

Cover Art



Adaptive introgression into maize at *ZmPLA1.2* controls phosphatidylcholine levels and induces earlier flowering

Adaptive introgression at *ZmPLA1.2* controls phosphatidylcholine levels and induces earlier flowering in maize

Fausto Rodríguez-Zapata^{a,b,1}, Allison C Barnes^{a,1}, Karla Blöcher-Juárez^{b,1}, Dan Gates^c, Andi Kur^a, Li Wang^d, Garrett M Janzen^d, Sarah Jensen^e, Juan M Estévez-Palmas^b, Taylor Crow^f, Rocío Aguilar-Rangel^b, Edgar Demesa-Arevalo^g, Sergio Pérez-Limón^b, Tara Skopelitis^g, David Jackson^g, Oliver Fiehn^h, Daniel Runcie^f, Edward S Buckler^e, Jeffrey Ross-Ibarra^c, Matthew Huford^d, Ruairidh JH Sawers^{b,i} and Rubén Rellán-Álvarez^{a, b,*}

^aDepartment of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, ^bNational Laboratory of Genomics for Biodiversity, Irapuato, México, ^cDepartment of Evolution and Ecology, Center for Population Biology and Genome Center, University of California, Davis, CA, ^dUS Department of Agriculture—Agricultural Research Service, Cornell University, Ithaca, NY, ^fDepartment of Plant Sciences, University of California, Davis, CA,

^dDepartment of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, USA, ^gCold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA,

^hWest Coast Metabolomics Center, University of California, Davis, CA, USA, ⁱDepartment of Plant Science, The Pennsylvania State University, PA, USA

After domestication from lowland teosinte in the warm, humid Mexican southwest, maize colonized the highlands of México and South America. In the highlands, maize was exposed to a range of novel environmental factors, including decreased phosphorous availability and lower temperatures that impose a strong selection on flowering time. Previous work has linked phospholipid metabolism to low phosphorous, low temperature stress and flowering time. Here, we combine linkage mapping and genome scans to identify *ZmPLA1.2*, a gene encoding a phospholipase A1 enzyme, as the major driver of phospholipid variation in highland maize. Our common garden experiments demonstrate that *ZmPLA1.2* exhibits a strong genotype by environment interaction and that the highland *ZmPLA1.2* allele leads to higher fitness possibly via a reduction of flowering time. We identify a putative loss of function allele in *ZmPLA1.2* that changes a strongly conserved amino acid. Metanalysis of bacterial growth conditions points to an conserved association between the identified amino acid and temperature, and our targeted CRISPR-CAS9 mutagenesis of *ZmPLA1.2* confirmed its effects on PC/LPC ratios and flowering time. Finally, we show that the highland *ZmPLA1.2* allele was introgressed from its wild relative teosinte *mexicana* and has been conserved in modern maize from the Northern US and Europe.

phospholipid metabolism; maize genetics; highland adaptation; convergent evolution

Elevation gradients are associated with changes in environmental factors that impose constraints on an organism's physiology. Organisms adapt to highland environments via selection of genetic variants that improve their physiological ability to cope with these constraints, including lower oxygen availability (1; 2; 3; 4), higher UV-radiation (5) and lower temperatures (6; 7). In particular, lower temperatures can significantly reduce growing season length and select for accelerated development.

After domestication from the wild relative teosinte *parviflora* (*Zea mays* spp. *parviflora*) (8; 9) in the lowland, subtropical environment of the Balsas River (Guerrero, México) around 9,000 BP, maize (*Zea mays* spp. *mays*) expanded throughout México and reached the highland valleys of central México around 6,500 BP (10).

In México, highland adaptation of maize was aided by significant adaptive introgression from a second subspecies of teosinte,

teosinte *mexicana* (*Zea mays* spp. *mexicana*) that had already adapted to the highlands of México thousands of years after the split from teosinte *parviflora* (11; 12). Phenotypically, the most evident signs of *mexicana* introgression into maize are the high levels of stem pigmentation and pubescence (13) that are supposed to protect against high UV radiation and low temperatures. The ability to withstand low temperatures and efficiently photosynthesize in early stages of seedling development is a key component of maize highland adaptation (14), and recent RNA-Seq analysis of the effects of the inversion *Inv4m*, introgressed from *mexicana*, support this hypothesis (15). *Inv4m* is also associated with shorter flowering times in highland maize (16; 17). Given the low growing degree unit accumulation in highland conditions, there has been selection for shorter flowering times in highland-adapted maize (17). Other possible adaptive traits originating from *mexicana* may include ability to grow in soils with low phosphorous (18; 19) that are characteristic of the acidic soils of the highland valleys of the Transmexican volcanic belt (20).

By the time that maize reached the Mexican highlands, its

Manuscript compiled: Friday 1st January, 2021

¹These authors contributed equally to this work.

^aDepartment of Molecular and Structural Biochemistry, North Carolina State University, 27607, Raleigh, NC. Email: rrellan@ncsu.edu

range had already expanded far to the south, including colonization of highland environments in the Andes (21; 22). Andean maize adaptation occurred without *mexicana* introgression, as there is no wild teosinte relative in South America. These multiple events of maize adaptation to highland environments constitute a good system to study the evolutionary and physiological mechanisms of convergent adaptation (23; 24).

In comparison to southward expansion, northward migration into the current United States, where summer daylength is longer compared to México, Central America, and South America, occurred at a much slower pace (25; 26). This is likely due to maize photoperiod sensitivity that leads to maladaptation in long day conditions (27). Indeed, a host of evidence suggests that maize cultivation in northern latitudes was enabled by selection of allelic variants of genes in the photoperiod pathway that lead to a reduction of photoperiod sensitivity and flowering time (28; 29; 30; 31; 32; 33; 27). There is evidence that early flowering alleles that confer an adaptive advantage in highland environments are also the result of highland *mexicana* introgression in highland maize (29) and were further selected in northern latitudes. Photoperiod insensitive maize from the Northern US and Canada was quickly introduced into Northern Europe as it was pre-adapted to northern latitudes and lower temperatures (34).

Plant glycerolipids, which include phospholipids, sulfolipids, galactolipids and other less polar lipids such as triacylglycerol, are involved in plant response to stresses typical of highland environments. For example, in low temperature, plants increase phospholipid concentration (35) and reduce the levels of unsaturated fatty acids in glycerolipids (36; 37), which may help maintain the fluidity of cell membranes. Under low phosphorus availability, plants tend to degrade phospholipids and increase the concentration of galactolipids and sulfolipids to free up phosphorus (38), particularly in older leaves. Finally, certain species of phosphatidylcholine (PC), the most abundant phospholipid (39), can bind to *Arabidopsis* flowering locus T and accelerate flowering time through a yet unknown mechanism (40), and glycerolipid content in maize has a predictive power for flowering time (41).

Furthermore, previous work in Mesoamerican and South American highland maize populations used F_{ST} statistics to identify loci that were under selection in both Mesoamerica and South America highland maize and showed little evidence of convergence between the two sub-continents but that glycerolipid pathways were some of the few pathways that were convergently selected in both highland environments (23).

In this paper we used several approaches to study the possible role of glycerolipid — and in particular phospholipid — metabolism in maize adaptation to highlands. Using measures of selection of at the gene and metabolite level in different maize landrace panels, we showed that pathways involved in the synthesis and degradation of phospholipids have clear signs of convergent selection in several highland population across the Americas. We identified that *ZmPLA1.2* (Phospholipase A1.2, Zm00001d039542) and *ZmLPCAT1* (lyso-phosphatidylcholine acyl-choline transferase 1, Zm00001d017584) showed strong, repeated signals of selection in maize adapted to several highland environments. *ZmPLA1.2* and *ZmLPCAT1* predicted enzymatic activities contribute to the synthesis and degradation of phosphatidylcholine and are compelling candidates to explain the high phosphatidylcholine (PC)/lyso-phosphatidylcholine(LPC) ratios we observed in highland mesoamerican landraces. In fact, QTL analysis of phospholipid content in a temperate by Mexican highland bi-

parental backcross revealed a major QTL explaining PC/LPC ratios that overlapped with *ZmPLA1.2* and showed that a loss of function in the highland allele leads to high PC/LPC ratios. We further confirmed this loss of function using a CRISPR-CAS9 knockout in a temperate inbred background that phenocopied the highland allele PC/LPC ratios. Using data from thousands of genotyped landrace testcrosses grown in common garden experiments at different elevations in México, we showed a strong genotype by environment effect of the *ZmPLA1.2* locus, where the highland allele leads to higher fitness in highland environments and reduced fitness at lower elevations. This GxE fitness effect is probably driven by the highland *ZmPLA1.2* allele that is associated with later flowering times in lowland environments and earlier flowering times in highland environments. *ZmPLA1.2* CRISPR-Cas9 (*ZmPLA1.2^{CR}*) mutants grown in similar conditions to lowland environments confirmed this effect. Lastly, we showed that the highland PT *ZmPLA1.2* locus is the result of teosinte *mexicana* introgression and that this introgression is further conserved in northern US and European Flints. These results suggest a potential beneficial effect of the *ZmPLA1.2* highland allele in cold, high latitude environments where early flowering time would be advantageous.

In summary, the results presented in this paper help us understand at the physiological and molecular level how an important crop like maize adapts to the unique conditions of highland environments, the role of wild relative introgression in this process, and the potential impact of this introgression in modern maize.

Results

Selection on glycerolipid metabolites

We grew a diversity panel composed of 120 highland and lowland landraces from mesoamerica and South America (Figure 1A) in highland and lowland Mexican common gardens and quantified phospholipid levels. Despite the intrinsic biological and environmental variability associated with analyzing open-pollinated varieties in field conditions, we could observe that mesoamerican highland landraces showed high PC/LPC ratios particularly when grown in highlands (Figure 1E). The differences observed in glycerolipid levels between highland and lowland maize could be the result of adaptive natural selection or random genetic drift in the process of maize adaptation to highland environments. To distinguish between these two competing scenarios, we compared each phenotype's population variance with genetic variance of neutral markers, an approach known as a Q_{ST} - F_{ST} comparison (42).

We calculated Q_{ST} - F_{ST} using DartSeq genotypic data from the same plants that were used to analyze glycerolipid levels, and we calculated the Q_{ST} - F_{ST} values for each glycerolipid species for highland/lowland populations of each continent. Mean Q_{ST} was greater than mean F_{ST} in mesoamerican and South American comparisons, though only the mesoamerican comparison is significant (two-tailed t-test, $p = 0.00073$; South American comparison $p = 0.12$). However, we observed several PC and LPC species with higher Q_{ST} values than the neutral F_{ST} in both sub continents (Supplementary Figure 1C-D). In particular, one of the species with the highest Q_{ST} values in both continents is PC-36:5.

Selection on phospholipid pathway genes

We used Genotyping By Sequencing (GBS) data from 2700 geo-referenced landraces from México generated by the SeeD project (16; 17) to run a *pcadapt* analysis that detects how strongly loci are contributing to patterns of differentiation between major

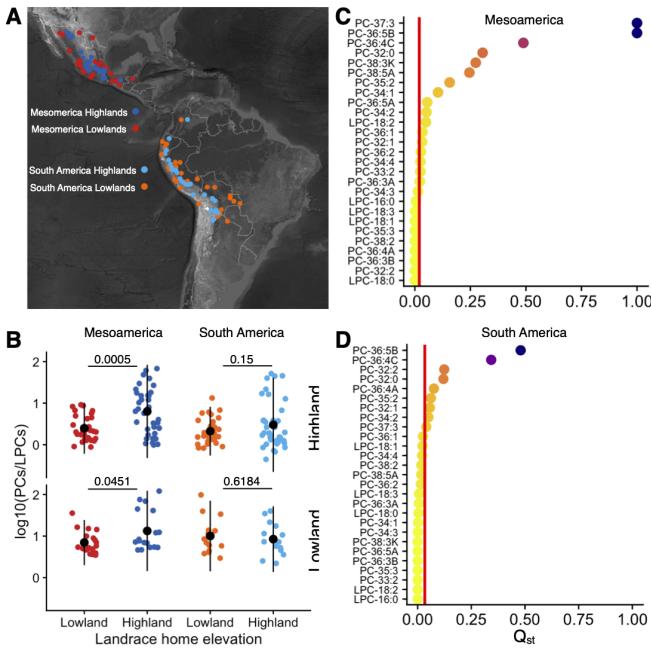


Figure 1 Phospholipid selection in highland maize. **A)** Map showing the geographical origin of the 120 accessions used in the common garden experiment to quantify glycerolipid levels. **B)** Logarithmic values of the PCs/LPCs ratio of highland and lowland landraces from MesoAmerica and South America. **C)** $Q_{ST}-F_{ST}$ analysis of phospholipid compounds between highland and lowland landraces from Mesoamerica **D)** and South America.

principal components of genetic variation (43). The first principal component of *pcadapt* polarized Mexican landraces based on elevation of the geographical origin of the landrace (Supplementary Figure S1A). Using this first principal component we identified outlier SNPs across the genome that are significantly associated with elevation of origin of the landrace and are potentially involved in elevation dependent local adaptation (Supplementary Figure S1B). We identified ≈ 600 genes involved in glycerolipid metabolism (See for details), and from those XX contained SNPs that were *pcadapt* PC1 outliers (top 5% -log(P)) (Supplementary File 1). We found that genes like *ZmPLA1.2* and *ZmLPCAT1* involved in phospholipid remodelling had SNPs within the coding sequence with significantly high -log(P) (Figure 2A-B) reflecting strong elevation dependent allele frequency changes (Figure 2C-D). We then used the same set of 600 genes involved in glycerolipid metabolism to identify selection signals using Population Branch Excess (PBE) (44) statistic across four (Southwestern US, Mexican Highland, Guatemalan Highlands and Andes) highland populations. We used PBE values already calculated by Li Wang and collaborators (24). We found that from the 153 glycerolipid-related genes that were PBE outliers in Mexican highlands 38 were also *pcadapt* PC1 outliers. We previously defined two possible explanations for the extent of convergent selection in genes repeatedly selected in several highland populations (24, 45). Convergent adaptation can be determined by a few number of genes that impose a *physiological* constraint on the genetic landscape routes that adaptation can take. On the other hand, convergent adaptation can potentially be determined by a large number of genes, but deleterious *pleiotropic* effects can constrain the number of genes that selection acts on. Using Yeaman's *et al.* C_{hyper} statistic

(45) that quantifies these two modes of convergent adaptation constraint, we found that the overlap observed among the presumably adapted genes in the four highland populations can not be explained just by physiological constraint ($C_{hyper} = 3.96$) and that most likely it includes certain degree of pleiotropic constraint. The overlap was higher for the US-Mesoamerica population pairs ($C_{hyper} = 4.79$), than between the Andean and US-mesoamerica pair ($C_{hyper} = 3.14$). These results are similar to our analysis in the flowering time pathway in the same populations (24). We identified a significant excess of genes that were targets of selection in more than two populations $P < 3 \times 10^{-5}$, Supplementary Figure S1C). The most over-represented intersection of selected glycerolipid genes was SWUS, MH, and GH ($p = 1 \times 10^{-15}$ Supplementary Figure S1C), perhaps indicating a set of genes specifically selected in this geographical region compared to the Andean material and/or closer kinship between those populations and therefore less statistical independence. We found 22 glycerolipid genes that were consistently PBE outliers in all four populations ($p = < 1 \times 10^{-10}$ Supplementary Figure S1C). We then performed an independent analysis for each of the 30 pathways and compared the average pathway (10 Kb window around genes of that pathway) PBE value with a genome wide genic random sampling distribution of PBE values. We found that 'phospholipid remodelling' and 'PC acyl editing' pathways had significantly high PBE values selected across all four populations indicating a possible adaptive role of phospholipid remodelling in maize highland adaptation (Supplementary Figure S1D and Supplementary File 1).

ZmLPCAT1 is an example of a gene selected, with shared outlier SNPs in the CDS region, in all four populations that is part of the 'phospholipid remodelling' and 'PC acyl editing' pathways (Figure 1E), Supplementary File 1). *ZmPLA1.2* performs the reverse reaction of *ZmLPCAT1* and is part of the 'PC acyl editing', 'triacylglycerol degradation' and 'phospholipases' pathways. *ZmPLA1.2* is a good example of an outlier gene in the Southwestern US and Mesoamerican populations. *ZmPLA1.2* has particularly high PBE values in the Mexican highland population and contains SNPs that are unique for each population and others that are shared across populations (Figure 1E). All taken together our data shows that phospholipid pathways, show clear signs of recurrent selection across several highland populations both at the metabolic (PC/LPC ratios) and gene level like *ZmPLA1.2* and *ZmLPCAT1*.

A major QTL explaining PC to LPC conversion overlaps with *ZmPLA1.2*

To break population structure and identify loci involved in phospholipid synthesis in highland maize, we developed a Backcross Inbred Line BC1S5 population, between B73 (a temperate inbred line) and Palomero Toluqueño (a Mexican highland landrace) using B73 as the recurrent parent (75% B73, 25% PT). Palomero Toluqueño (PT) accession Mexi5 (CIMMYTMA 2233) is a popcorn (Palomero means popcorn in Spanish) from the Toluca valley in México (Figure 3A). The Hilo landrace panel and the B73 x PT BC1S5 mapping population were grown on the same highland and lowland common gardens and samples for glycerolipid analysis were collected.

In highland conditions, with typical 5 growth degree units across the growth season, Palomero Toluqueño showed higher fitness than B73 (Figure 3A-B). While B73 typically flowers around 65 days after planting in US temperate conditions and Mexican lowland conditions, B73 flowers around 150 days after planting in our Toluca field (Figure 3A) Using the sum of LPC

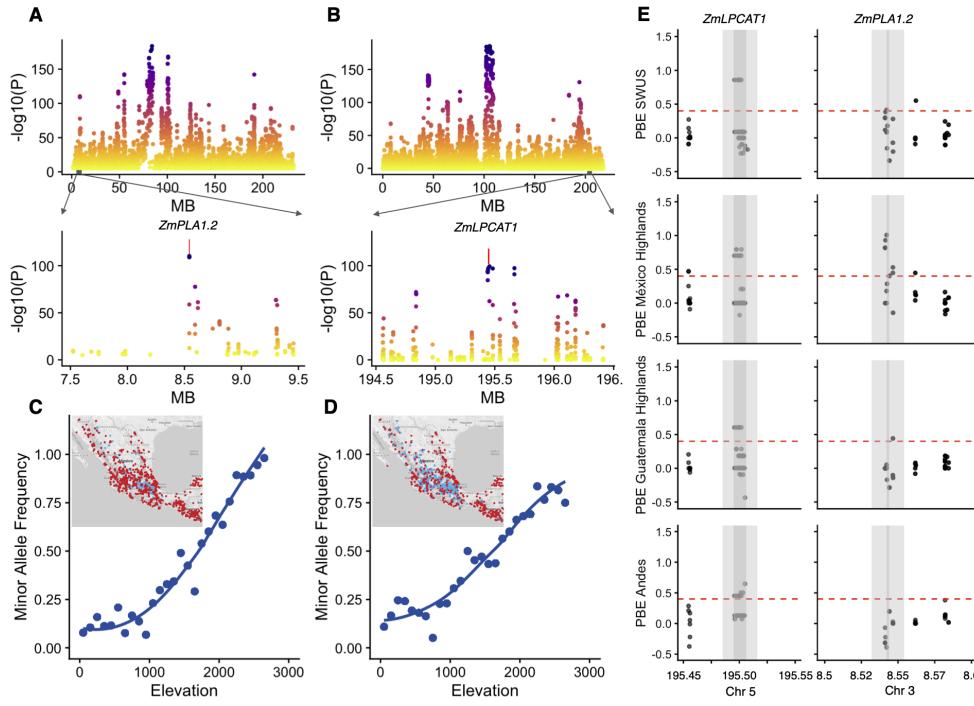


Figure 2 Highland selection in genes determining PC/LPC ratios. Manhattan plots of minus $\log_{10}(P)$ -values *pcadapt* outliers. **A**) and **B**) *pcadapt* PC1 outliers plots of chromosome 3 and 5, respectively. Lower panels are zoomed areas of outlier SNPs that co-localize with the physical position of the coding sequences (marked with a red line) of *ZmPLA1.2* and *ZmLPCAT1*. **C**) and **D**) show the geographic and elevation dependent minor allele frequencies of the highland (blue) and lowland (red) alleles of one of the outlier SNPs in the coding sequence of *ZmPLA1.2* and *ZmLPCAT1*. **E**) PBE values of SNPs in *ZmLPCAT1* and *ZmPLA1.2*

species, we found a major QTL peak ($LOD = 9.2$, 53% of phenotypic variance explained) located at 8.5 Mb of chromosome 3 (AGPv3) *qLPCs3@8.5* (Figure 3B). We also found a major QTL peak *qPCs3@8.5* in the same locus as *qLPCs3@8.5* when we use the sum of PC species (PCs) (Figure 3B), ($LOD = 5.6$, 37% of phenotypic variance explained). The PCs/LPCs ratio also showed a major QTL *qPCs3/LPCs3@8.5* on the same locus as *qLPCs3@8.5* and *qPCs3@8.5* with an even larger LOD , ($LOD = 24.5$, 87% of phenotypic variance explained). We searched for epistatic effects in LPCs, PCs, and PCs/LPCs ratios through a combination of R/qtl *scantwo* and *stepwise* functions (46) but no additional significant QTLs were found. *qLPCs3@8.5*, *qPCs3@8.5* and *qPCs3/LPCs3@8.5* were robust to environmental effects and were found in BILs grown in highland and lowland environments. The additive effect of the PT allele at these QTLs lead to high levels of PCs, low levels of LPCs and consequently high PC/LPC ratios while we observed the opposite effect for the B73 allele (Figure 3C, top panel). Individual PC and LPC QTLs at this locus show the same additive PT allele effect behaviour than the summary *qLPCs3@8.5* and *PCs3@8.5* (Figure 3C, top panel and Supplementary S2). All individual LPC QTLs at the *qLPCs3* locus correspond to LPCs that contain at least one double bond in the fatty acid (Figure 3D, Supplementary Fig 2, Supplementary file 3). The summary *qPCs3@8.5* was driven mainly by PC species with more than 2 fatty acid double bonds such as PC 36:5 (Figure 3C and Supplementary S2 bottom panel)

We then sought to identify the potential candidate gene underlying the QTLs at Chr 3. The QTL 7.9–10 Mb 1.5 LOD drop confidence interval contained 72 genes. We hypothesized that the metabolic phenotypes we observed could be due to a gene that is involved in the process of PC-LPC conversion. There are 75 genes

in the maize genome with predicted phospholipase activity (Supplementary Figure S3A) and half of them have predicted PLA1 activity (Supplementary Figure S3A). We identified *ZmPLA1.2* (Chr3:8,542,107..8,544,078), right at the QTL peak, as the most likely candidate (Figure 3B). *ZmPLA1.2* has a predicted Phospholipase A1-Igamma1 activity and can be classified, based on its two closest *Arabidopsis* orthologs (At1g06800 and At2g30550), as a PC hydrolyzing PLA1 Class I Phospholipase (47). PLA1 phospholipases hydrolyze phospholipids in the sn-1 position and produce a lyso-phospholipid and a free fatty acid as a result (Supplementary Figure S3B).

If *ZmPLA1.2* is the underlying causal gene of the QTL, the metabolic phenotypes observed would be consistent with a loss or impaired function of the *ZmPLA1.2-PT* allele that leads to higher levels of PCs and low levels of LPCs in PT. We generated a CRISPR-CAS9 *ZmPLA1.2* (*ZmPLA1.2^{CR}*) knockout mutant in B104, a temperate inbred derived from B73, and measured PC and LPC species in WT and mutant plants grown under greenhouse control conditions. *ZmPLA1.2^{CR}* phenocopied (Figure 3C-D bottom panels) the PT allele effect of the BILs, further confirming that the *ZmPLA1.2-PT* is a loss of function allele that underlies the QTL in chr3 @8.5 Mb.

ZmPLA1.2 loss of function could be due to a mis-regulation of *ZmPLA1.2* expression in highland landraces and/or to a mutation affecting the enzymatic activity of *ZmPLA1.2*. We analyzed *ZmPLA1.2* expression in B73, PT and the corresponding F1 in plants grown under high and low temperatures simulating highland and lowland conditions (Figure 3E). Under cold conditions *ZmPLA1.2-B73* was up-regulated but *ZmPLA1.2-PT* was not (Figure 3E). F1 plants showed a similar expression pattern to B73 plants. *ZmPLA1.2* on the F1 showed a pattern of expression con-

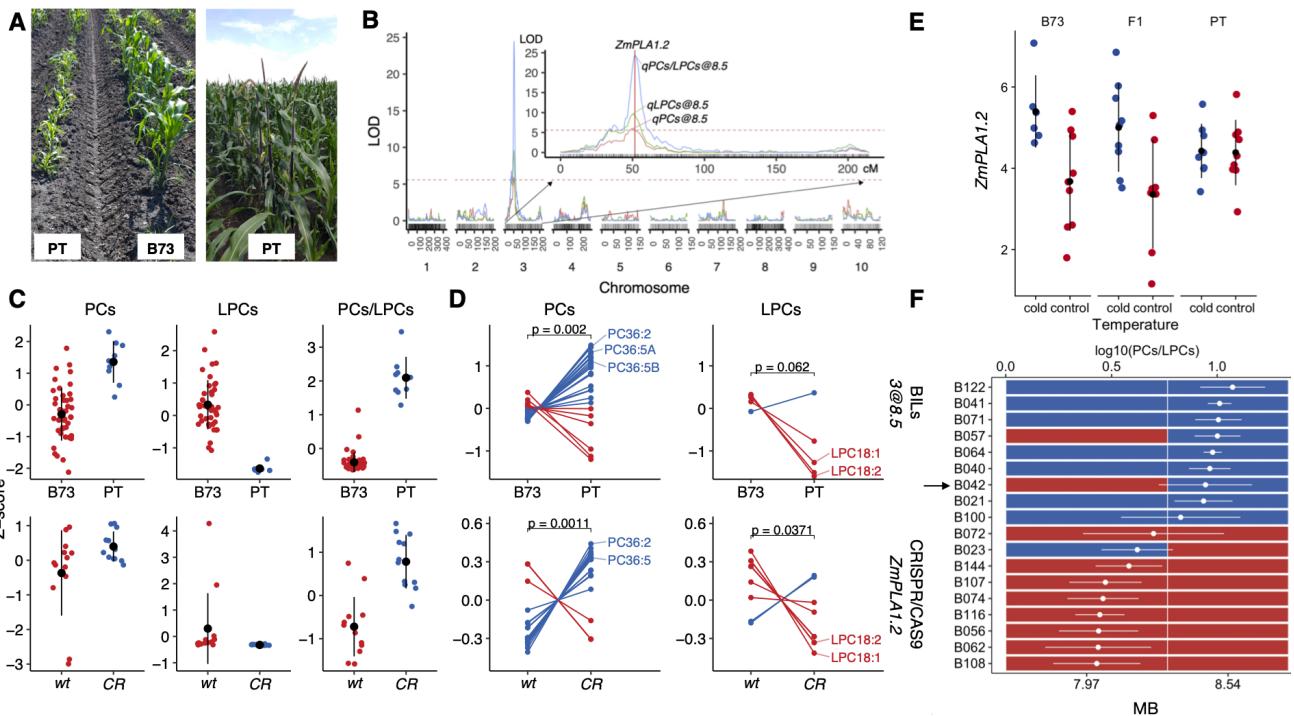


Figure 3 *ZmPLA1.2* is a major QTL explaining PC/LPC conversion. **A)** PT and B73 plants growing in the highland Metepec field. **B)** QTL analysis using data collected from plants growing in the highland and lowland fields of PCs, LPCs and PCs/LPCs ratio identified overlapping major QTLs at 8.5 Mb in chromosome 3. QTL peak coincides with the physical location of *ZmPLA1.2*. **C)** PCs, LPCs and PC/LPCs z-scores effect sizes of BILs at chr 3 8.5 Mb that are either homozygous B73 or PT (top row) and CRISPR-CAS9 *ZmPLA1.2*^{CR} mutant and wild plants (bottom row). **D)** Individual PCs and LPCs species z-scores effect sizes of BILs at at chr 3 8.5 Mb (top row) and CRISPR-CAS9 *ZmPLA1.2*^{CR} mutant. **E)** PC/LPC ratios of several BILs including B042 that shows a recombination of event 500 bp upstream of the ATG. **F)** *ZmPLA1.2* Expression analysis of B73, PT and their F1 grown in control and cold temperatures in growth chamber conditions.

sistent with a dominant B73 effect and this was also the case when we analyzed PC/LPCs ratios in the few B73 x PT BC1S5 BILs that are heterozygous at the *qPC/LPC3@8.5* locus (Supplementary Figure S5B). Loss of function could also be the result of enzymatic malfunction of the *ZmPLA1.2*-PT allele and, in fact, PBE and *pcadapt* outlier SNPs are located within the CDS of *ZmPLA1.2* and not within the regulatory region. We Sanger sequenced B73 x PT BILs (B021, B042, B122) that are homozygous PT on the *ZmPLA1.2* locus. We identified a recombination point 500 base pairs upstream of the ATG of *ZmPLA1.2* (Figure 3F, Supplementary Figure S4) in BIL B042. BIL B042 then contains a *ZmPLA1.2*-PT CDS but its promoter region, 500 base pairs upstream of the ATG, is mainly *ZmPLA1.2*-B73. PC/LPC levels on the B042 BIL were similar to other BILs that are homozygous PT at the 8.54 Mb marker in the QTL peak (Figure 3F). This result supports the hypothesis that the metabolic effect we see is due to a misfunction or a loss of function of the *ZmPLA1.2*-PT enzyme rather than changes in the *ZmPLA1.2*-PT regulatory region.

We analyzed nucleotide diversity in the promoter and CDS of *ZmPLA1.2* of highland and lowland landraces from México and South America and we did not observe any obvious pattern when we compared highland vs lowland landrace (Supplementary Figure S5A).

Fitness effects of *ZmPLA1.2*-PT

***ZmPLA1.2* shows strong elevation-dependent antagonistic pleiotropy in Mexican landraces.** We re-analyzed phenotypic data from the F1 Association Mapping panel (16) and (17) and fit a model to estimate the effect of *ZmPLA1.2*-PT allele on sev-

eral fitness trait's intercept and slope on trial elevation using *GridLMM* (48). We found that genetic variation in *ZmPLA1.2* showed significant effects of genotype by environment interactions on several fitness related traits. (Figure 4A). *ZmPLA1.2* showed clear antagonistic pleiotropy effects on flowering time traits (Figure 4A). The highland *ZmPLA1.2*-PT allele was associated to an increase of around one day of male flowering time (DTA) and 1/4 of day in the Anthesis to Silking Interval (ASI) in low elevation environments while, at high elevations, the highland allele was associated with a decrease of DTA and ASI of one and 1/4 of a day, respectively. Yield related traits such as fresh ear weight and grain weight per hectare showed typical conditional neutrality effects where the highland allele had no effects in lowland environments but led to higher yield values in highland environments.

***ZmPLA1.2*^{CR} mutants phenocopy the effect of the highland allele in flowering time.** We then grew the *ZmPLA1.2* CRISPR-CAS9 mutant in long day conditions in North Carolina and measured flowering time. The *ZmPLA1.2*^{CR} mutant phenocopied the effect of the highland allele in Mexican lowland conditions and lead to an increase of flowering time of around 1 day (Figure 4B). We are currently performing further experiments in conditions simulating highland environments to test if the *ZmPLA1.2*^{CR} mutant shows a similar reduction in flowering time as what we observed with in Figure 4A, confirming an interaction between *ZmPLA1.2* and environment.

Association of *ZmPLA1.2*^{CR} with phosphorus levels. We measured phosphorus content in flag leaves of highland lan-

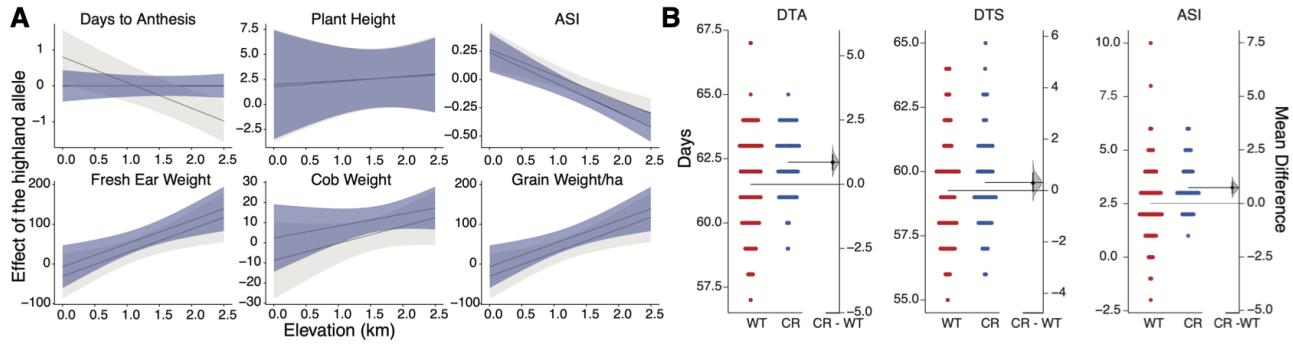


Figure 4 Fitness effects of *ZmPLA1.2-PT* and *ZmPLA1.2^{CR}*. **A)** We used BLUPs and GBS data from 2700 landraces from (17) evaluated in 23 common gardens at different elevations in México. We modeled each trait as a function of *ZmPLA1.2-PT* genotype, trial elevation, and tester line, with controls for main effects and responses to elevation of the genomic background. Gray lines and ribbons show estimates of the effect of the highland allele of *ZmPLA1.2-PT* as a function of common garden elevation \pm 2SE, using the *GridLMM* package (48). Purple lines show estimates of the *ZmPLA1.2-PT* effect in a model that additionally included effects of Days-to-Anthesis. **B)** Flowering traits measured in the CRISPR-CAS9 *ZmPLA1.2^{CR}* mutant is long day conditions during the Summer of 2020 in Clayton, NC

draces from México, Perú and B73 grown in control field conditions. Phosphorus levels in highland landraces were higher than in B73 (Supplementary Fig S6A). In México, in geographical locations of accessions of the Seeds dataset, the probability of finding Andosol soils, a type of volcanic soil that has low pH values and low phosphorus availability increased with elevation (Supplementary Fig S6B). It is possible that higher levels of phosphorus in highland landraces might be associated with adaptation to soils with low phosphorus availability. To test if *ZmPLA1.2* is associated with phosphorus leaf levels we measured phosphorus content in leaves *ZmPLA1.2* CRISPR-CAS9 mutants but we did not observe any differences when compared to wild type plants (Supplementary Fig S6C). Principal component analysis analysis of ion levels in mutant and WT plants revealed no differences between the two genotypes (Supplementary Fig S6D).

Identification of causal SNP in *ZmPLA1.2^{CR}*. Using Sanger sequencing of PT BILs at the *ZmPLA1.2* locus we identified several non-synonymous SNPs within the CDS (407, 520, 553, 610, 631, 1028, 1315, 1342, and 1345 from the ATG) that could have an effect on *ZmPLA1.2*. We focused our attention on SNP 631 on the flap lid domain that leads to a conservative substitution from isoleucine to valine (V211I, Figure 5A). The flap lid domain is located in a lipase 3 domain that is highly conserved across the tree of life. We identified 982 observations of the PF01764 lipase 3 PFAM domain in 719 prokaryote species using PfamScan (49; 50), and then calculated bacterial optimal growth temperatures from their tRNA sequences (51). We then tested if genetic variation in PF01764 residues was significantly associated with bacterial optimal growth temperatures. We found that all of the significant associations were located in the flap lid region (Figure 5A, bold letters). We then specifically analyzed variation in the 211 residue and observed that the PT allele (V) was associated with lower bacterial optimal growth temperatures than the B73 allele (I) (Figure 5B) suggesting that the SNP we identified in PT leading to V211I may be associated with adaptation to low temperatures that highland maize is usually exposed to.

Class I Phospholipases are targeted to the chloroplast and in fact we identified a Chloroplast Transit Peptide at the beginning of the CDS of the gene (Figure 5A) using ChloroP (52). We further confirmed chloroplast localization by transiently expressing

the *ZmPLA1.2* Chloroplast Transit Peptide fused with GFP in *Nicotiana benthamiana* leaves (Supplementary Figure S7).

ZmPLA1.2-PT is an introgression of teosinte mexicana and is conserved in Flint inbreds

We then explored if the SNP leading to the V211I residue change was unique to PT or was conserved in other highland maize. The PT allele was conserved in highland landraces from México and Guatemala and was segregating in Southwestern US landraces. The B73 allele was fixed in lowland Mexican, South American and Andean landraces (Figure 6A). These results are consistent with the PBE results we have observed before (Figure 2E). This is a typical pattern of *teosinte mexicana* introgression (24) and indeed the PT allele was present in both *teosinte mexicana* accessions but only in 1/4 of the *teosinte parviflora* accessions available in Hapmap 3 (53) (Figure 6A). This lead us to ask whether the PT allele was the result of *teosinte mexicana* introgression in highland maize or selected from *parviflora* standing variation.

To test for *mexicana* introgression, we used f_d data from (12). f_d data around the *ZmPLA1.2* indicated that the region was introgressed from *mexicana* into highland maize. (Figure 6B). We then performed a haplotype network analysis using SNP data of the *ZmPLA1.2* CDS from 1160 Mexican homozygous accessions from the SeeD Dataset (16) and the *teosinte* TIL lines from Hapmap 3 (53). We identified nine haplotype groups that clustered mainly based on elevation. (Figure 6C) The two major groups (II) and (VI) contained mainly lowland and highland landraces respectively. The two *mexicana* TIL lines (TIL08 and 25) were located in group IV (Figure 6C) together with highland landraces primarily collected in the Trans-Mexican Volcanic Belt (30/36 from highlands of Jalisco, Michoacán, México, Puebla and Veracruz). We then checked whether this *mexicana* *ZmPLA1.2* haplotype introgressed in mesoamerican highland maize is also present in modern maize inbreds. To address this, we constructed a phylogenetic tree using Hapmap 3 inbred lines including those from the 282 inbred panel, Teosinte Inbred Lines, German Lines and PT. We identified two main groups, one containing the *ZmPLA1.2-PT* haplotype and the other one containing the *ZmPLA1.2-B73* haplotype. Palomero Toluqueño and the *teosinte mexicana*'s TIL-08 and TIL-25 clustered together with Northern European (Figure 6D) Flints like EP1, UH008, and UH009. Other northern US flints

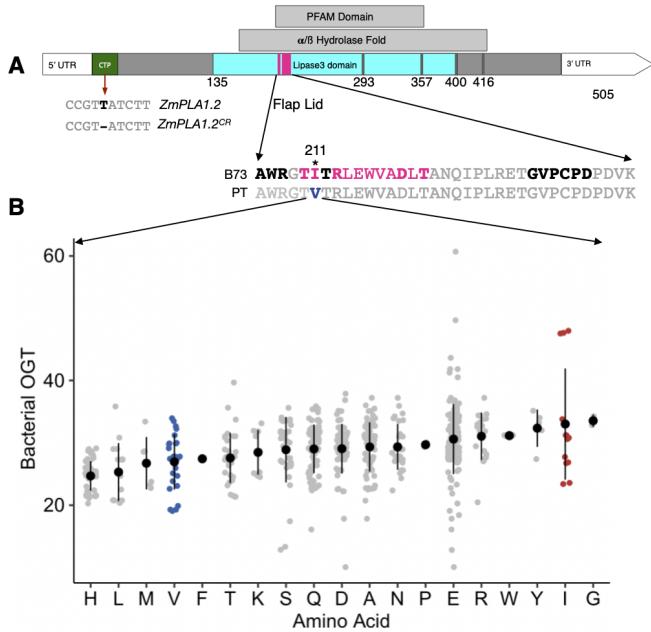


Figure 5 Analysis of *ZmPLA1.2* flap lid domain SNP effects. A) *ZmPLA1.2* promoter and coding sequence showing different features. CTP represents the chloroplast transit peptide. The domains of *ZmPLA1.2* were analyzed from UniProt identifier A0A1D6MIA3. The Lipase3-PFAM - PF01764, and alpha/beta Hydrolase fold were identified using InterPro, and Lipase3-CDD, shown in cyan, including flap lid, shown in magenta, and S293, D357, and H400 catalytic triad were identified from CDD. H416 was identified as a substitute for H400 in the catalytic triad by protein modelling. B) Bacterial optimal growth temperature (OGT) of different residues at the position in the flap-lid domain where the B73/PT mutation leading to a V211I conversion. Across the whole lipase domain, residues in bold in the flap lid domain (A) showed high correlation with bacterial OGT.

like CM7 are also closely related to the mexicana *ZxPLA1.2* haplotype. These data suggest that after introgression into highland maize, the *ZxPLA1.2* haplotype was conserved in Flint materials adapted to cold environments in the North of the US, Canada and Europe. Building on previous reports that indicate a role of highly unsaturated PC species in determining flowering time (40; 41) and considering the significant *ZmPLA1.2*-PT induced accumulation of those PC species, we asked if genetic variation of *ZmPLA1.2* could be associated with flowering time traits in modern maize. We used a large gene expression dataset obtained from the 282 maize diversity panel that was sampled at several developmental stages (54), and phenotypic datasets collected from the same panel grown in long and short day conditions. We found that *ZmPLA1.2* and *ZmLPCAT1* expression are usually inversely correlated in most of the tissues (Supplementary Figure S8), further supporting the idea that these two enzymes are co-regulated. We also found significant associations of *ZmPLA1.2* expression in aerial tissues similar with several flowering time traits. The magnitude of this associations is similar with other well known genes that are involved in determining flowering time (Supplementary Figure S8) like *ZmZCN8* and *ZmRAP2.7*. Furthermore, in both long and short days conditions, the inbred lines that carry the *ZmPLA1.2*-PT allele showed lower levels of expression and shorter flowering times than the inbreds that

carry the *ZmPLA1.2*-B73 allele (Figure 6E).

Discussion

Understanding the genetic, molecular and physiological basis of crop species adaptation to different environments and the role that wild relatives have on these processes is relevant to identify favorable genetic variation that can be used to improve modern crops. The repeated events of maize adaptation to highland environments constitute an excellent natural experiment to study crop local adaptation processes. Recent studies (24; 23; 15) have helped improved our understanding of the genetics of maize highland adaptation, however, the molecular, physiological and genetic mechanisms of maize highland adaptation remain largely unknown. Phospholipids are key structural components of plant membranes that also function as signalling molecules in adaptation to stresses prevalent in highland environments (47; 55) such as low phosphorus availability (56; 57; 38) and low temperature (35; 36; 58). Additionally, in Arabidopsis, accumulation of certain highly unsaturated phosphatidylcholine (PC) species can accelerate flowering time (40), a major driver of maize adaptation to highland environments (16; 17).

Here we show that genes involved in the synthesis and degradation of PC species, such as *ZmPLA1.2* and *ZmLPCAT1*, have been repeatedly selected in maize highland populations leading to high PC/LPC ratios. We then identified *ZmPLA1.2* as the major candidate gene explaining a major effect PC/LPC QTL in a B73 x Palomero Toluqueño BIL population. The allelic effects of this QTL indicate that the highland PT allele is a loss of function, probably due to single point mutation in the flap-lid domain of *ZmPLA1.2* that leads to high PC/LPC ratios by impairing PC to LPC conversion. Genetic variation at *ZmPLA1.2* has a strong GxE effect dependent on elevation that leads to a positive effect of the highland allele in highland environments probably by shortening flowering time. This effect was further confirmed with a CRISPR-CAS9 missense mutant, *ZmPLA1.2*^{CR}. We then found that the highland *ZmPLA1.2* haplotype is the result of teosinte *mexicana* introgression that is still conserved in Northern USA, Canada and EU Flints.

We found that genes of pathways involved in the synthesis and degradation of phospholipids were repeatedly selected in several highland maize populations of North America, Central America and South America (Figure 1 and 2). *ZmLPCAT1* and *ZmPLA1.2* were two of the genes that showed strongest, repeated signals of selection measured by PBE and *pcadapt* in highland populations (Figure 2). A previous study found that *ZmLPCAT1* showed high *Fst* values when comparing highland and lowland landraces (23). The predicted function of *ZmLPCAT1* is to synthesize PC species via the acylation of LPC species while the predicted function of *ZmPLA1.2* is to hydrolyze PCs into LPCs and free fatty acids by cleaving at the sn1-position. Selection in these two genes is probably driving the high PC/LPC ratio that we found in highland Mexican landraces (Figure 1). Natural variation in maize *ZmPLA1.2* is associated with lipid content (59) and flowering time (60; 27). In B73, *ZmPLA1.2* is one of the most highly expressed phospholipases and its expression pattern is almost entirely restricted to vegetative leaves (V4-V9) (Supplementary Figure S3C) (61), the same type of leaves that we sampled for glycerolipid analysis. Additionally, *ZmPLA1.2* is highly expressed in B73 and other temperate inbreds under low temperature conditions and is down-regulated in heat conditions (Supplementary Figure S3D) (62). In our experiments, B73 and the B73 x PT F1 were up-regulated in cold conditions, while the

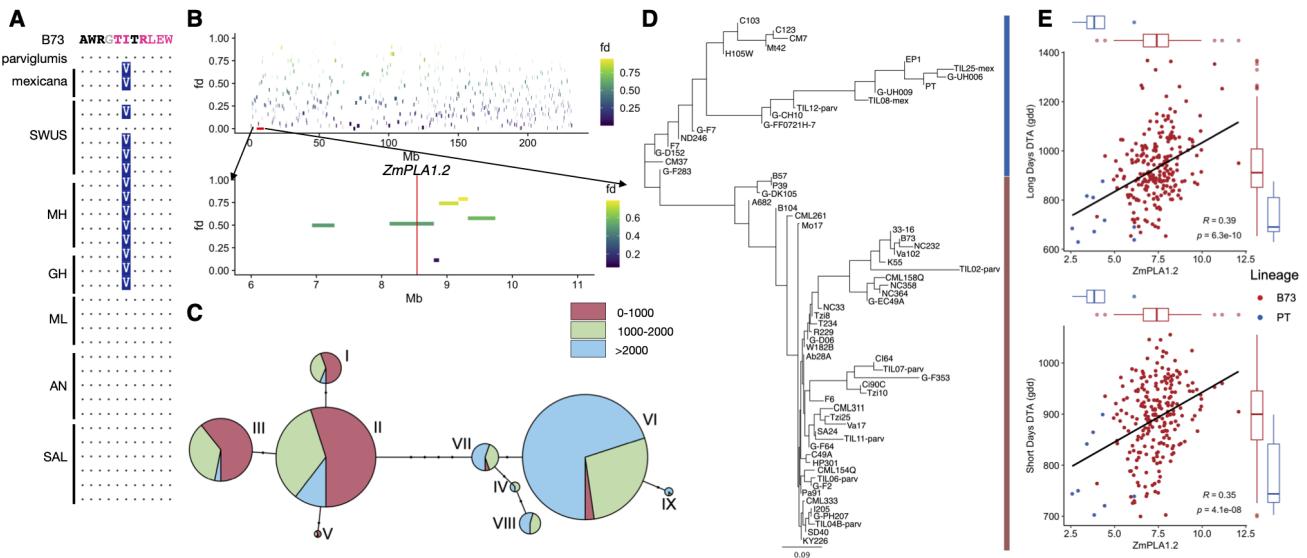


Figure 6 Introgession of teosinte *mexicana* into maize *ZmPLA1.2*. **A)** Alignments around the V211I mutation in the flap-lid domain in B73, *mexicana* and *parviflumis* and highland and lowland landraces Southwestern US, Mesoamerican highlands and lowlands and Andes and South America lowlands. **B)** f_d analysis of *mexicana* introgression. Data was obtained from (12). **C)** Haplotype network analysis of *ZmPLA1.2* CDS SNPs using 1060 Mexican homozygous individuals from the Seeds dataset. **D)** Cluster analysis of *ZmPLA1.2* CDS using a sample of Hapmap3 inbreds and Palomero Toluqueño. **E)** *ZmPLA1.2-PT* expression correlation with DTA in short and long days. Inbred lines from the PT lineage shown in panel C are colored in blue while inbred lines from the B73 lineage are colored in red. Data from (54)

expression of PT showed similar levels than B73 in control conditions, it was not up-regulated in cold conditions (Figure 3E). Our data suggest that Mexican highland environments have low phosphorus availability and that highland landraces from México and Perú contain significantly higher levels of phosphorus (Supplementary Figure S6A-B), however data from the *ZmPLA1.2^{CR}* mutant does not support a role of *ZmPLA1.2* in maize adaptation to low phosphorus. In Arabidopsis, *LPCAT1* is involved in the determination of phosphorus leaf content levels under low Zn conditions (63). In maize a GWAS hit 45 KB upstream of *ZmLPCAT1* for flowering time under low phosphorus conditions further supports the possible role of natural variation of *ZmLPCAT1* in plant adaptation to low phosphorus availability. We are currently exploring if *ZmLPCAT1* elevation dependent genetic variation could be involved in adaptation could be involved in highland maize adaptation to low phosphorus availability.

Our QTL analysis of PC/LPC ratios in a B73 × PT mapping population and in the *ZmPLA1.2^{CR}* mutant support the hypothesis that the highland *ZmPLA1.2-PT* allele results in a loss of function enzyme that rewires highland Mexican maize PC metabolism leading to high PC/LPC ratios (Figures 3). Adaptive loss of function mutations can be an effective way to gain new metabolic functions in new environmental conditions (64). Our data supports an enzymatic loss of function due to a single conservative amino acid substitution located in the flap lid domain of *ZmPLA1.2* that can impact substrate accessibility and/or substrate binding (Figure 5A). Indeed, flap lid domains are targets of biotechnological modification of these types of enzymes (65).

Why were the metabolic changes induced by *ZmPLA1.2* selected in highland maize? PC metabolism is intimately connected to multiple stress response and developmental pathways; alterations in PC amounts and PC/LPC ratios impact overall plant fitness. The large *qPC/LPC3@8.5* QTL is driven by individual QTLs of PC and LPC species with high levels of unsaturated

fatty acids (Supplementary Figure S2). Some of these species, like PC 36:5 and LPC 18:1 (Figure 3D, Supplementary File 3), show similar patterns during Arabidopsis cold acclimation (36) and sorghum low temperature response (58). PC 36:5 also showed very high Q_{ST} values when comparing highland and lowland landraces from both Mesoamerica and South America (Figure 1C-D, Supplementary File 2). In maize *ZmPLA1.2* expression is circadianly regulated (66) and peaks at the end of the day. In Arabidopsis, highly unsaturated PC (34:3, 34:4, 36:5, 36:6) species increase during the dark (67), coinciding with low expression levels of *PLA1.2* (66). Furthermore, Yuki Nakamura and colleagues elegantly showed that PC 36:5 and 36:6 species accumulate during the night and can bind to Arabidopsis flowering time locus T (FT) accelerating flowering time (40) by cellular mechanisms still not well understood. The Arabidopsis FT ortholog in maize, *ZCN8* (68) underlies a major flowering time and photoperiod sensitivity QTL (27). Additive mutations in the regulatory region of *ZCN8*, including a teosinte *mexicana* introgression, lead to higher expression of *ZCN8* contributing to maize adaptation to long days in temperate conditions (69). Our results show that *ZmPLA1.2* drives the accumulation of PC species (like PC 36:5) that accumulate in Arabidopsis at the end of the day and that can bind to FT. We hypothesize that accumulation of these PC species in highland maize could drive early flowering in a similar way that occurs in Arabidopsis.

The flap-lid mutation in *ZmPLA1.2-PT* residue 211 is conserved in highland Mexican and Guatemalan landraces and in teosinte *mexicana* (Figure 6A) but is still segregating in teosinte *parviflumis*. We showed that indeed, *ZmPLA1.2-PT* is an introgression from teosinte *mexicana* (Figure 6B) and that this introgression has been conserved in modern inbred lines particularly northern US, Canada and EU Flints (Figure 6D). *ZmPLA1.2* in inbreds carrying the *ZxPLA1.2 mexicana* haplotype show low levels of expression and earlier flowering times (Figure 6E) (54). More-

over, genetic variation in the regulatory region of *ZmPLA1.2* was significantly associated with photoperiod sensitivity in the maize Nested Association Mapping population (27). Other grasses *Trip-sacum dactyloides* adapted to temperate latitudes also show accelerated rates of evolution in genes involved in PC metabolism (70). The effect of genetic variation of *ZmPLA1.2* on flowering time in Mexican landraces indicated a strong GxE interaction where the highland allele lead to a reduction of flowering time and ASI in highland environments (Figure 4A) similar to the effect observed to the well known teosinte *mexicana* introgression of inversion *inv4m* (15).

Other flowering time loci analyzed do not show this clear GxE effect (17). Interestingly, the effect of the highland allele shows typical conditional neutrality in yield-related traits, with increased fitness of the *ZmPLA1.2-PT* allele in highlands (Figure 4A) Conversely, the previously characterized *mexicana* introgression *textitinv4m* (12; 71; 11) shows no effect in highlands and negative effect in lowlands (?).

To the best of our knowledge this is the first documented case of an adaptive teosinte *mexicana* introgression with a characterized metabolic phenotype and fitness advantage in highlands.

Materials and Methods

Populations used in the analysis. Highland and lowland populations used for Population Branch Excess analysis consisted of three to six accessions from each of the highland and lowland populations and have been previously described in (24; 72). The 120 Landraces from the HiLo Diversity Panel were selected and ordered from the CIMMYT germplasm bank to maximize a good latitudinal gradient sampling across Mesoamerica and South America. For each highland (>2000 masl) landrace a lowland (<1000 masl) sample was selected at the same latitude (<0.5°) to form 60 highland/lowland pairs, 30 from each continent. The list of the accessions used is provided in Supplementary file 4. B73 x Palomero Toluqueño Backcross Inbred Lines (BILs) were developed by crossing B73 with a single Palomero Toluqueño plant (Mexi5 accession, CIMMYTMA 2233) that was then backcrossed with B73 once and selfed five times (BC1S5). We used individual landrace accession genotype and fitness data from the CIMMYT Seeds of Discovery project (SeeD) (17) to calculate *padapt* (43) values and GxE effects of *ZmPLA1.2-PT*.

Field Experimental Conditions and sampling Two replicates of the HiLo Diversity panel accessions and three replicates of the B73 x PT BILs were planted in a field located in Metepec, Edo de México, (19°13'28.7"N 99°32'51.6"W) within the Trans-Mexican volcanic belt. The field is at 2610 meters above sea level (masl), and the range of average monthly temperatures along the year vary from 5 °C to 21.5 °C with an average annual of 13.6 °C. We collected 50 mg of fresh tissue (10 discs) using a leaf puncher from the tip of the second youngest leaf above the last leaf with a fully developed collar around V4-V6 developmental stage. Tissue discs were immediately flash frozen in liquid nitrogen. We collected all samples from a field in a single day between 10:00 am and 12:00 pm, approximately 3 h after sunrise. Samples were transported in dry ice to the lab and stored at -80°C until extraction.

Glycerolipid analysis We crushed frozen samples in a tissue grinder Retsch (Haan, Germany) for 40 seconds at a frequency of 30 1/s. We performed lipid extraction following Matyash and collaborators (73). First, we added 225 µL of cold methanol (MeOH) to each sample. For the blanks, MeOH previously prepared with a Quality Control (QC) mix was added (Supplementary File 5).

We vortexed each sample for 10 seconds, keeping the rest of the material on ice. Then, we added 750 µL of cold methyl tert-butyl ether (MTBE). The MTBE added to the blanks contained 22:1 cholesterol ester as internal standard (Supplementary File 5). We vortexed each sample for 10 seconds, followed by 6 minutes of shaking at 4°C in the orbital mixer. We next added 188 µL of LC/MS grade water at room temperature (RT), and vortexed samples for 20 seconds. We centrifuged the samples for 2 min at 14000 rcf and recovered 700 µL of supernatant from the upper organic phase. We then split the supernatant into two aliquots of 350 µL, one for lipid profiling and the other for preparation of pools to be used along the lipid profiling. Finally, samples were dried with a speed vacuum concentration system. We resuspended dried samples in 110 µL of MeOH-Toluene 90:10 (with the internal standard CUDA, 50 ng/mL). We vortexed samples at low speed for 20 s and then sonicated at RT for 5 min. We then transferred aliquots of 50 µL per sample into an insert within an amber glass vial. The UHPLC-QTOF MS/MS utilized were Agilent 1290 and Agilent 6530, respectively. Before analyzing the samples, a new Waters Acquity charged surface hybrid (CSH) C18 2.1x100 mm 1.7 µm column was set. The column was initially purged for 5 min. We coupled the UHPLC column with a Van-Guard pre-column (Waters Acquity CSH C18 1.7 µm). We injected six "no sample injections" at the beginning of each run to condition the column, followed by ten samples, one pool (made out of the mix of the second aliquot of all the samples contained per UHPLC plate) and one blank. We injected 1.67 µL per sample into UHPLC-QTOF MS/MS ESI (+); the running time per sample was 15 min. Mobile phase "A" consisted of 60:40 acetonitrile:water, 10 mM of ammonium formate and 0.1% formic acid. Mobile phase "B" consisted of 90:10 isopropanol:acetonitrile, 10 mM ammonium formate and 0.1% of formic acid. The flow rate was 0.6 mL/min and the column compartment was maintained at 65° C. Initial conditions were 15% B; the gradient uniformly increased until reaching 100%. At 12.10 min the mobile phase composition returned to initial conditions. The mass spectrometer (Q-TOF MS/MS) was operated in positive electrospray ionization mode (ESI) For the source parameters, ESI gas temperature was set at 325 °C, nebulizer pressure at 35 psig, gas flow at 11L/min, capillary voltage at 3500 V, nozzle voltage at 1000V, and MS TOF fragmentor and skimmer at 120 and 65 V, respectively. Under the acquisition parameters a mass range between 60 and 1700 m/z was set. As for reference mass parameters, a detection window of 100 ppm and a minimum height of 1000 counts were set. We performed a retention time (rt) correction of the acquired data using Agilent MassHunter Qualitative Analysis B.06.00 version and Microsoft Excel. To extract ion chromatograms (EICs) of the internal standards within the run we used Agilent MassHunter Qualitative Analysis. We identified the time of the highest intensity point of each EIC, which then was used as the current retention time of the experiment. We used the method retention time for internal standards and the current rt and we fitted a polynomial regression to calculate new retention times using retention times from 501 lipids of a MS1 m/z-rt library (See Supplementary File 6). In MSDIAL (74), identification of lipids is based on two approaches: the MSP file and MS/MS identification setting included in MSDIAL and the use of a post identification file containing accurate m/z and rt for a list of lipids. In this study we used both identification approaches. Under positive ion mode, the MSP file and MS/MS identification setting has a total of 51 lipid classes that can be selected for identification. The post identification file that we used was the retention time-

corrected MS1-MS2 mz-rt lipid library that we explained before. We used MSDIAL (74) version 3.40. To use MSDIAL, the raw data was converted from .d to .abf format with Reifycs Abf converter (<https://www.reifycs.com/AbfConverter/>). The MSDIAL alignment results were filtered out based on whether compounds intensity was ten times above blank intensity. Then, filtered data was normalized using Systematic Error Removal using Random Forest (SERRF) (75). This normalization is based on the quality-control pool samples. Normalized features were filtered out considering a coefficient of variation (CV) equal or less than 30% among the pools. To curate the data for duplicate features, isotopes and ion-adducts, we utilized MS-FLO (76). Curated data was also normalized using the sum of all known metabolite signal (mTIC). After data processing and normalization, lipid intensities were used for further analysis.

Q_{ST} - F_{ST} analysis of glycerolipid data Quantitative trait differentiation (Q_{ST}) was contrasted to the distribution of F_{ST} for neutral genetic markers (77). Highland/Lowland contrasts were considered separately for Mesoamerica and South America.

A linear mixed effects model (R package lmer, function lmer) was used to partition phenotypic variance between population pairs (Mesoamerica/South America, all highland/all lowland, Mesoamerican highland/Mesoamerican lowland, South American highland/South American lowland).

$$TRAIT \sim 1 + (1|POPULATION) + (1|GARDEN/BLOCK) + (1|BATCH)$$

Within-population and between-population variances were calculated with the R function VarCorr (R package lme4, 78), and were used to calculate Q_{ST} following the equation below:

$$Q_{ST} = \sigma_{GB}^2 / (\sigma_{GB}^2 + 2\sigma_{GW}^2)$$

in which σ_{GB}^2 and σ_{GW}^2 are the between- and within-population genetic variance components, respectively (42). Pairwise F_{ST} was calculated with the R function fst.each.snp.hudson (R package dartsR, 79).

pcadapt analysis of biological adaptation in Mexican landraces In order to conduct genome scans for signatures of adaptation we used the pcadapt (43) package. pcadapt identifies adaptive loci by measuring how strongly loci are contributing to patterns of differentiation between major axes of genetic variation. Under simple models pcadapt captures major patterns of F_{ST} but is conducted in a way that does not require population delimitation (80). As the genome scan comparison requires a focal SNP to be compared to the first K principal components of the genotype data, it can be biased by large regions of low recombination that drive the major axes of variation in the principal components. Thus, when SNPs from these low recombination regions are compared against principal components driven by linked loci spurious signals may arise. To prevent this bias from occurring, we used custom scripts to calculate the principal component step separately based upon all the chromosomes except for the chromosome of the focal SNPs being tested. The genotype data we used for this analysis was GBS data from roughly 2,000 landraces of Mexican origin collected by CIMMYT (www.cimmyt.org) as part of the SeeDs of discovery initiative (<https://www.cimmyt.org/projects/seeds-of-discovery-seed/>). From this, we calculated the strength of association between each SNP and the first five principal components (excluding the chromosome of the focal SNP) using the communality statistic as implemented in pcadapt version 3.0.4.

Glycerolipid pathways selection We compiled a list of genes pertinent to glycerolipid metabolism starting with a search

of all genes belonging to the *Zea mays* 'Glycerophospholipid metabolism' and 'Glycerolipid metabolism' KEGG pathways (81) (map identifiers: zma00564 and zma00561). With the NCBI Entrez gene identifiers in KEGG we retrieved the AGPv4 transcript identifiers used in Corncyc 8.0 (82; 83) from an id cross reference file found in MaizeGDB () (82). This resulted in a list of 300 genes comprising 51 Corncyc pathways. Then we discarded Corncyc pathways tangentially connected to the KEGG glycerolipid metabolism list (sharing just one enzyme with the initial KEGG list) or that we judged to belong to different biological processes (e.g 'long chain fatty acid synthesis', 'anthocyanin biosynthesis'). Finally, we added manually the 'phosphatidylcholine biosynthesis V' pathway that was missing. The list of 30 selected Corncyc pathways included genes outside the initial KEGG search results and raised the number of genes to 557. In addition to this, 37 genes were found to have an enzymatic activity related to phospholipid metabolism but not placed into any particular pathway, i.e orphan enzymes, consisting mostly of alcohol dehydrogenases. Sixteen additional genes found in KEGG were not annotated at all in Corncyc probably due to differences between AGPv4 and RefSeq pseudogene annotation of the maize genome. The list of all possible candidates coming either from KEGG or Corncyc that were orphan enzymes or were unannotated in Corncyc amounted to 594 genes (Supplementary File 1). This process is documented in the 0_get_glycerolipid_genes.R script of the pgplipid R package accompanying this paper (84).

Population Branch Excess Analysis Population Branch Excess quantifies changes in allele frequencies in focal populations relative to two independent "outgroup" populations. We used *Zea mays spp. parviflora* as one of the outgroup populations for all four highland groups. The other outgroup was Mexican lowlands in the case of Southwestern US, Mexican highlands and Guatemalan highlands; and South American lowlands in the case of the Andes population. We used calculated PBE SNP values for the 4 populations (described in detail in (24)) and we tested for selection outliers SNPs in the 594 phosphoglycerolipid candidates and the 30 Corncyc pathways (556 genes). We first defined PBE outlier SNPs as the top 5% of the PBE score distribution; this fraction corresponds to approximately 50000 out of 1 million genotyped SNPs in each population. Following (author?) (24), we defined a gene as a PBE outlier if it contained an outlier SNP within the coding sequence or 10 Kbp upstream/downstream. Then we tested for over-representation of genes selected in particular subsets of populations using Fisher's exact test with the 32283 protein genes from the maize genome (Supplemental Figure 1a) (85) as background. For each pathway, we first selected all SNPs within CDS regions and 10Kb upstream and downstream of genes in the pathway and we calculated the mean pathway PBE score. We then constructed a null distribution by drawing 10000 samples without replacement of n SNPs from those found within or around 10Kb upstream and downstream of all protein coding genes and we obtained the mean PBE for this null distribution. With the set of PBE outliers for glycerolipid metabolism in the 4 populations we tested for evidence of physiological or pleiotropic constraint using the C_χ^2 statistic (45).

QTL analysis of phospholipid levels We analyzed glycerophospholipid QTLs in a mapping population of 57 BILs (BC1S5) from the cross B73 x (PT). These BILs were grown in a highland site in Metepec, Edo de México at 2600 masl during the Summer of 2016 and in Puerto Vallarta, Jalisco at 50 masl during the Winter of 2016/17. We analyzed the samples using UHPLC-QTOF, as above, and 67 leaf lipid species were identi-

fied. For QTL analysis we calculated the mean across all fields of individual lipid mass signal. We also used as phenotypes the sum total of the following lipid classes: diacylglycerol, triacylglycerol, PC and LPC. Furthermore, we also included the log base 10 transformed ratios of LPCs/PCs and the ratios of their individual species. We did a simple single marker analysis with “scanone” using Haley-Knott regression, and assessed the QTL significance with 1000 permutations.

CRISPR-CAS9 editing of *ZmPLA1.2* and analysis of the effect of *Zmpla1.2^{CR}* mutant on flowering time CRISPR/Cas9 was used to create a *ZmPLA1.2* gene knockout through *Agrobacterium* mediated transformation of background line B104 (86; 87). Guide RNA was designed as described in (88) for the B73 reference genome v4. B104 and B73 sequence for *ZmPLA1.2* were identical. The gRNA cassette was cloned into pGW-Cas9 using Gateway cloning. Two plants from the T0 transgenic event were identified through genomic PCR amplification and Sanger sequencing and were self-pollinated. Plants were genotyped using forward primer CAGTTCTCATCCATGCACG and reverse primer CCTGATGAGAGCTGAGGTCC. Several T1 plants containing the *ZmPLA1.2^{CR}* event were planted for lipid analysis in greenhouses at North Carolina State University during Spring 2020 and then self-pollinated. Cas9 positivity was tested for using 0.05% Glufosinate ammonium contained in Liberty herbicide. T2 seeds from CAS9 free T1 plants were collected and used for flowering time analysis in Clayton, NC during the 2020 Summer.

Expression analysis of *ZmPLA1.2* in B73, PT and B73xPT F1s in conditions simulating highland environments Gene expression data was generated from leaf tissue from B73, PT and the B73xPT F1. Plants were grown following the same protocol as in (15). Briefly, kernels were planted in growth chambers set to imitate spring temperature conditions in the Mexican lowlands (22°C night, 32°C day, 12 hr light) and highlands (11°C night, 22°C day, 12 hr light). Leaf tissue was sampled from the V3 leaf the day after the leaf collar became visible, between 2 and 4 hours after lights came on. Tissue was immediately placed in a centrifuge tube, frozen using liquid nitrogen, and stored at -70°C.

RNAseq libraries were constructed, sequenced, and analyzed following (15). Briefly, randomly primed, strand specific, mRNA-seq libraries were constructed using the BRaD-seq (89) protocol. Multiplexed libraries were sequenced on 1 lane of an Illumina HiSeq X. Low quality reads and adapter sequences were removed using Trimmomatic v.0.36 (90), and the remaining paired reads were aligned and quantified using kallisto v.0.42.3 (91). Gene counts were normalized using the weighted trimmed mean of M-values (TMM) with the calcNormFactors function in edgeR (92) and converted to log2CPM.

Sanger sequencing of *ZmPLA1.2* in BILs homozygous at the *qPC/LPC3@8.5* locus We identified 3 BILs homozygous for the PT allele at the *ZmPLA1.2* locus and we developed 6 sets of primers to Sanger sequence across the CDS and the gene promoter. The location of primers in the gene are shown in Supplementary file 7.

Nucleotide Diversity of *ZmPLA1.2* We estimated nucleotide diversity for the promoter and coding regions of *ZmPLA1.2* using WGS data from highland and lowland accessions from México and South America obtained from (72) and the R package PopGenome (93). FASTA files were partitioned into 4 populations by origin supplemental for accessions included) and then subset into coding and promoter regions, which was defined as 3Kb upstream of the cds. Data were imported into PopGenome

using the option include.unknown=FALSE in order to prevent bias by excluding missing and ambiguous nucleotide data. Nucleotide diversity was measured separately within each population and averaged by the number of sites in the coding and promoter regions.

Association of *ZmPLA1.2* with agronomic traits We re-analyzed phenotypic data from the F1 Association Mapping (FOAM) panel of Romero-Navarro *et al* (16) and Gates *et al* (17) to more fully characterize association signatures of *ZmPLA1.2*. Full descriptions of this experiment and data access are described in those references. We downloaded BLUPs for each trait and line from Germinate 3, and subset the data to only those lines with GBS genotype data from México. We fit a similar model to the GWAS model used by (17) to estimate the effect of the *ZmPLA1.2-PT* allele on the trait’s intercept and slope on trial elevation, accounting for effects of tester ID in each field and genetic background and family effects on the trait intercept and slope using four independent random effects. We implemented this model in the R package GridLMM (48). We extracted effect sizes and covariances conditional on the REML variance component estimates and used these to calculate standard errors for the total *ZmPLA1.2-PT* effect as a function of elevation. To test whether the phenotypic effects of *ZmPLA1.2-PT* on yield components could be explained as indirect effects via flowering time, we additionally re-fit each model using Days-To-Anthesis as a covariate with an independent effect in each trial.

Effect of *Zmpla1.2^{CR}* mutant on flowering time Seeds of T2 CAS9-free plants were grown in isolation during summer 2020 in Clayton, NC. Female (Days to Silking) and male (Days to Anthesis) flowering time were calculated from the day of planting until the first silks and anther pollen shed could be observed in each individual plant. Plants homozygous *Zmpla1.2^{CR}* and WT derived from the same T1 families were planted for this experiment. Effect size analysis was performed using the dabestR package (94).

Phosphorus analysis and phosphorus soil availability data We analyzed phosphorus content in B73 and 5 Mexican and 5 Peruvian highland landraces (Supplementary file 9) grown in control field conditions in the 2018 Puerto Vallarta, México, Winter nursery. Samples were analyzed using ICP-MS according to (95). Frequency of Andosol soils at different elevations was calculated using the soilP package (96). Phosphorus content in flag leaves of the the same *Zmpla1.2^{CR}* and WT mutants used for the flowering time experiment were analyzed using ICP-OES at the North Carolina Department of Agriculture.

Bacterial optimal growth temperature association with *ZmPLA1.2* flap-lid domain allelic variation The maize *ZmPLA1.2* protein sequence was compared to prokaryotes with the same sequence to determine whether the identified residue change in maize and accompanying association with low temperature survival was consistent with observations in other organisms. Pfam domain PF01764 was identified in the B73 protein sequence using the HMMER3 web server, and 982 observations of the PF01764 Pfam domain were identified in 719 prokaryote species using PfamScan (49; 50). The optimal growth temperature of these species was predicted using tRNA sequences as in (51). Maize and prokaryote PF01764 domain sequences were aligned with hmmlalign from the hmmer3 package (97), and the aligned Pfam sequences were recoded to reflect nine amino acid physicochemical properties (98). Sequences were filtered to remove gaps in the domain alignment and observations with only partial domain sequences, then clustered based on sequence similarity,

974 resulting in two clusters of observations within the domain. For
975 each cluster, positions in the filtered alignment were associated
976 with prokaryote optimal growth temperatures using a linear
977 regression with all 9 amino acid physicochemical properties.
978 Seventeen sites in and around the flap-lid region of the protein
979 passed a 10% FDR significance threshold, including the single
980 residue change p.Ile211Val previously identified in *ZmPLA1.2-PT*.
981 Welch's two-sided t-test was used to compare the optimal growth
982 temperatures of prokaryote species with the *ZmPLA1.2-B73* allele
983 to the optimal growth temperatures of prokaryote species with
984 the *ZmPLA1.2-B73* allele at this site.

985 **Subcellular localization of *ZmPLA1.2*** We fused the
986 *ZmPLA1-B73* Chloroplast Transit Peptide (52 aminocacids) with
987 GFP. Three constructs encoding subcellular localization signals
988 were used as control; Cytoplasm (C-GFP), nucleus (N-GFP), and
989 Chloroplast (P-GFP). All of them were under control of the 35S
990 promoter. These constructs were transiently expressed in *Nicotiana benthamiana* leaf cells.

991 **Teosinte *mexicana* introgression in highland maize** To evaluate if the *ZmPLA1.2-PT* allele was the result of standing variation
992 from teosinte *parviflumis* or introgression from *mexicana* we used Patterson's *D* statistic and genome-wide f_d to calculate
993 ABBA-BABA patterns. The data was obtained from (12). The
994 material used in (12) included whole genome sequence data from
995 three highland outbred individuals: two Palomero Toluqueño
996 and one Mushito de Michoacán; three lowland landraces: Nal
997 Tel (RIMMA0703) and Zapalote Chico (RIMMA0733) obtained
998 from (72) and BKN022 from (53); two *mexicana* inbreds: TIL08
999 and TIL25; three *parviflumis* inbreds: TIL01, TIL05, TIL10 and
1000 *Tripsacum* TDD39103 (53) as an outlier.

1001 **Haplotype network analysis of *ZmPLA1.2* in Mexican
1002 maize landraces and teosintes.** We extracted SNP genotypes
1003 for *ZmPLA1.2* from the TIL teosinte accessions in the HapMap 3
1004 imputed data (53) and the 3700 Mexican landraces in the SEEDS
1005 dataset. With the set of 1060 accessions that were homozygous
1006 at all sites in this genomic region we calculated a haplotype
1007 network depicting the minimal spanning tree for haplotypes
1008 covering 90% of the input accessions with the R package pegas
1009 (99), and haplotype frequencies for three elevation classes in the
1010 landraces (0-1000,1000-2000, >3000 masl).

1011 **Clustering analysis of *ZmPLA1.2* in maize inbreds and
1012 teosintes.** Using v3 of the B73 genome, *ZmPLA1.2* SNPs were ob-
1013 tained from Palomero Toluqueño (this paper), the 282-panel (100),
1014 and teosinte inbred lines the German inbreds from HapMap 3
1015 (53). The selection was made to have a good representation
1016 of tropical, temperate and european inbred lines together with
1017 teosintes and palomero toluqueño lines. SNPs were aligned using
1018 Geneious2020.0.5 and a neighbor-joining cluster analysis was
1019 generated. To facilitate visualization and interpretation of the
1020 tree we condensed cluster branches from lines with identical
1021 haplotypes and from similar geographic locations. The full tree
1022 is available as Supplementary file 8.

1023 **Expression analysis of candidate genes and association
1024 with flowering traits in the 282 panel** We used gene expression
1025 RNA-Seq data obtained from the 282 panel at different develop-
1026 mental stages (54) and BLUP values of several flowering and
1027 photoperiod sensitivity traits (27) to study the correlation of *Zm-*
1028 *PLA1.2* expression values with flowering time traits.

1029 **Acknowledgments**

1030 This work was supported by Conacyt Young Investigator (CB-
1031 238101), Conacyt National Problems (APN-2983), UC-Mexus and

1032 NC State startup funds awarded to Rubén Rellán-Álvarez. Field-
1033 work and mapping population development was supported by
1034 NSF-PGR award 1546719 to Jeffrey Ross-Ibarra, Ruairidh Sawers,
1035 Daniel Runcie and Matthew Hufford. Fausto Rodríguez-Zapata
1036 and Blöcher-Juárez were supported by Conacyt doctoral fellow-
1037 ships. Allison Barnes was supported by NSF-PGRP grant 201703.
1038 Andi Kur was supported as an AgBioFEWS fellow by the NSF
1039 award 1828820. We want to acknowledge the work of the Rellán-
1040 Álvarez, Sawers, Hufford, Flint-Garcia labs and CIMMYT and
1041 Puerto Vallarta Winter Nursery crews that have helped generate
1042 and evaluate the mapping populations developed in this
1043 manuscript. We specially want to thank the indigenous people
1044 of the Americas and the ingenuity through which they domes-
1045 ticated and made possible the spread and adaptation of maize
1046 throughout the continent. Field experiments were carried out on
1047 Wixárika (Puerto Vallarta, México); Nahuatl, Matlatzinca, and
1048 Mazahua (Metepec, México); and Tuscarora (Clayton, NC) land.

1049 **References**

- 1050 [1] C Natarajan, et al., Predictable convergence in hemoglobin
1051 function has unpredictable molecular underpinnings. *Science* **354**, 336–339 (2016).
1052 [2] X Yi, et al., Sequencing of 50 human exomes reveals adap-
1053 tation to high altitude. *Science* **329**, 75–78 (2010).
1054 [3] A Bigham, et al., Identifying signatures of natural selection
1055 in tibetan and andean populations using dense genome
1056 scan data. *PLoS genetics* **6**, e1001116 (2010).
1057 [4] X Liu, et al., EPAS1 gain-of-function mutation contributes
1058 to high-altitude adaptation in tibetan horses. *Molecular
1059 biology and evolution* (2019).
1060 [5] J Yang, et al., Genetic signatures of high-altitude adaptation
1061 in tibetans. *Proceedings of the National Academy of Sciences of
1062 the United States of America* **114**, 4189–4194 (2017).
1063 [6] JP Velotta, CE Robertson, RM Schweizer, GB McClelland,
1064 ZA Chevtron, Adaptive shifts in gene regulation underlie
1065 a developmental delay in thermogenesis in High-Altitude
1066 deer mice. *Molecular biology and evolution* **37**, 2309–2321
1067 (2020).
1068 [7] F Cicconardi, et al., Genomic signature of shifts in selection
1069 in a subalpine ant and its physiological adaptations (2020).
1070 [8] Y Matsuoka, et al., A single domestication for maize shown
1071 by multilocus microsatellite genotyping. *Proceedings of the
1072 National Academy of Sciences of the United States of America*
1073 **99**, 6080–6084 (2002).
1074 [9] DR Piperno, AJ Ranere, I Holst, J Iriarte, R Dickau, Starch
1075 grain and phytolith evidence for early ninth millennium
1076 BP maize from the central balsas river valley, mexico. *Pro-
1077 ceedings of the National Academy of Sciences* **106**, 5019–5024
1078 (2009).
1079 [10] DR Piperno, KV Flannery, The earliest archaeological
1080 maize (*zea mays l.*) from highland mexico: new accelerator
1081 mass spectrometry dates and their implications. *Pro-
1082 ceedings of the National Academy of Sciences of the United States
1083 of America* **98**, 2101–2103 (2001).
1084 [11] MB Hufford, et al., The genomic signature of crop-wild
1085 introgression in maize. *PLoS genetics* **9**, e1003477 (2013).
1086 [12] E Gonzalez-Segovia, et al., Characterization of introgres-
1087 sion from the teosinte *zea mays* ssp. *mexicana* to mexican
1088 highland maize. *PeerJ* **7**, e6815 (2019).
1089 [13] N Lauter, C Gustus, A Westerbergh, J Doebley, The inheri-
1090 tance and evolution of leaf pigmentation and pubescence
1091 in teosinte. *Genetics* **167**, 1949–1959 (2004).
1092

- [14] AK Hardacre, HA Eagles, Comparisons among populations of maize for growth at 13°C. *Crop science* **20**, 780–784 (1980).
- [15] T Crow, et al., Gene regulatory effects of a large chromosomal inversion in highland maize. *PLoS genetics* **16**, e1009213 (2020).
- [16] JA Romero Navarro, et al., A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature genetics* **49**, 476 (2017).
- [17] DJ Gates, D Runcie, GM Janzen, AR Navarro, others, Single-gene resolution of locally adaptive genetic variation in mexican maize. *bioRxiv* (2019).
- [18] JA Aguirre-Liguori, et al., Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (zea mays parviglumis and zea mays mexicana). *Molecular ecology* **24**, 2663 (2019).
- [19] MA Fustier, et al., Signatures of local adaptation in lowland and highland teosintes from whole genome sequencing of pooled samples. *Molecular ecology* (2017).
- [20] P Krasilnikov, et al., *The Soils of Mexico*. (Springer Science), (2013).
- [21] J Stephen Athens, et al., Early prehistoric maize in northern highland ecuador. *Latin American Antiquity* **27**, 3–21 (2016).
- [22] A Grobman, et al., Preceramic maize from paredones and huaca prieta, peru. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1755–1759 (2012).
- [23] S Takuno, et al., Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics*, genetics.115.178327 (2015).
- [24] L Wang, et al., Molecular parallelism underlies convergent highland adaptation of maize landraces. *bioRxiv* (2020).
- [25] RR da Fonseca, et al., The origin and evolution of maize in the southwestern united states. *Nature plants* **1**, 14003 (2015).
- [26] K Swarts, et al., Genomic estimation of complex traits reveals ancient maize adaptation to temperate north america. *Science* **357**, 512–515 (2017).
- [27] HY Hung, et al., ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize (2012).
- [28] Z Liang, et al., Conventional and hyperspectral time-series imaging of maize lines widely used in field trials. *Giga-Science* **7**, 1–11 (2018).
- [29] L Guo, et al., Stepwise cis-regulatory changes in ZCN8 contribute to maize Flowering-Time adaptation. *Current biology: CB* **28**, 3005–3015.e4 (2018).
- [30] ND Coles, MD McMullen, PJ Balint-Kurti, RC Pratt, JB Holland, Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics* **184**, 799–812 (2010).
- [31] C Huang, et al., ZmCCT9 enhances maize adaptation to higher latitudes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E334–E341 (2018).
- [32] Q Yang, et al., CACTA-like transposable element in Zm-CCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16969–16974 (2013).
- [33] S Salvi, et al., Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11376–11381 (2007).
- [34] JT Brandenburg, et al., Independent introductions and admixtures have contributed to adaptation of european maize and its american counterparts. *PLoS genetics* **13**, e1006666 (2017).
- [35] T Degenkolbe, et al., Differential remodeling of the lipidome during cold acclimation in natural accessions of arabidopsis thaliana. *The Plant journal: for cell and molecular biology* **72**, 972–982 (2012).
- [36] R Welti, et al., Profiling membrane lipids in plant stress responses: Role of phospholipase D α in freezing-induced lipid changes in arabidopsis. *The Journal of biological chemistry* **277**, 31994–32002 (2002).
- [37] DV Lynch, PL Steponkus, Plasma membrane lipid alterations associated with cold acclimation of winter rye seedlings (secale cereale l. cv puma). *Plant physiology* **83**, 761–767 (1987).
- [38] H Lambers, et al., Proteaceae from severely phosphorus-impovery soils extensively replace phospholipids with galactolipids and sulfolipids during leaf development to achieve a high photosynthetic phosphorus-use-efficiency. *The New phytologist* **196**, 1098–1108 (2012).
- [39] Y Gu, et al., Biochemical and transcriptional regulation of membrane lipid metabolism in maize leaves under low temperature. *Front. Plant Sci.* **8**, 2053 (2017).
- [40] Y Nakamura, et al., Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nature communications* **5**, 3553 (2014).
- [41] C Riedelsheimer, Y Brotman, M Méret, AE Melchinger, L Willmitzer, The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. *Scientific reports* **3**, 2479 (2013).
- [42] T Leinonen, RJS McCairns, RB O'Hara, J Merilä, Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature reviews. Genetics* **14**, 179–190 (2013).
- [43] K Luu, E Bazin, MGB Blum, pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources* **17**, 67–77 (2017).
- [44] JE Pool, DT Braun, JB Lack, Parallel evolution of cold tolerance within drosophila melanogaster. *Molecular biology and evolution* **34**, 349–360 (2017).
- [45] S Yeaman, AC Gerstein, KA Hodgins, MC Whitlock, Quantifying how constraints limit the diversity of viable routes to adaptation. *PLOS Genetics* **14**, e1007717 (2018) Publisher: Public Library of Science.
- [46] KW Broman, H Wu, S Sen, GA Churchill, R/qtL: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
- [47] SB Ryu, Phospholipid-derived signaling mediated by phospholipase a in plants. *Trends in plant science* **9**, 229–235 (2004).
- [48] DE Runcie, L Crawford, Fast and flexible linear mixed models for genome-wide genetics. *PLOS Genetics* **15**, 1–24 (2019).
- [49] SC Potter, et al., HMMER web server: 2018 update (2018).
- [50] S El-Gebali, et al., The pfam protein families database in 2019. *Nucleic acids research* **47**, D427–D432 (2019).
- [51] E Cimen, SE Jensen, ES Buckler, Building a tRNA thermometer to estimate microbial adaptation to temperature. *Nucleic Acids Research* (2020).

- 1220 [52] O Emanuelsson, H Nielsen, G von Heijne, ChloroP, a neural
1221 network-based method for predicting chloroplast trans-
1222 it peptide and their cleavage sites. *Protein Sci.* **8**, 978–984
1223 (1999).
- 1224 [53] R Bukowski, et al., Construction of the third generation
1225 *zea mays* haplotype map. *GigaScience* (2017).
- 1226 [54] KAG Kremling, et al., Dysregulation of expression cor-
1227 relates with rare-allele burden and fitness loss in maize.
1228 *Nature* **555**, 520–523 (2018).
- 1229 [55] Y Nakamura, Plant phospholipid diversity: Emerging
1230 functions in metabolism and Protein–Lipid interactions.
1231 *Trends in plant science* **22**, 1027–1040 (2017).
- 1232 [56] EJ Veneklaas, et al., Opportunities for improving
1233 phosphorus-use efficiency in crop plants: Tansley review.
1234 *The New phytologist* **195**, 306–320 (2012).
- 1235 [57] A Cruz-Ramirez, et al., The xipotl mutant of *Arabidopsis*
1236 reveals a critical role for phospholipid metabolism in root
1237 system development and epidermal cell integrity. *The Plant*
1238 *cell* **16**, 2020–2034 (2004).
- 1239 [58] SR Marla, et al., Comparative transcriptome and lipidome
1240 analyses reveal molecular chilling responses in Chilling-
1241 Tolerant sorghums. *Plant Genome* **10** (2017).
- 1242 [59] C Riedelsheimer, et al., Genomic and metabolic prediction
1243 of complex heterotic traits in hybrid maize. *Nature genetics*
1244 **44**, 217–220 (2012).
- 1245 [60] C Chen, et al., PICARA, an analytical pipeline providing
1246 probabilistic inference about a priori candidates genes un-
1247 derlying genome-wide association QTL in plants. *PloS one*
1248 **7**, e46596 (2012).
- 1249 [61] SC Stelplug, et al., An Expanded Maize Gene Expression
1250 Atlas based on RNA Sequencing and its Use to Explore
1251 Root Development. *The plant genome* **9**, 0 (2016).
- 1252 [62] AJ Waters, et al., Natural variation for gene expression
1253 responses to abiotic stress in maize. *The Plant Journal* **89**,
1254 706–717 (2017).
- 1255 [63] M Kisko, et al., LPCAT1 controls phosphate homeostasis
1256 in a zinc-dependent manner. *eLife* **7** (2018).
- 1257 [64] AK Hottes, et al., Bacterial adaptation through loss of func-
1258 tion. *PLoS genetics* **9**, e1003617 (2013).
- 1259 [65] FI Khan, et al., The lid domain in lipases: Structural and
1260 functional determinant of enzymatic properties. *Frontiers*
1261 *in bioengineering and biotechnology* **5**, 16 (2017).
- 1262 [66] S Khan, SC Rowe, FG Harmon, Coordination of the maize
1263 transcriptome by a conserved circadian clock. *BMC Plant*
1264 *Biol.* **10**, 126 (2010).
- 1265 [67] S Maatta, et al., Levels of *arabidopsis thaliana* leaf phos-
1266 phatidic acids, phosphatidylserines, and most Trienoate-
1267 Containing polar lipid molecular species increase during
1268 the dark period of the diurnal cycle. *Front. Plant Sci.* **3**, 49
1269 (2012).
- 1270 [68] CM Lazakis, V Coneva, J Colasanti, ZCN8 encodes a poten-
1271 tial orthologue of *arabidopsis* FT florigen that integrates
1272 both endogenous and photoperiod flowering signals in
1273 maize. *J. Exp. Bot.* **62**, 4833–4842 (2011).
- 1274 [69] T Guo, et al., Optimal designs for genomic selection in
1275 hybrid crops. *Molecular plant* **12**, 390–401 (2019).
- 1276 [70] L Yan, et al., Parallels between natural selection in the
1277 cold-adapted crop-wild relative *tripsacum dactyloides* and
1278 artificial selection in temperate adapted maize. *The Plant*
1279 *journal: for cell and molecular biology* (2019).
- 1280 [71] T PyhÄjÄrvi, MB Hufford, S Mezmouk, J Ross-Ibarra,
1281 Complex Patterns of Local Adaptation in Teosinte. *Genome*
1282 *Biology and Evolution* **5**, 1594–1609 (2013).
- [72] L Wang, et al., The interplay of demography and selection
during maize domestication and expansion. *Genome biology*
18, 215 (2017).
- [73] V Matyash, G Liebisch, TV Kurzhalia, A Shevchenko, D
Schwudke, Lipid extraction by methyl-tert-butyl ether for
high-throughput lipidomics. *Journal of lipid research* **49**,
1137–1146 (2008).
- [74] H Tsugawa, et al., MS-DIAL: data-independent MS/MS
deconvolution for comprehensive metabolome analysis.
Nature methods **12**, 523–526 (2015).
- [75] S Fan, et al., Systematic error removal using random forest
for normalizing large-scale untargeted lipidomics data.
Analytical Chemistry **91**, 3590–3596 (2019) PMID: 30758187.
- [76] BC DeFelice, et al., Mass spectral feature list optimizer
(ms-flo): A tool to minimize false positive peak reports
in untargeted liquid chromatography–mass spectroscopy
(lc-ms) data processing. *Analytical Chemistry* **89**, 3250–3255
(2017) PMID: 28225594.
- [77] MC Whitlock, Evolutionary inference from qst. *Molecular*
ecology **17**, 1885–1896 (2008).
- [78] D Bates, M Maechler, B Bolker, S Walker, lme4: Linear
mixed-effects models using eigen and s4. r package version
1.1-7 (2014).
- [79] B Gruber, PJ Unmack, OF Berry, A Georges, dartr: an r
package to facilitate analysis of snp data generated from
reduced representation genome sequencing. *Molecular*
ecology resources **18**, 691–699 (2018).
- [80] N Duforet-Frebbourg, E Bazin, MG Blum, Genome scans for
detecting footprints of local adaptation using a bayesian
factor model. *Molecular biology and evolution* **31**, 2483–2495
(2014).
- [81] M Kanehisa, Y Sato, M Furumichi, K Morishima, M Tan-
abe, New approach for understanding genome variations
in KEGG. *Nucleic Acids Research* **47**, D590–D595 (2019)
Publisher: Oxford Academic.
- [82] JL Portwood, et al., MaizeGDB 2018: the maize multi-
genome genetics and genomics database. *Nucleic Acids*
Research **47**, D1146–D1154 (2019) Publisher: Oxford Aca-
demic.
- [83] JR Walsh, et al., The quality of metabolic pathway re-
sources depends on initial enzymatic function assignments:
a case for maize. *BMC Systems Biology* **10**, 129 (2016).
- [84] FR Zapata, sawers-rellan-labs/pglipid: Glycerophospho-
lipid Pathway analysis (2020).
- [85] M Wang, Y Zhao, B Zhang, Efficient Test and Visualization
of Multi-Set Intersections. *Scientific Reports* **5**, 16923 (2015)
Number: 1 Publisher: Nature Publishing Group.
- [86] Q Wu, et al., The maize heterotrimeric G protein β sub-
unit controls shoot meristem development and immune
responses. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1799–1805
(2020).
- [87] SN Char, et al., An agrobacterium-delivered CRISPR/Cas9
system for high-frequency targeted mutagenesis in maize.
Plant Biotechnol. J. **15**, 257–268 (2017).
- [88] VA Brazelton, Jr, et al., A quick guide to CRISPR sgRNA
design tools. *GM Crops Food* **6**, 266–276 (2015).
- [89] BT Townsley, MF Covington, Y Ichihashi, K Zumstein, NR
Sinha, Brad-seq: Breath adapter directional sequencing: a
streamlined, ultra-simple and fast library preparation pro-
tocol for strand specific mrna library construction. *Frontiers*
in plant science **6** (2015).

- 1344 [90] AM Bolger, M Lohse, B Usadel, Trimmomatic: a flexible
1345 trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–
1346 2120 (2014).
- 1347 [91] NL Bray, H Pimentel, P Melsted, L Pachter, Near-optimal
1348 probabilistic rna-seq quantification. *Nature biotechnology*
1349 **34**, 525 (2016).
- 1350 [92] MD Robinson, DJ McCarthy, GK Smyth, edger: a biocon-
1351 ductor package for differential expression analysis of digi-
1352 tal gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- 1353 [93] B Pfeifer, U Wittelsbürger, SE Ramos-Onsins, MJ Lercher,
1354 PopGenome: an efficient swiss army knife for population
1355 genomic analyses in R. *Molecular biology and evolution* **31**,
1356 1929–1936 (2014).
- 1357 [94] J Ho, T Tumkaya, S Aryal, H Choi, A Claridge-Chang,
1358 Moving beyond P values: data analysis with estimation
1359 graphics. *Nat. Methods* **16**, 565–566 (2019).
- 1360 [95] IR Baxter, et al., Single-kernel ionomic profiles are highly
1361 heritable indicators of genetic and environmental influ-
1362 ences on elemental accumulation in maize grain (*zea mays*).
1363 *PloS one* **9**, e87628 (2014).
- 1364 [96] F Rodríguez-Zapata, R Sawers, R Rellán-Álvarez, SoilP:
1365 A package to extract phosphorus retention data from soil
1366 maps and study local adaptation to phosphorus availabil-
1367 ity (2018).
- 1368 [97] SR Eddy, Accelerated profile HMM searches. *PLoS com-
1369 putational biology* **7**, e1002195 (2011).
- 1370 [98] Z Li, J Tang, F Guo, Identification of 14-3-3 proteins
1371 Phosphopeptide-Binding specificity using an Affinity-
1372 Based computational approach. *PloS one* **11**, e0147467
1373 (2016).
- 1374 [99] E Paradis, pegas: an R package for population genetics
1375 with an integratedâmodular approach. *Bioinformatics* **26**,
1376 419–420 (2010).
- 1377 [100] SA Flint-Garcia, et al., Maize association population: a
1378 high-resolution platform for quantitative trait locus dissec-
1379 tion: High-resolution maize association population. *The
1380 Plant journal: for cell and molecular biology* **44**, 1054–1064
1381 (2005).

Supplementary Tables and Figures

pop1	pop2	C_{hyper}	p
US	MH	3.82	1.11E-04
US	GH	6.17	8.00E-10
MH	GH	4.37	1.24E-05
US	AN	3.51	3.37E-04
MH	AN	2.73	4.43E-03
GH	AN	3.16	1.16E-03

Table 1 Pairwise C_{hyper} statistic for population comparisons.

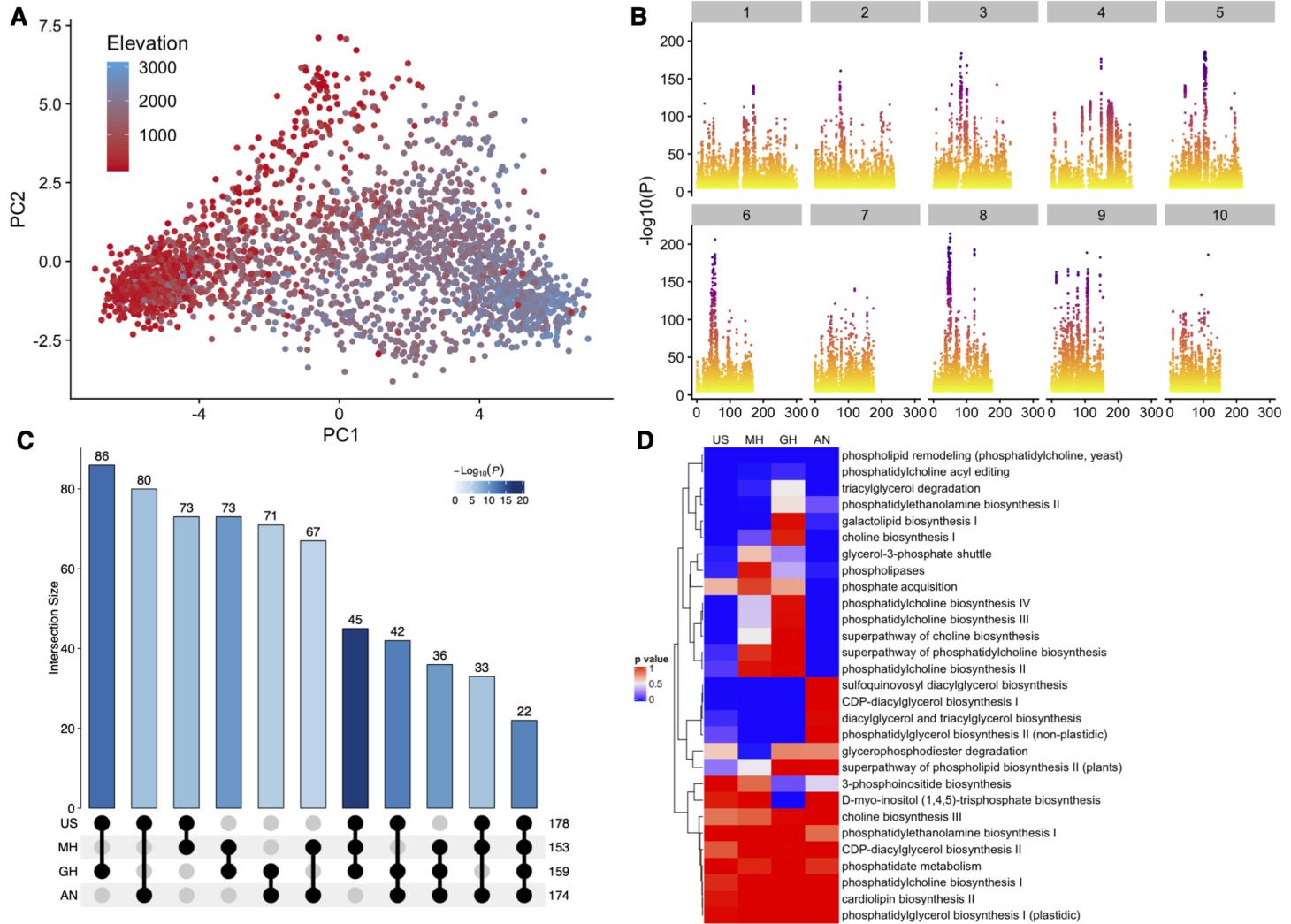


Figure S1 *Pcadapt* analysis of Mexican landraces. We used GBS data from the Mexican landraces of the SEEDs dataset (16) and run a *pcadapt* analysis (43) that identified **A**) elevation as the major driver of population differentiation polarizing PC1. **B**) Genome wide analysis of *Pcadapt* PC1 outliers. **C**) Population Branch Excess Analysis of glycerolipid pathway genes. From the initial set of 211 genes we used 186 genes with 6219 non redundant SNPs and 683162 non redundant SNPs for the SW US group; 186 genes with 6106 non redundant SNPs and 664555 non redundant SNPs for the Mexican Highland group; 185 genes with 5912 non redundant SNPs and 641186 non redundant SNPs for the Guatemalan Highlands group; and 184 genes with 5698 non redundant SNPs and 614783 non redundant SNPs for the Andes group. **D**) Highland selection of Glycerolipid pathways using PBE. See methods for details. For each pathway, we first selected all SNPs in the CDS regions and 10Kb upstream and downstream of the gene and we calculated the mean pathway PBE score. We then constructed a null distribution by drawing 10000 samples without replacement of n SNPs from those found within or around 10Kb upstream and downstream of all protein coding genes and we obtained the mean PBE for this null distribution. For each pathway, the heat map shows p-values corresponding to the probability of sampling from the null distribution a set of n SNPs with the observed glycerolipid pathway mean PBE score or higher.

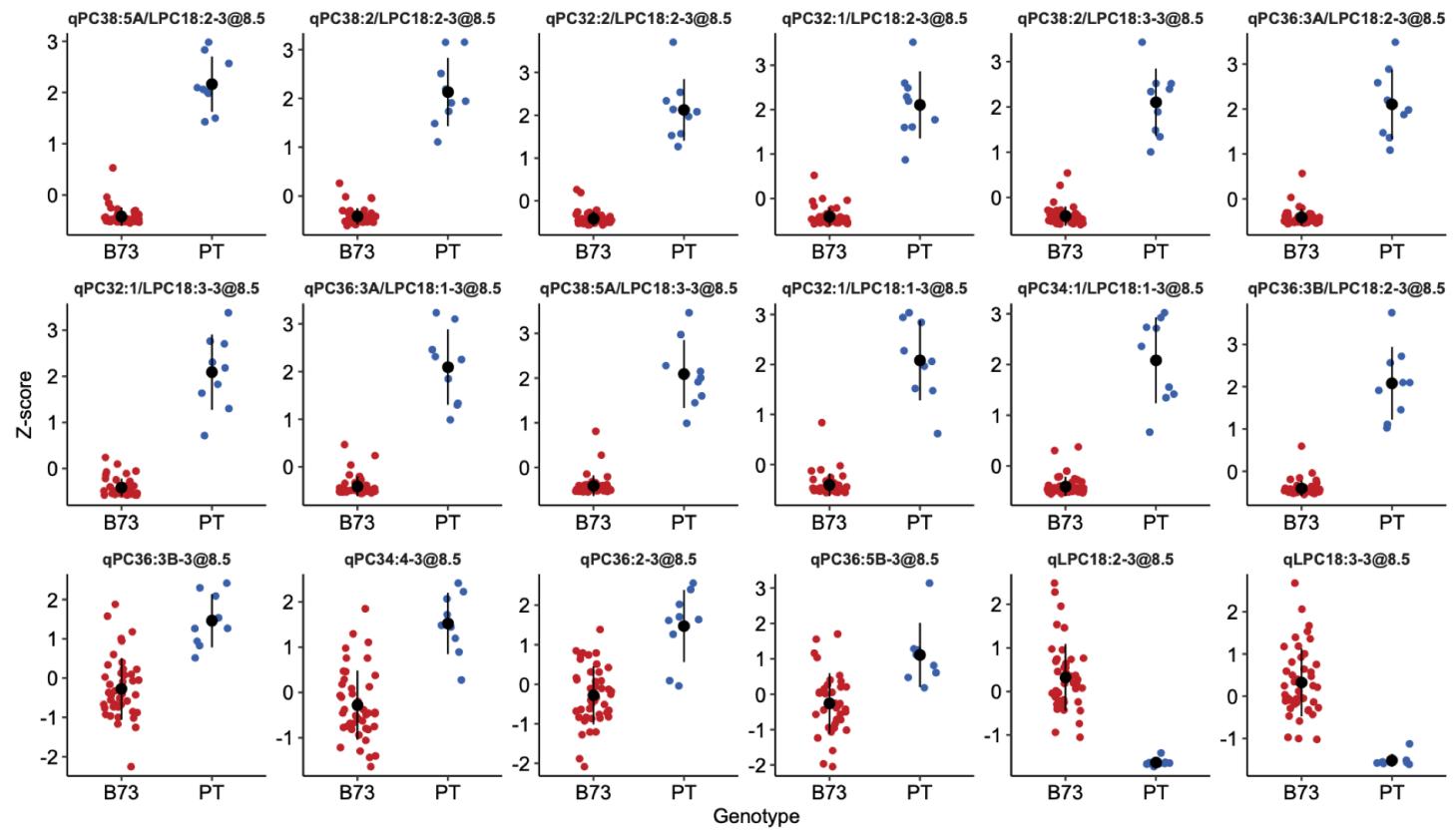


Figure S2 Effect sizes of several individual PC, LPC, and PC/LPC QTL peaks.

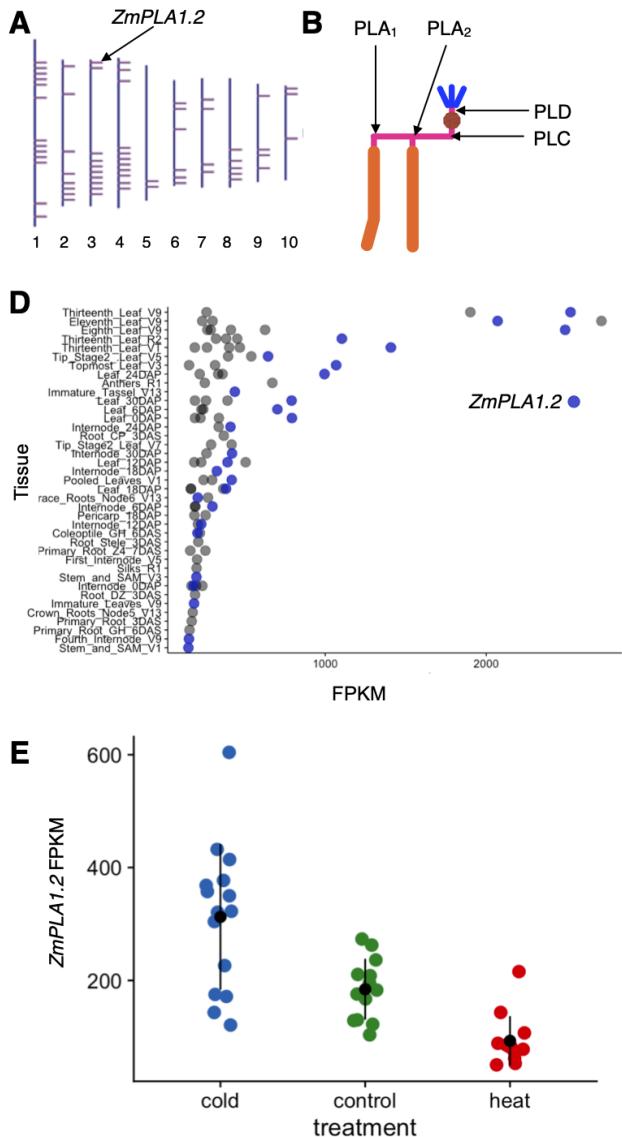


Figure S3 A) Genomic Location of genes coding for enzymes predicted Phospholipase A1 activity. B) Site of action of the different types of phospholipases. C) Subcellular localization of *ZmPLA1.2*. D) B73 expression levels of genes coding for enzymes with predicted Phospholipase A1 activity across different tissues. *ZmPLA1.2* is indicated in blue. E) *ZmPLA1.2* expression levels of temperate inbreds B73, Mo17, Oh43, and Ph207 under control, control and heat stress. Values taken from (62).

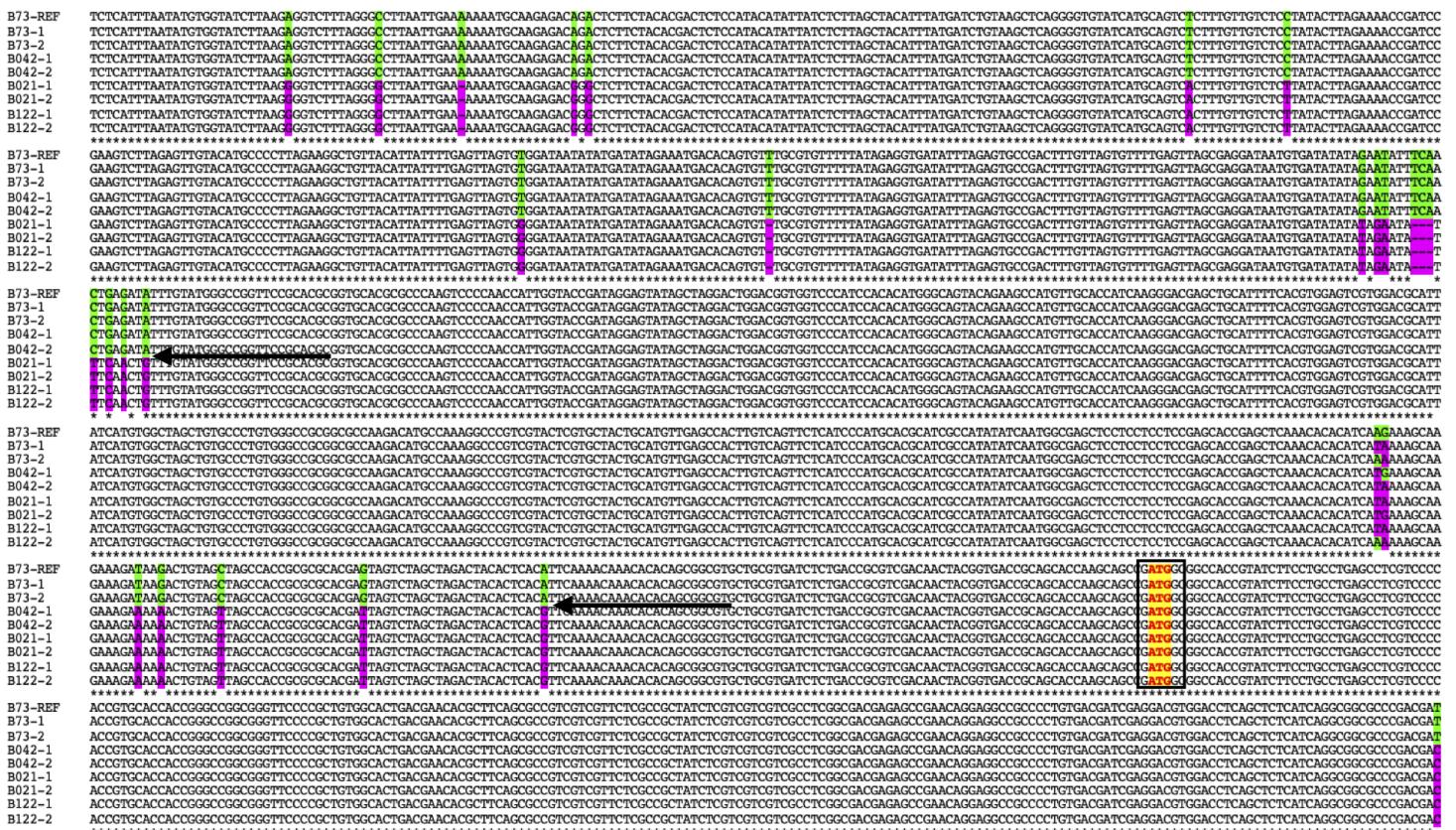


Figure S4 Sanger sequencing of the promoter and start of the *ZmPLA1.2* sequence obtained from B73 plants and 3 BILs (B042, B021 and B122). A recombination point 500 bp upstream the ATG in B104 is indicated by arrows. B73 alleles are marked in green and PT alleles are marked in pink.

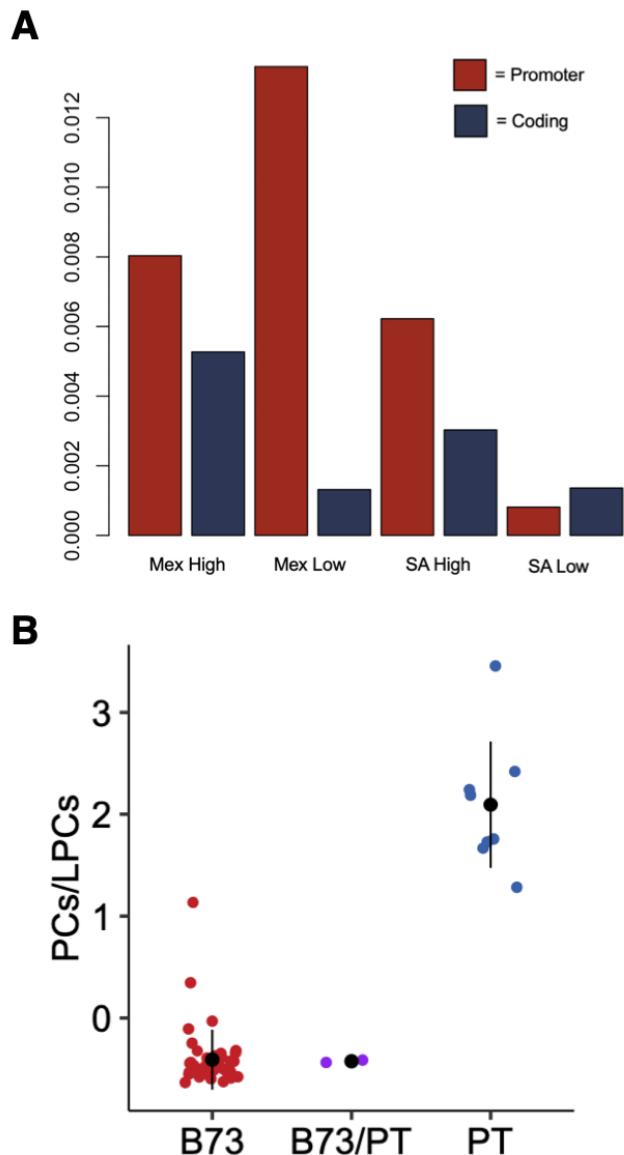


Figure S5 A) Nucleotidy diversity analysis of the promoter and CDS region of *ZmPLA1.2* using whole genome sequencing data of highland and lowland landraces México and South American. **B)** Effect sizes of PC/LPC levels at BILs homozyzogous B73, PT and heterozygous qPC/LPC3@8.5.

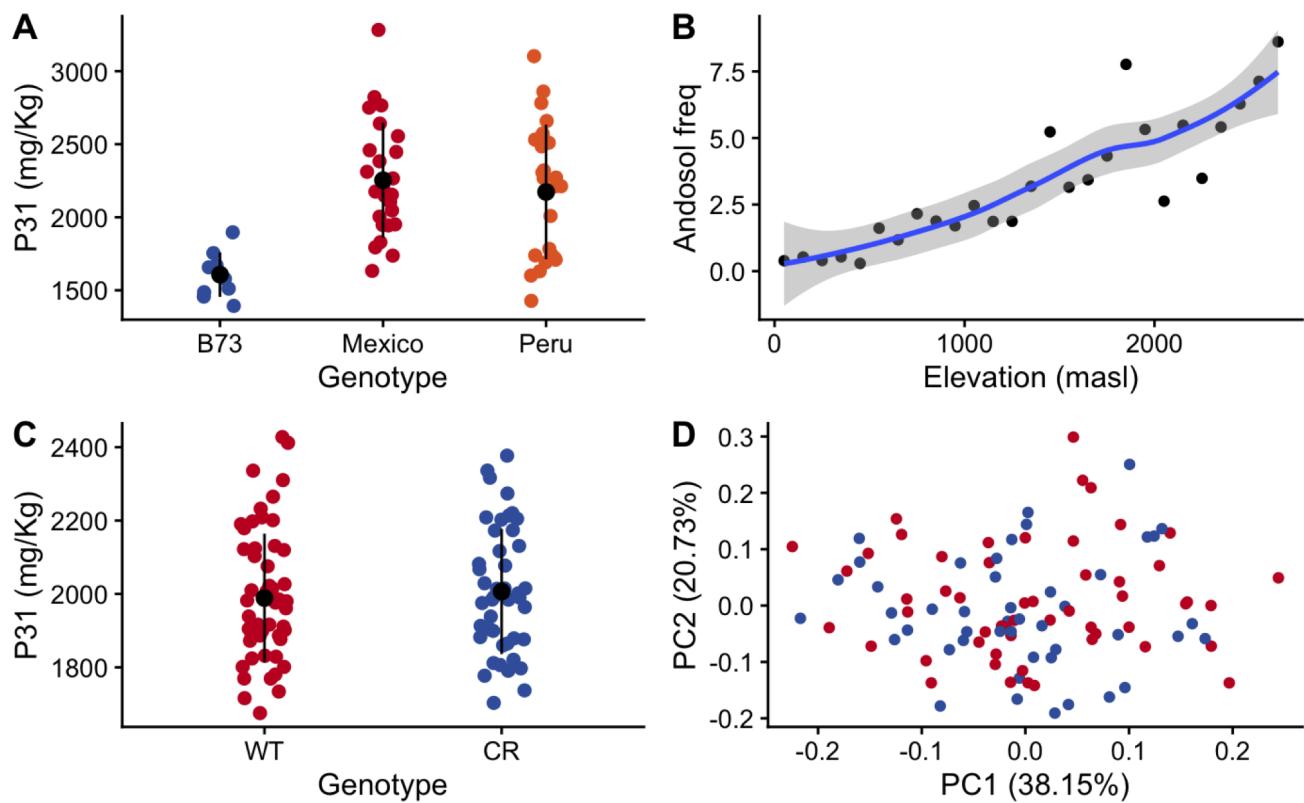


Figure S6 A) Flag leaf phosphorus levels of B73 and 10 highland landraces each from México and Perú grown in control conditions. B) Andosol soil frequency measured using the geographic coordinates of landrace accessions from the SEEDS dataset calculated using the soilP package (96). C) Phosphorus content on the *ZmPLA1.2^{CR}* mutants grown in long day conditions in Clayton, NC. D) PCA analysis of ionomics data of the *ZmPLA1.2^{CR}* mutants and control plants grown in long day conditions in Raleigh.

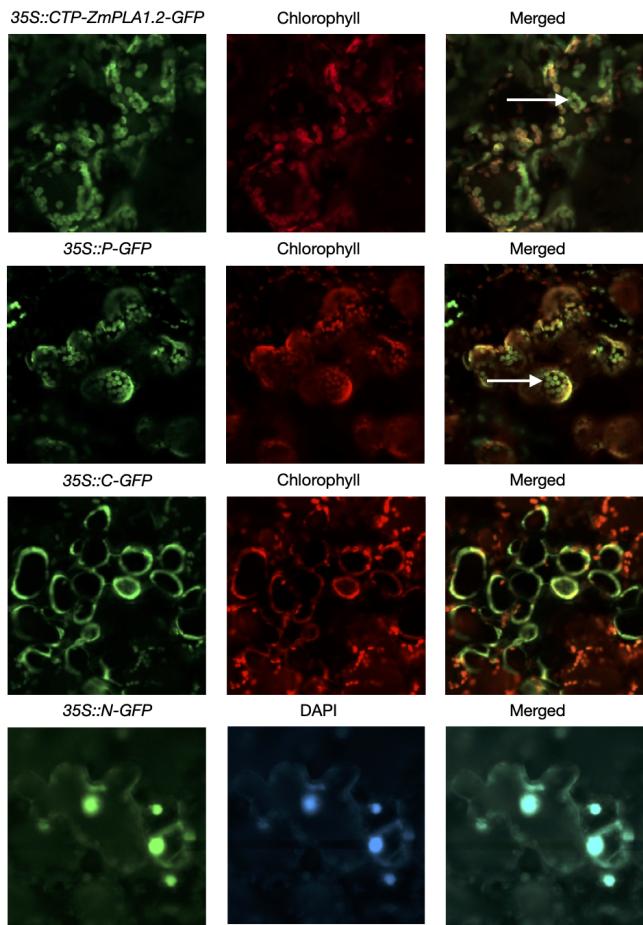


Figure S7 Subcellular localization of *ZmPLA1.2*.

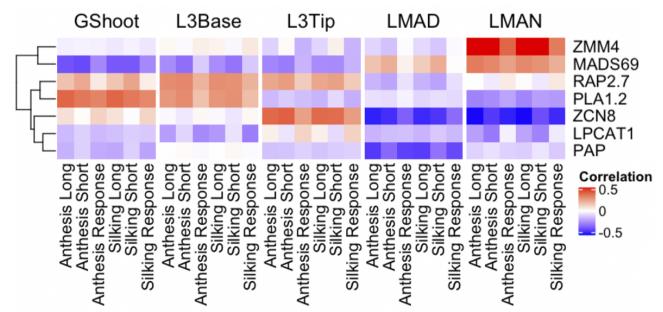


Figure S8 Flowering time and phospholipid related gene expression correlation with flowering time traits in aerial tissues in the 282 panel. Data obtained from (54).