5214	929	2	819	43	4.99%	176	0	176	0.00%
5225	931	3	815	39	4.57%	262	0	262	0.00%
5229	931	2	710	31	4.18%	314	0	314	0.00%
5233	931	3	1012	58	5.42%	440	2	438	0.45%
5239	931	3	915	25	2.66%	264	2	262	0.76%
5250	936	3	1135	20	1.73%	258	2	256	0.78%
5258	936	3	1228	67	5.17%	420	3	417	0.71%
5261	936	3	916	33	3.48%	260	1	259	0.38%
TOTALS	3 events	36	11060	468	4.06%	3628	12	3616	0.33%

Extended Data Table 4 | Haploid induction and kernel abortion rate data for several ears crossed by T1 plants biallelic for frame-shift mutations in GRMZM2G471240 (*MTL*) but lacking the TALEN tDNA

Event	Construct - individual	Mutation(s)	Ears	Kernel Characteristics			Ploidy Analysis Data			
				Avg. viable	Avg. aborted	Percent aborted	Total Embryos	Putative Haploids	Confirmed Haploids	HIR
39A	22808-3954	Biallelic (13 bp & 28 bp dels)	4	162	128	44.1%	579	37	35	6.04%
23A	22808-3924	Biallelic (8 bp & 5 bp dels)	2	114	116	50.4%	128	18	16	12.50%
81A	22808-3932	Homozygous (13 bp del)	2	165	129	43.9%	169	18	15	8.88%
81A	22808-3317	Homozygous (13 bp del)	2	183	108	37.1%	343	19	19	5.54%
81A	22808-3303	Homozygous (13 bp del)	1	189	100	34.6%	176	7	7	3.98%
38A	22808-4108	Biallelic (11 bp & 5 bp dels)	4	147	102	40.1%	379	28	26	6.86%
Total		Totals	15	160	113.833	41.6%	1774	127	118	6.65%

Extended Data Table 5 | Proteins off and on in NP2222 and NP2222-HI pollen samples, including MTL, which is found in NP2222 but not NP2222-HI pollen

	log2 LFQ						Majority protein ID	best BlastP match (S prot plants)		
	NP2222			NP2222-HI						
	rep 1	rep 2	rep 3	rep 1	rep 2	rep 3				
Absent in NP2222-HI	23.3	23.3	23.4	ND	ND	ND	GRMZM2G028905	L-fucose alpha-1,3-D-xylosyltransferase		
	22.9	24.1	23.9	ND	ND	ND	GRMZM2G046743	Lysine histidine transporter 1		
	23.2	23.1	23.2	ND	ND	ND	GRMZM2G310362	Polyadenylate-binding protein 5		
	24.2	24.2	24.1	ND	ND	ND	GRMZM2G130121	Chaperone protein ClpB2, chloroplastic		
	23.5	23.5	23.6	ND	ND	ND	GRMZM2G375807	ABC transporter D; COMATOSE		
	24.1	23.7	23.9	ND	ND	ND	GRMZM2G396212	Phospho-2-dehydro-3-deoxyheptonate aldolase 1		
	23.7	23.8	23.9	ND	ND	ND	GRMZM2G467907	RNA-binding protein 47		
	23.8	23.8	23.9	ND	ND	ND	GRMZM2G471240	Matrilineal		
	ND	ND	ND	23.9	23.8	23.9	GRMZM2G013607	Ferredoxin-6, chloroplastic		
	ND	ND	ND	22.1	22.1	22.2	GRMZM2G030971	Phospholipase A I		
Absent in NP2222	ND	ND	ND	24.5	24.6	24	GRMZM2G064967	Mannan endo-1,4-beta-mannosidase		
	ND	ND	ND	24	24.3	24.2	GRMZM2G143613	F-box protein		
	ND	ND	ND	24.4	24.3	24.1	GRMZM2G166906	HOTHEAD (synth long-chain a-dicarboxylic FAs)		
	ND	ND	ND	23.2	23.5	23.6	GRMZM2G181259	beta-D-xylosidase 2		

*ND, Not detected

Extended Data Table 6 | List of genes significantly downregulated in RWK and NP2222-HI pollen compared with their near isogenic lines

Gramene ID	Annotation	Log 2 value	
		RWK vs RWK-	NP2222-HI vs NP2222
AC207265.3_FG002	ENTH, 1-phosphatidyl inositol, clathrin adaptor	-1.29	-2.25
GRMZM2G008033	unannotated	-1.43	-4.04
GRMZM2G042154	peptide transporter PTR2	-1.22	-0.66
GRMZM2G042582	PRL3 adenosine 5'-phosphosulfate reductase	-0.5	-0.32
GRMZM2G050933	Cyclin D6	-2.2	-0.65
GRMZM2G081639	POLD1 - Putative DNA polymerase D, REV3	-0.98	-0.45
GRMZM2G094072	CFM2 (CRM family member 2)	-2.6	-0.89
GRMZM2G114113	expressed protein	-1.49	-1.13
GRMZM2G129354	ARF GAP-like Zinc finger protein	-0.54	-0.26
GRMZM2G145300	TRAF-like, MATH domain containing	-1.08	-6.24
GRMZM2G151564	calcium binding tetratricopeptide repeat	-0.64	-0.48
GRMZM2G156126	triacylglycerol lipase	-0.76	-1.32
GRMZM2G159155	Myb TF or DNA methyl transferase	-0.67	-1.18
GRMZM2G388892	DNA-directed RNA polymerase I subunit A2	-1.23	-0.55
GRMZM2G440198	PIF, ping pong	-0.97	-2.44
GRMZM5G808624	PANK2 pantothenate kinase 2	-0.9	-1.09
GRMZM5G809195	IAA14 auxin responsive	-1.59	-1.33
GRMZM5G828945	uncharacterized protein	-0.55	-0.34