

Cover Art



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

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

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, Sarah Jensen^e, Juan M Estévez-Palmas^b, Garrett M Janzen^d, 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, ^eUS 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 lower temperatures that impose a strong selection on flowering time. Linkage mapping and genome scans identified a highland maize loss of function of *ZmPLA1.2*, a gene encoding a phospholipase A1 enzyme as the major driver of high PC/LPC ratios. We then showed that the highland *ZmPLA1.2* allele was introgressed from teosinte *mexicana* and that it has been conserved in maize from the Northern US and Europe. We showed that genetic variation in *ZmPLA1.2* exhibits a strong genotype by environment interaction in common garden experiments and that the highland *ZmPLA1.2* allele leads to higher fitness in highland environments possibly via a reduction of flowering time. 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.

phospholipid metabolism; maize genetics; highland adaptation; convergent evolution

1 Elevation gradients are associated with changes in environmental
 2 factors that impose constraints on an organism's physiology.
 3 Organisms adapt to highland environments via selection
 4 of genetic variants that improve their physiological ability to
 5 cope with these constraints, including lower oxygen availability (1; 2; 3; 4), higher UV-radiation (5) and lower temperatures
 6 (6; 7). In particular, lower temperatures can significantly reduce
 7 growing season length and select for accelerated development.
 8

9 After domestication from the wild relative teosinte parviflora (Zea mays spp. *parviflora*) (8; 9) in the lowland, subtropical
 10 environment of the Balsas River (Guerrero, México) around 9,000
 11 BP, maize (Zea mays spp. *mays*) expanded throughout México and
 12 reached the highland valleys of central México around 6,500 BP
 13 (10).

14 In México, highland adaptation of maize was aided by significant adaptive introgression from a second subspecies of teosinte,
 15 teosinte *mexicana* (Zea mays spp. *mexicana*) that had already
 16 adapted to the highlands of México thousands of years after
 17 the split from teosinte *parviflora* (11; 12). Phenotypically, the
 18 most evident signs of *mexicana* introgression into maize are the
 19

20 high levels of stem pigmentation and pubescence (13) that are
 21 supposed to protect against high UV radiation and low temperatures.
 22 The ability to withstand low temperatures and efficiently
 23 photosynthesize in early stages of seedling development is a key
 24 component of maize highland adaptation (14), and recent RNA-
 25 Seq analysis of the effects of the inversion *Inv4m*, introgressed
 26 from *mexicana*, support this hypothesis (15). *Inv4m* is also asso-
 27 ciated with shorter flowering times in highland maize (16; 17).
 28 Given the low growing degree unit accumulation in highland
 29 conditions, there has been selection for shorter flowering times
 30 in highland-adapted maize (17). Other possible adaptive traits
 31 originating from *mexicana* may include ability to grow in soils
 32 with low phosphorous (18; 19) that are characteristic of the acidic
 33 soils of the highland valleys of the Transmexican volcanic belt
 34 (20).

35 By the time that maize reached the Mexican highlands, its
 36 range had already expanded far to the south, including coloniza-
 37 tion of highland environments in the Andes (21; 22). Andean
 38 maize adaptation occurred without *mexicana* introgression, as
 39 there is no wild teosinte relative in South America. These multi-
 40 ple events of maize adaptation to highland environments consti-
 41 tute a good system to study the evolutionary and physiological
 42 mechanisms of convergent adaptation (23; 24).

43 In comparison to southward expansion, northward migra-
 44

Manuscript compiled: Thursday 24th December, 2020

¹These authors contributed equally to this work.

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

tion 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; 34; 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 (35).

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 (36) and reduce the levels of unsaturated fatty acids in glycerolipids (37; 38), 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 (39), particularly in older leaves. Finally, certain species of phosphatidylcholine (PC), the most abundant phospholipid (40; 41; 42), can bind to *Arabidopsis* flowering locus T and accelerate flowering time through a yet unknown mechanism (43), and glycerolipid content in maize has a predictive power for flowering time (44). 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 biparental 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

Phospholipid pathways are under selection in highland maize
We analyzed selection of glycerolipid-related pathways both at the gene and metabolite level in highland maize using several approaches.

Population Branch Excess analysis of glycerolipid pathway genes. Li Wang and collaborators (24) find strong indications of convergent adaptation to four highland populations: Southwestern US (SWUS), Mexican Highland (MH), Guatemalan Highlands (GH) and Andes using the Population Branch Excess (PBE) statistic (45). In this paper we used the calculated PBE values from (24) to evaluate the extent of convergent adaptation in glycerolipid-related pathways in the same four highland populations (Figure 1A).

Among the ≈ 600 genes involved in glycerolipid metabolism (See), we identified a significant excess of genes that were targets of selection in more than two populations ($p < 2.87e - 5$, Figure S1A). The most over-represented intersection of selected genes was SWUS-MH-GH ($p = 1 \times 10^{-15}$), 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. These three populations also showed the highest number of genes that were recurrently selected in at least three populations. We found 22 genes that were consistently PBE outliers in all four populations ($p = < 1 \times 10^{-10}$). 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 (Figure 1B, Supplementary S1B and Supplementary File 1). 'Triacylglycerol degradation' and 'galactolipid biosynthesis' were recurrently selected in the Southwestern US, Mexican Highlands and the Andes while others involved in sulfolipid, diacylglycerol and phosphatidylglycerol biosynthesis were selected in the Southwestern US and Mesoamerican populations (Figure 1B).

ZmLPCAT1 is an example of a gene selected, with shared outlier SNPs in the CDS region, in all four populations that is part of

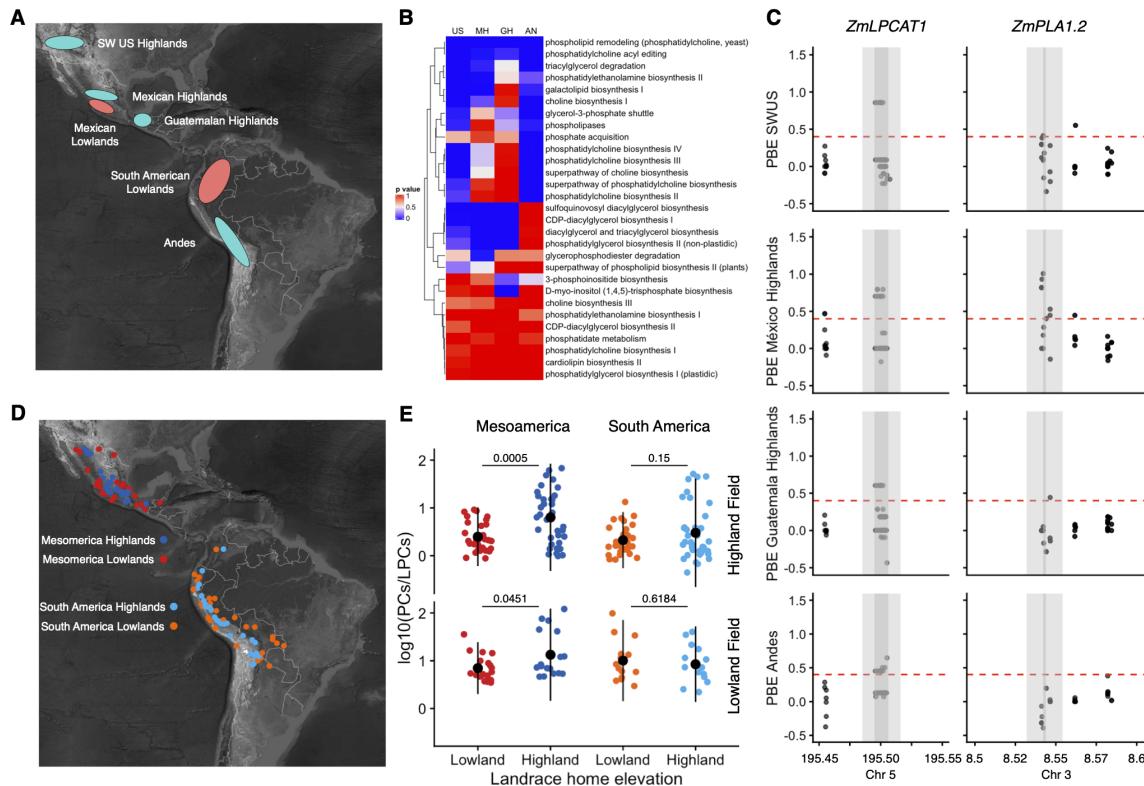


Figure 1 Glycerolipid pathway selection. A) Populations used in the Population Branch Excess (PBE) analysis. B) Highland selection of Glycerolipid pathways using PBE. See methods for details. C) PBE values of SNPs in *ZmLPCAT1* and *ZmPLA1.2*. D) Map showing the geographical origin of the 120 accessions used in the common garden experiment to quantify glycerolipid levels. E) Logarithmic values of the PCs/LPCs ratio of highland and lowland landraces from MesoAmerica and South America.

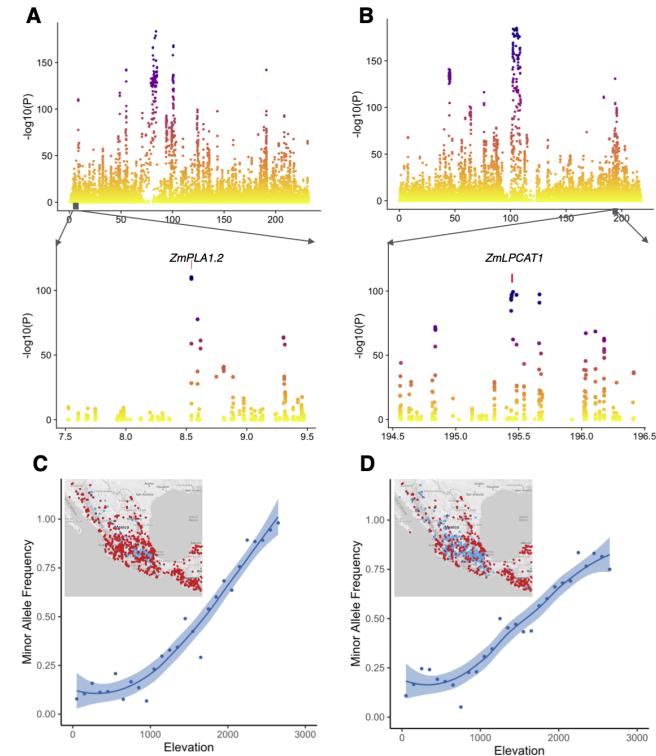
the 'phospholipid remodelling' and 'PC acyl editing' pathways ((Figure 1C), Supplementary File 1). *ZmPLA1.2* performs the reverse reaction of *ZmLPCAT1* and is part of the PC acyl editing, triacylglycerol degradation and phospholipases pathways. *ZmPLA1.2* is a good example of an outlier gene in the 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 1C).

Modes of convergent adaptation constraint. We previously defined two possible explanations for the extent of convergent selection in genes repeatedly selected in several highland populations (24; 46). 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 (46) 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).

Selection on glycerolipid metabolites. We then evaluated selection at the phospholipid level. We grew a diversity panel composed of 120 highland and lowland landraces from mesoamerica and South America (Figure 1D) 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 (47). 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 S1C-D, Supplementary File 2). In particular, one of the species with the highest Q_{ST} values is PC 36:5.

223 ***pcadapt* analysis of biological adaptation of Mexican landraces**

224 We then used Genotyping By Sequencing (GBS) data from 2700
 225 geo-referenced landraces from México generated by the SeeD
 226 project (16; 17) to run a *pcadapt* analysis that detects how strongly
 227 loci are contributing to patterns of differentiation between major
 228 principal components of genetic variation (48). The first principal
 229 component of *pcadapt* polarized Mexican landraces based on ele-
 230 vation of the geographical origin of the landrace (Supplementary
 231 Figure S2A). Using this first principal component we identified
 232 outlier SNPs across the genome that are significantly associated
 233 with elevation of origin of the landrace and are potentially in-
 234 volved in elevation dependent local adaptation (Supplementary
 Figure S2B). We found that from the 153 glycerolipid-related



235 **Figure 2** Manhattan plots of minus log10(P-values) *pcadapt* outliers. A) and B) *pcadapt* PC1 outliers plots of chromosome 3 and
 236 5, respectively. Lower panels are zoomed areas of outlier SNPs
 237 that co-localize with the physical position of the coding se-
 238 quences (marked with a red line) of *ZmPLA1.2* and *ZmLPCAT1*.
 239 C) and D) show the geographic and elevation dependent mi-
 240 nor allele frequencies of the highland (blue) and lowland (red)
 241 alleles of one of the outlier SNPs in the coding sequence of
 242 *ZmPLA1.2* and *ZmLPCAT1*

243 genes that were PBE outliers in Mexican highlands, 38 of them
 244 where also *pcadapt* PC1 outliers (top 5% -log(P)) (Supplementary
 245 File 1). We also found that genes involved in phospholipid re-
 246 modelling had significantly high -log(P) values indicating strong
 247 selection with elevation. In fact, both *ZmPLA1.2* and *ZmLPCAT1*
 248 contained SNPs with very high *pcadapt* -log10(P) values within
 the genes' coding regions (Figure 2A-B) reflecting strong eleva-
 tion dependent allele frequency changes (Figure 2C-D).

249 All taken together our data shows that phospholipid path-
 250 ways, in particular genes involved in determining PC/LPC ratios,
 251 like *ZmPLA1.2* and *ZmLPCAT1*, show clear signs of recurrent
 252 selection across several highland populations both at the genetic
 253 and metabolic level.

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

256 To break population structure and identify loci involved in phos-
 257 pholipid synthesis in highland maize, we developed a Backcross
 258 Inbred Line BC1S5 population, between B73 (a temperate inbred
 259 line) and Palomero Toluqueño (a Mexican highland landrace)
 260 using B73 as the recurrent parent (75% B73, 25% PT). Palome-
 261 ro Toluqueño (PT) accession Mexi5 (CIMMYTMA 2233) is a
 262 popcorn (Palomero means popcorn in Spanish) from the Toluca
 263 valley in México (Figure 3A). The Hilo landrace panel and the
 264 B73 x PT BC1S5 mapping population were grown on the same
 265 highland and lowland common gardens and samples for gly-
 266 cerolipid analysis were collected. In highland conditions, with
 267 typical 5 growth degree units across the growth season, Palome-
 268 ro Toluqueño shows higher fitness than B73 (Figure 3A-B). While
 269 B73 typically flowers around 65 days after planting in US tem-
 270 perate conditions and Mexican lowland conditions, B73 flowers
 271 around 150 days after planting in our Toluca field (Figure 3A)
 272 Using the sum of LPC species, we found a major QTL peak
 273 (LOD = 9.2, 53% of phenotypic variance explained) located at
 274 8.5 Mb of chromosome 3 (AGPv3) *qLPCs3@8.5* (Figure 3B). We
 275 also found a major QTL peak *qPCs3@8.5* in the same locus as
 276 *qLPCs3@8.5* when we use the sum of PC species (PCs) (Figure
 277 3B), (LOD = 5.6, 37% of phenotypic variance explained). The
 278 PCs/LPCs ratio also showed a major QTL *qPCs3/LPCs3@8.5* on
 279 the same locus as *qLPCs3@8.5* and *qPCs3@8.5* with an even larger
 280 LOD, (LOD = 24.5, 87% of phenotypic variance explained). We
 281 searched for epistatic effects in LPCs, PCs, and PCs/LPCs ratios
 282 through a combination of R/qtl scantwo and stepwise functions
 283 (?) but no additional significant QTLs were found. *qLPCs3@8.5*,
 284 *qPCs3@8.5* and *qPCs3/LPCs3@8.5* were robust to environmental
 285 effects and were found in BILs grown in highland and lowland
 286 environments. The additive effect of the PT allele at these QTLs
 287 lead to high levels of PCs, low levels of LPCs and consequently
 288 high PC/LPC ratios while we observed the opposite effect for the
 289 B73 allele (Figure 3C, top panel). Individual PC and LPC QTLs at
 290 this locus show the same additive PT allele effect behaviour than
 291 the summary *qLPCs3@8.5* and *qPCs3@8.5* (Figure 3C, top panel
 292 and Supplementary S4). All individual LPC QTLs at the *qLPCs3*
 293 locus correspond to LPCs that contain at least one double bond
 294 in the fatty acid (Figure 3D, Supplementary Fig 3, Supple-
 295 mentary file 3). The summary *qPCs3@8.5* was driven mainly by PC
 296 species with more than 2 fatty acid double bonds such as PC
 297 36:5 (Figure 3C and Supplementary S4 bottom panel) We then
 298 sought to identify the potential candidate gene underlying the
 299 QTLs at Chr 3. The QTL 7.9-10 Mb 1.5 LOD drop confidence
 300 interval contained 72 genes. We hypothesized that the metabolic
 301 phenotypes we observed could be due to a gene that is involved
 302 in the process of PC-LPC conversion. There are 75 genes in
 303 the maize genome with predicted phospholipase activity (Sup-
 304 plementary Figure S3A) and half of them have predicted PLA1
 305 activity (Supplementary Figure S3A). We identified *ZmPLA1.2*
 306 (Chr3:8,542,107..8,544,078), right at the QTL peak, as the most
 307 likely candidate (Figure 3B). *ZmPLA1.2* has a predicted Phospho-
 308 lipase A1-Igamma1 activity and can be classified, based on its
 309 two closest Arabidopsis orthologs (At1g06800 and At2g30550),
 310 as a PC hydrolyzing PLA1 Class I Phospholipase (49). PLA1
 311 phospholipases hydrolyze phospholipids in the sn-1 position
 312 and produce a lyso-phospholipid and a free fatty acid as a result
 313 (Supplementary Figure S3B). Class I Phospholipases are targeted
 314 to the chloroplast and in fact we identified a Chloroplast Transit
 315 Peptide at the beginning of the CDS of the gene (Figure 4A) using

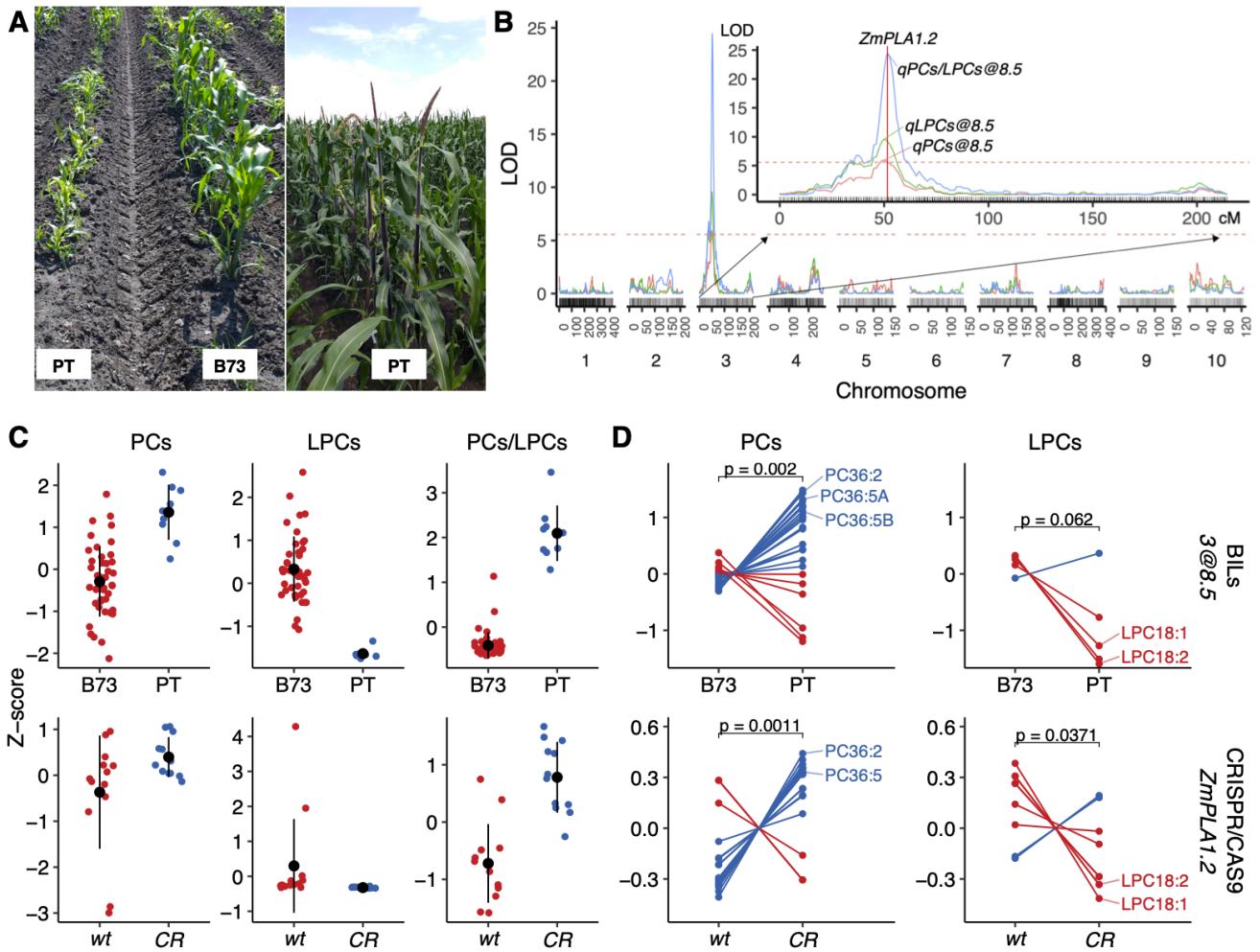


Figure 3 QTL analysis of phospholipid content in a B73 x PT BIL population. A) PT and B73 plants growing in the highland Metepec field. B) QTL analysis using data collected from plants growing in the highland and lowland fields of PCs, LPCs and PCs/LPCs ratio identified overlapping major QTLs at 8.5 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 chr 3 8.5 Mb (top row) and CRISPR-CAS9 *ZmPLA1.2^{CR}* mutant

ChloroP (50). We further confirmed chloroplast localization by transiently expressing the *ZmPLA1.2* Chloroplast Transit Peptide fused with GFP in *Nicotiana benthamiana* leaves (Supplementary Figure S3C). 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 then 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.

Mode of action of *ZmPLA1.2*

Our PBE, *pcadapt*, $Q_{ST}-F_{ST}$ and QTL data strongly suggest that the PC-LPC balance is under selection in highland Mexican maize and that *ZmPLA1.2*, and to a minor extent, *ZmLPCAT1* are themselves under selection and are major drivers of the lipid changes observed in highland maize. Furthermore, the QTL data suggest

that the highland PT allele is a loss of function of *ZmPLA1.2*. This 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 4B). Under cold conditions *ZmPLA1.2-B73* was up-regulated but *ZmPLA1.2-PT* was not (Figure 4B). F1 plants showed a similar expression pattern to B73 plants. *ZmPLA1.2* on the F1 showed a pattern of expression consistent 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 S6A). 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 4A, Supplementary Figure S5) in BIL

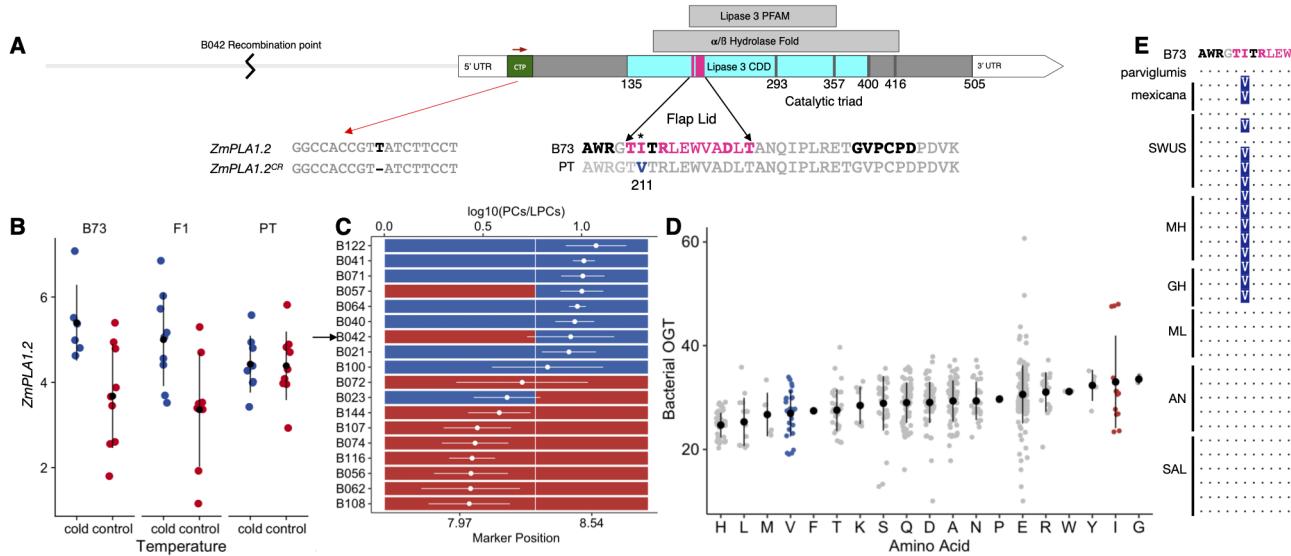


Figure 4 Analysis of the underlying causes of *ZmPLA1.2* variation on PC/LPC ratios. A) *ZmPLA1.2* promoter and coding sequence showing different features. CTP represents the chloroplast transit peptide. The domains of *ZmPLA1.2* were analyzed from UniProt identifier A0A1D6MIA3. The Lipase3-PFAM - PF01764, 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) *ZmPLA1.2* Expression analysis of B73, PT and their F1 grown in control and cold temperatures in growth chamber conditions. C) PC/LPC ratios of several BILs including B042 that shows a recombination of event 500 bp upstream of the ATG. D) 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. E) 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

351 BIL B042 then contains a *ZmPLA1.2-PT* CDS but its pro-
 352 moter region, 500 base pairs upstream of the ATG, is mainly
 353 *ZmPLA1.2-B73*. PC/LPC levels on the B042 BIL were similar to
 354 other BILs that are homozygous PT at the 8.54 Mb marker in the
 355 QTL peak (Figure 4C). This result supports the hypothesis that
 356 the metabolic effect we see is due to a malfunction or a loss of
 357 function of the *ZmPLA1.2-PT* enzyme rather than changes in
 358 the *ZmPLA1.2-PT* regulatory region. We analyzed nucleotide
 359 diversity in the promoter and CDS of *ZmPLA1.2* of highland
 360 and lowland landraces from México and South America and we
 361 did not observe any obvious pattern when we compared high-
 362 land vs lowland landrace (Supplementary S6B). Using Sanger
 363 sequencing of PT BILs at the *ZmPLA1.2* locus we identified sev-
 364 eral non-synonymous SNPs within the CDS (407, 520, 553, 610,
 365 631, 1028, 1315, 1342, and 1345 from the ATG) that could have
 366 an effect on *ZmPLA1.2*. We focused our attention on SNP 631
 367 on the flap lid domain that leads to a conservative substitution
 368 from isoleucine to valine (V211I, Figure 4A). The flap lid domain
 369 is located in a lipase 3 domain that is highly conserved across
 370 the tree of life. We identified 982 observations of the PF01764
 371 lipase 3 PFAM domain in 719 prokaryote species using PfamScan
 372 (51; 52), and then calculated bacterial optimal growth tempera-
 373 tures from their tRNA sequences (53). We then tested if genetic
 374 variation in PF01764 residues was significantly associated with
 375 bacterial optimal growth temperatures. We found that all of the
 376 significant associations were located in the flap lid region (Figure
 377 4A, bold letters). We then specifically analyzed variation in the
 378 211 residue and observed that the PT allele (V) was associated
 379 with lower bacterial optimal growth temperatures than the B73
 380 allele (I) (Figure 4E) suggesting that the SNP we identified in
 381 PT leading to V211I may be associated with adaptation to low

temperatures that highland maize is usually exposed to. We
 382 then explored if this residue change was unique to PT or was
 383 conserved in other highland maize. The PT allele was conserved
 384 in highland landraces from México and Guatemala and was seg-
 385 regating in Southwestern US landraces. The B73 allele was fixed
 386 in lowland Mexican, South American and Andean landraces
 387 (Figure 4F). These results are consistent with the PBE results we
 388 have observed before (Figure 1C). This is a typical pattern of
 389 teosinte *mexicana* introgression (24) and indeed the PT allele was
 390 present in both teosinte *mexicana* accessions but only in 1/4 of the
 391 teosinte *parviflumis* accessions available in Hapmap 3 (55)
 392 (Figure 4F). This lead us to ask whether the PT allele was the
 393 result of teosinte *mexicana* introgression in highland maize or
 394 selected from *parviflumis* standing variation.

Introgession of *mexicana*

To test for *mexicana* introgression, we used f_d data from (12). f_d data around the *ZmPLA1.2* indicated that the region was intro-
 397 gressed from *mexicana* into highland maize. (Figure 5A). We then
 398 performed a haplotype network analysis using SNP data of the
 399 *ZmPLA1.2* CDS from 1160 Mexican homozygous accessions from
 400 the SeeD Dataset (16) and the teosinte TIL lines from Hapmap 3
 401 (55). We identified nine haplotype groups that clustered mainly
 402 based on elevation. (Figure 5B) The two major groups (II) and
 403 (VI) contained mainly lowland and highland landraces respec-
 404 tively. The two *mexicana* TIL lines (TIL08 and 25) were located
 405 in group IV (Figure 5A) together with highland landraces pri-
 406 marily collected in the Trans-Mexican Volcanic Belt (30/36 from
 407 highlands of Jalisco, Michoacán, México, Puebla and Veracruz).
 408 We then checked whether this *mexicana* *ZxPLA1.2* haplotype
 409 introgressed in mesoamerican highland maize is also present in

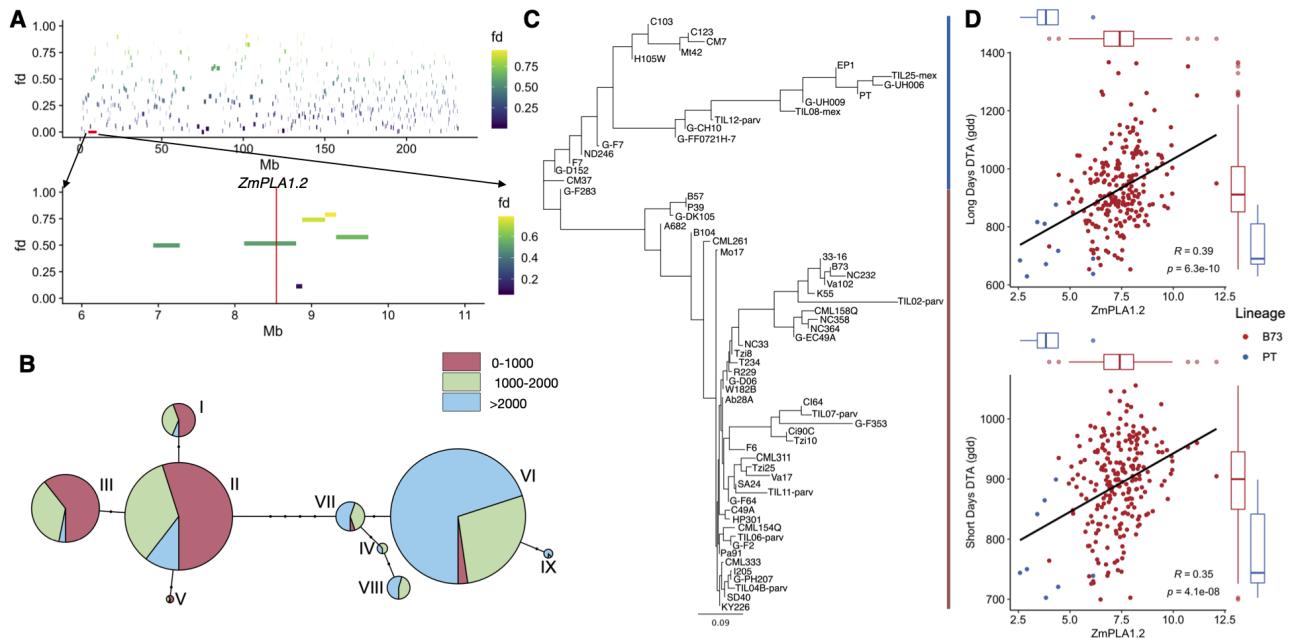


Figure 5 Introgession of teosinte *mexicana* into maize ZmPLA1.2. A) f_d analysis of *mexicana* introgression. Data was obtained from (12). B) Haplotype network analysis of ZmPLA1.2 CDS SNPs using 1060 Mexican homozygous individuals from the Seeds dataset. C) Phylogenetic tree of ZmPLA1.2 CDS using a sample of Hapmap3 inbreds and Palomero Toluqueño D) 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)

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 5C) flints like EP1, UH008, and UH009. Other northern US flints 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.

Fitness effects of ZmPLA1.2-PT

ZmPLA1.2 expression in modern maize is associated with flowering time. Building on previous reports that indicate a role of highly unsaturated PC species in determining flowering time (43; 44) 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 S7), 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 S7) 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 5D).

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 several fitness trait's intercept and slope on trial elevation using GridLMM (56). We found that genetic variation in ZmPLA1.2 showed significant effects of genotype by environment interactions on several fitness related traits. (Figure 6A). ZmPLA1.2 showed clear antagonistic pleiotropy effects on flowering time traits (Figure 6A). 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 phenocopied 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 6B). 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 6A, confirming an interaction between ZmPLA1.2 and environment.

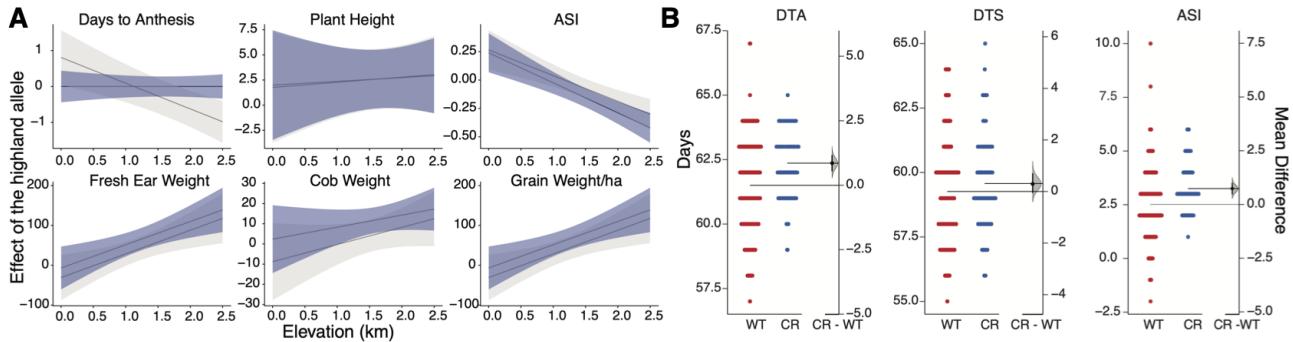


Figure 6 Fitness effects of *ZmPLA1.2-PT* and *ZmPLA1.2^{CR}*. A) We used BLUPs and GBS data from 2700 landraces from (17) evaluated in 23 common gardens at different elevations in México. We modeled each trait as a function of *ZmPLA1.2-PT* genotype, trial elevation, and tester line, with controls for 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 (56). 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

Association of *ZmPLA1.2^{CR}* with phosphorus levels. We measured phosphorus content in flag leaves of highland landraces from México, Perú and B73 grown in control field conditions. Phosphorus levels in highland landraces were higher than in B73 (Supplementary Fig S8A). 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 S8B). 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 S8C.) Principal component analysis analysis of ion levels in mutant and WT plants revealed no differences between the two genotypes (Supplementary Fig S8D.)

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 (49; 57) such as low phosphorus availability (58; 59; 39) and low temperature (36; 37; 60). Additionally, in Arabidopsis, accumulation of certain highly unsaturated phosphatidylcholine (PC) species can accelerate flowering time (43), 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 ma-

jor 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. 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. Finally we observed that 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.

We found that 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 1A-B, Supplementary S1A-B). A comparable pattern has been found during the adaptation of *Tripsacum dactyloides* to temperate latitudes. In this close relative to maize genes involved in PC metabolism showed accelerated rates of evolution (61). This further suggests that PC metabolism is important for adaptation to low temperature conditions that are prevalent at high elevation and latitude. *ZmLPCAT1* and *ZmPLA1.2* were two of the genes that showed strongest, repeated signals of selection measured by PBE and *pcadapt* in highland populations (Figures ??). 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). In Arabidopsis, *LPCAT1* is involved in the determination of phosphorus leaf content levels under low Zn conditions (62). 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. In fact, a previous study found that *ZmLPCAT1* showed high *Fst* values when comparing highland and lowland landraces (23). Natural variation in maize *ZmPLA1.2* is also associated with lipid content (63) and flowering time (64; 27). In B73, *ZmPLA1.2* is one of the most highly

559 expressed phospholipases and its expression pattern is almost
560 entirely restricted to vegetative leaves (V4-V9) (Supplementary
561 Figure S3C) (65), the same type of leaves that we sampled for glycerolipid
562 analysis. Additionally, *ZmPLA1.2* is highly expressed in
563 B73 and other temperate inbreds under low temperature conditions
564 and is downregulated in heat conditions (Supplementary
565 Figure S3D) (66).

566 Our QTL analysis of PC/LPC ratios in a B73 x PT mapping
567 population and in the *ZmPLA1.2^{CR}* mutant support the
568 hypothesis that the highland *ZmPLA1.2-PT* allele results in a loss
569 of function enzyme that rewrites highland Mexican maize PC
570 metabolism leading to high PC/LPC ratios (Figures 3). Adaptive
571 loss of function mutations can be an effective way to gain new
572 metabolic functions in new environmental conditions (67). Our
573 data supports an enzymatic loss of function due to a single conser-
574 vative amino acid substitution located in the flap lid domain
575 of *ZmPLA1.2* that can impact substrate accessibility and/or sub-
576 strate binding (Figure 4a). Indeed, flap lid domains are targets of
577 biotechnological modification of these types of enzymes (68).

578 Why were the metabolic changes induced by *ZmPLA1.2* se-
579 lected in highland maize? PC metabolism is intimately connected
580 to multiple stress response and developmental pathways; alter-
581 ations in PC amounts and PC/LPC ratios impact overall plant
582 fitness. The large *qPC/LPC3@8.5* QTL is driven by individual
583 QTLs of PC and LPC species with high levels of unsaturated
584 fatty acids (Figure S3). Some of these species, like PC 36:5 and
585 LPC 18:1 (Figure 3D, Supplementary File 3), show similar pat-
586 terns during *Arabidopsis* cold acclimation (37) and sorghum low
587 temperature response (60). PC 36:5 also showed very high Q_{ST}
588 values when comparing highland and lowland landraces from
589 both mesoamerica and South America (Supplementary Figure
590 S1C-D, Supplementary File 2). In maize *ZmPLA1.2* expression
591 is circadianly regulated (69) and peaks at the end of the day. In
592 *Arabidopsis*, highly unsaturated PC (34:3, 34:4, 36:5, 36:6) species
593 increase during the dark (70), coinciding with low expression lev-
594 els of *PLA1.2* (69). Furthermore, Yuki Nakamura and colleagues
595 elegantly showed that PC 36:5 and 36:6 species accumulate dur-
596 ing the night and can bind to *Arabidopsis* flowering time locus
597 T (FT) accelerating flowering time (43). The *Arabidopsis* FT or-
598 tholog in maize, *ZCN8* (71) underlies a major flowering time
599 and photoperiod sensitivity QTL (27). Additive mutations in
600 the regulatory region of *ZCN8*, including a teosinte *mexicana*
601 introgression, lead to higher expression of *ZCN8* contributing
602 to maize adaptation to long days in temperate conditions (72).
603 Our results show that *ZmPLA1.2* drives the accumulation of PC
604 species (like PC 36:5) that accumulate in *Arabidopsis* at the end
605 of the day and that can bind to FT. We hypothesize that accumula-
606 tion of these PC species in highland maize could drive early
607 flowering in a similar way that occurs in *Arabidopsis*.

608 Our results showed that the flap-lid mutation in *ZmPLA1.2-PT*
609 residue 211 is conserved in highland Mexican and
610 Guatemalan landraces and in teosinte *mexicana* (Figure 4E) but
611 is still segregating in teosinte *parviflora*. We showed that indeed,
612 *ZmPLA1.2-PT* is an introgression from teosinte *mexicana*
613 (Figure 5A) and that this introgression has been conserved in
614 modern inbred lines particularly northern US, Canada and EU
615 Flints (Figure 5C). *ZmPLA1.2* in inbreds carrying the *ZxPLA1.2*
616 *mexicana* haplotype show low levels of expression and earlier
617 flowering times (Figure 5D) (54). Moreover, genetic variation in
618 the regulatory region of *ZmPLA1.2* was significantly associated
619 with photoperiod sensitivity in the maize Nested Association
620 Mapping population (27). The effect of genetic variation of *Zm-*

621 *PLA1.2* on flowering time in Mexican landraces indicated a strong
622 GxE interaction where the highland allele lead to a reduction of
623 flowering time and ASI in highland environments similar to the
624 effect observed to the well known teosinte *mexicana* introgression
625 of inversion *inv4m* (15). Other flowering time loci analyzed
626 in the same experiment do not show this clear GxE effect (17).
627 Interestingly, for yield-related traits, the effect of the highland
628 allele shows typical conditional neutrality with increased fitness
629 of the *ZmPLA1.2-PT* allele in highlands. Conversely, the previ-
630 ously characterized *mexicana* introgression *textitinv4m* (12; 73; 11)
631 shows no effect in highlands and negative effect in lowlands (?).

632 To the best of our knowledge this is the first documented case
633 of an adaptive teosinte *mexicana* introgression with a character-
634 ized metabolic phenotype and fitness advantage in highlands.

Materials and Methods

Populations used in the analysis

635 Highland and lowland populations used for Population Branch
636 Excess analysis consisted of three to six accessions from each of
637 the highland and lowland populations and have been previously
638 described in (24; 34).

639 The 120 Landraces from the HiLo Diversity Panel were se-
640 lected and ordered from the CIMMYT germplasm bank to maxi-
641 mize a good latitudinal gradient sampling across Mesoamerica
642 and South America. For each highland (>2000 masl) landrace
643 a lowland (<1000 masl) sample was selected at the same lati-
644 tude (<0.5°) to form 60 highland/lowland pairs, 30 from each
645 continent. The list of the accessions used is provided in Supple-
646 mentary file 4.

647 B73 x Palomero Toluqueño Backcross Inbred Lines (BILs) were
648 developed by crossing B73 with a single Palomero Toluqueño
649 plant (Mex5 accession, CIMMYTMA 2233) that was then back-
650 crossed with B73 once and selfed five times (BC1S5).

651 We used individual landrace accession genotype and fitness
652 data from the CIMMYT Seeds of Discovery project (SeeD) (17) to
653 calculate *pcadapt* (48) values and GxE effects of *ZmPLA1.2-PT*.

Field Experimental Conditions and sampling

654 Two replicates of the HiLo Diversity panel accessions and three
655 replicates of the B73 x PT BILs were planted in a field located in
656 Metepec, Edo de México, (19°13'28.7"N 99°32'51.6"W) within the
657 Trans-Mexican volcanic belt. The field is at 2610 meters above
658 sea level (masl), and the range of average monthly temperatures
659 along the year vary from 5 °C to 21.5 °C with an average annual
660 of 13.6 °C. Fifty mg of fresh tissue was collected using a leaf
661 punch from the tip of the second youngest leaf above the last
662 leaf with a fully developed collar around V4-V6 developmen-
663 tal stage. Tissue discs were immediately flash frozen in liquid
664 nitrogen. All samples were collected in a single day between
665 10:00 am and 12:00 pm, approximately 3 h after sunrise. Samples
666 were transported in dry ice to the lab and stored at -80°C until
667 extraction.

Glycerolipid extraction and UHPLC-QTOF MS/MS analysis

668 Frozen material was homogenized in a tissue grinder Retsch
669 (Haan, Germany) for 40 seconds at a frequency of 30 1/s. After
670 grinding, samples were extracted. We performed lipid extraction
671 following Matyash and collaborators (74). First, 225 µL of cold
672 methanol (MeOH), was added to each sample. For the blanks,
673 MeOH previously prepared with a Quality Control (QC) mix was
674 added (Supplementary File 5). Each sample was vortexed for
675 10 seconds, keeping the rest of the material on ice. Then 750 µL
676 of cold methyl tert-butyl ether (MTBE) were added. The MTBE
677 were added. The MTBE

added to the blanks contained 22:1 cholesterol ester as internal standard (Supplementary File 5). Each sample was vortexed for 10 seconds, followed by 6 minutes of shaking at 4°C in the orbital mixer. 188 μ L of LC/MS grade water at room temperature (RT) was added, and the samples were vortexed for 20 seconds. We centrifuged the samples for 2 min at 14000 rcf and recovered 700 μ L of supernatant from the upper organic phase. We then split the supernatant into two aliquots of 350 μ L, 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. Dry samples were resuspended in 110 μ L of MeOH-Toluene 90:10 (with the internal standard CUDA, 50 ng/mL). Samples were vortexed at low speed for 20 s and then sonicated at RT for 5 min. Aliquots of 50 μ L per sample were transferred 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 Acuity charged surface hybrid (CSH) C18 2.1x100 mm 1.7 μ m column was set. The column was initially purged for 5 min. The UHPLC column was coupled to a VanGuard pre-column (Waters Acuity CSH C18 1.7 μ m). Six “no sample injections” were injected 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 μ L 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 maintained at 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.

Glycerolipid data processing

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. In Microsoft Excel, using the method retention time for internal standards and the current rt, a polynomial regression was obtained and used for calculating new retention times for 501 lipids listed in a MS1 m/z-rt library from Dr. Oliver Fiehn's laboratory (See Supplementary File 6). In MSDIAL (75), 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 (75) 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) (76). 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 (77). Curated data was also normalized using sum known metabolite signal (mTIC). After data processing and normalization, lipid intensities were used for further analysis.

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 (78) (map identifiers: zma00564 and zma00561). With the NCBI Entrez gene identifiers in KEGG we retrieved the AGPv4 transcript identifiers used in Corncyc 8.0 (79; 80) from an id cross reference file found in MaizeGDB () (79). 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 pseudo-gene 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 (81).

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. parviglumis* 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. After calculating genome wide PBE scores for the 4 populations (described in detail in (24)), we tested for selection outliers SNPs in the 594 phosphoglycerolipid candidates and the 30 Corncyc pathways (556 genes). We first defined a 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. We defined a gene as a PBE outlier if it contained an outlier SNP within the coding sequence or 10 Kbp upstream/downstream (24). Then we tested for over-

804 representation of genes selected in particular subsets of populations
 805 using Fisher's exact test using the 32283 protein genes from
 806 the maize genome (Supplemental Figure 1a) (82) as background.
 807 For each pathway, we first selected all SNPs that include CDS
 808 regions and 10Kb upstream and downstream of the gene and we
 809 calculated the mean pathway PBE score. We then constructed a
 810 null distribution by drawing 10000 samples without replacement
 811 of n SNPs from those found within or around 10Kb upstream
 812 and downstream of all protein coding genes and we obtained the
 813 mean PBE for this null distribution. With the set of PBE outliers
 814 for glycerolipid metabolism in the 4 populations we tested for
 815 evidence of physiological or pleiotropic constraint using the C^2_{χ}
 816 statistic (46).

817 ***QST*-FST analysis of glycerolipid data**

818 Quantitative trait divergence (Q_{ST}) was contrasted to the
 819 distribution of F_{ST} for neutral genetic markers (83). Highland/
 820 Lowland contrasts were considered separately for
 821 Mesoamerica and South America.

822 A linear mixed effects model (R package lmer, function lmer)
 823 was used to partition phenotypic variance between population
 824 pairs (Mesoamerica/South America, all highland/all lowland,
 825 Mesoamerican highland/Mesoamerican lowland, South Ameri-
 826 can highland/South American lowland).

$$827 \text{TRAIT} \sim 1 + (1|\text{POPULATION}) + \\ 828 (1|\text{GARDEN/BLOCK}) + (1|\text{BATCH})$$

829 Within-population and between-population variances were
 830 calculated with the R function VarCorr (R package lme4, 84), and
 831 were used to calculate Q_{ST} following the equation below:

$$832 Q_{ST} = \sigma_{GB}^2 / (\sigma_{GB}^2 + 2\sigma_{GW}^2)$$

833 in which σ_{GB}^2 and σ_{GW}^2 are the between- and within-population
 834 genetic variance components, respectively (47).

835 Pairwise F_{ST} was calculated with the R function
 836 fst.each.snp.hudson (R package dartR, 85). Q_{ST} values
 837 were considered significantly high or low if they were greater or
 838 less than than two standard deviations from the mean F_{ST} .

839 ***pcadapt* analysis of biological adaptation in Mexican landraces**

840 In order to conduct genome scans for signatures of adaptation
 841 we used the pcadapt (48) package. *pcadapt* identifies adaptive
 842 loci by measuring how strongly loci are contributing to patterns
 843 of differentiation between major axes of genetic variation. Under
 844 simple models the principal component based models of pcadapt
 845 capture major patterns of F_{ST} outlier tests but are conducted in a
 846 way that does not require population delimitation (86). As the
 847 genome scan comparison requires a focal SNP to be compared to
 848 the first K principal components of the genotype data, it can be
 849 biased by large regions of low recombination that drive the major
 850 axes of variation in the principal components. Thus, when SNPs
 851 from these low recombination regions are compared against
 852 principal components driven by linked loci spurious signals may
 853 arise. To prevent this bias from occurring, we used custom scripts
 854 to calculate the principal component step separately based upon
 855 all the chromosomes except for the chromosome of the focal
 856 SNPs being tested. The genotype data we used for this analysis
 857 was GBS data from roughly 2,000 landraces of Mexican origin
 858 collected by CIMMYT (www.cimmyt.org) as part of the SeeDs of
 859 discovery initiative (<https://www.cimmyt.org/projects/seeds-of-discovery-seed/>). From this, we calculated the strength of

861 association between each SNP and the first five principal compo-
 862 nents (excluding the chromosome of the focal SNP) using the
 863 communality statistic as implemented in *pcadapt* version 3.0.4.

864 ***QTL* analysis of phospholipid levels**

865 We analyzed glycerophospholipid QTLs in a mapping popula-
 866 tion of 57 BILs (BC1S5) from the cross B73 x (PT). These BILs
 867 were grown in a highland site in Metepec, Edo de Mexico at 2600
 868 masl during the Summer of 2016 and in Puerto Vallarta, Jalisco at
 869 50 masl during the Winter of 2016/17. We analyzed the samples
 870 using UHPLC-QTOF, as above, and 67 leaf lipid species were
 871 identified. For QTL analysis we calculated the mean across all
 872 fields of individual lipid mass signal. We also used as pheno-
 873 types the sum total of the following lipid classes: diacylglycerol,
 874 triacylglycerol, PC and LPC. Furthermore, we also included the
 875 log base 10 transformed ratios of LPCs/PCs and the ratios of
 876 their individual species. We did a simple single marker analysis
 877 with "scanone" using Haley-Knott regression, and assessed the
 878 QTL significance with 1000 permutations.

879 ***CRISPR-Cas9* editing of *ZmPLA1.2* and analysis of the effect 880 of *Zmpla1.2^{CR}* mutant on flowering time**

881 CRISPR/Cas9 was used to create a *ZmPLA1.2* gene knockout
 882 through agrobacterium mediated transformation of background
 883 line B104 (87; 88). Guide RNA was designed as described in (89)
 884 for the B73 reference genome v4. B104 and B73 sequence for
 885 *ZmPLA1.2* were identical. gRNA cassette was cloned into pGW-
 886 Cas9 using Gateway cloning. Two plants from the T0 transgenic
 887 event were identified through genomic PCR amplification and
 888 Sanger sequencing and were self-pollinated. Several T1 plants
 889 containing the *ZmPLA1.2^{CR}* event were selected and planted for
 890 lipid analysis in greenhouses at the North Carolina State Univer-
 891 sity during the 2020 Spring, genotyped using forward primer
 892 CAGTTCTCATCCATGCACG and reverse primer CCTGAT-
 893 GAGAGCTGAGGTCC and self-pollinated. Cas9 positivity was
 894 tested for using 0.05% Glufosinate ammonium contained in Lib-
 895 ery herbicide. T2 seeds from CAS9 free T1 plants were collected
 896 and used for flowering time analysis in Clayton, NC during the
 897 2020 Summer.

898 ***Subcellular localization of ZmPLA1.2***

899 We fused the 52 *ZmPLA1-B73* Chloroplast Transit Peptide was
 900 fused. Three constructs encoding subcellular localization signals
 901 were used as control; Cytoplasm (C-GFP), nucleus (N-GFP), and
 902 Chloroplast (P-GFP). All of them under control of the 35S pro-
 903 moter. This constructs were transiently expressed in *Nicotiana*
 904 *benthamiana* leaf cells.

905 ***Sanger sequencing of ZmPLA1.2 in BILs homozygous at the 906 qPC/LPC3@8.5 locus***

907 We identified 3 BILs homozygous PT at the *ZmPLA1.2* locus and
 908 we developed 6 set of primers to Sanger sequence across the CDS
 909 and the gene promoter. Primers with location in the gene are
 910 shown in Supplementary file 7.

911 ***Expression analysis of ZmPLA1.2 in B73, PT and B73xPT F1s 912 in conditions simulating highland environments***

913 Gene expression data was generated from leaf tissue from B73,
 914 PT and the B73xPT F1. Plants were grown following the same
 915 protocol as in (15). Briefly, kernels were planted in growth cham-
 916 bers set to imitate spring temperature conditions in Mexican
 917 lowlands (22°C night, 32°C day, 12 hr light) and highlands (11°C
 918 night, 22°C day, 12 hr light). Leaf tissue was sampled from the
 919 V3 leaf the day after the leaf collar became visible, between 2
 920 and 4 hours after lights came on. Tissue was immediately placed

921 in a centrifuge tube, frozen using liquid nitrogen, and stored at
922 -70°C.

923 RNAseq libraries were constructed, sequenced, and analyzed
924 following (15). In summary, randomly primed, strand specific,
925 mRNA-seq libraries were constructed using the BRaD-seq (90)
926 protocol. Multiplexed libraries were sequenced on 1 lane of
927 a Illumina HiSeq X platform. Low quality reads and adapter
928 sequences were removed using Trimmomatic v.0.36 (91), and
929 the remaining paired reads were aligned and quantified using
930 kallisto v.0.42.3 (92). Gene counts were normalized using the
931 weighted trimmed mean of M-values (TMM) with the *calcNorm-*
932 *Factors* function in *edgeR* (93), and converted to log2CPM.

933 **Bacterial optimal growth temperature association with Zm- 934 PLA1.2 flap-lid domain allelic variation**

935 The maize *ZmPLA1.2* protein sequence was compared to prokary-
936 otes with the same sequence to determine whether the identified
937 residue change in maize and accompanying association with
938 low temperature survival was consistent with observations in
939 other organisms. Pfam domain PF01764 was identified in the
940 B73 protein sequence using the HMMER3 web server, and 982
941 observations of the PF01764 Pfam domain were identified in
942 719 prokaryote species using PfamScan (51; 52). The optimal
943 growth temperature of these species was predicted using tRNA
944 sequences as in (53). Maize and prokaryote PF01764 domain
945 sequences were aligned with hmalign from the hmmer3 package
946 (94), and the aligned Pfam sequences were recoded to reflect nine
947 amino acid physicochemical properties (95). Sequences were fil-
948 tered to remove gaps in the domain alignment and observations
949 with only partial domain sequences, then clustered based on se-
950 quence similarity, resulting in two clusters of observations within
951 the domain. For each cluster, positions in the filtered alignment
952 were associated with prokaryote optimal growth temperatures
953 using a linear regression with all 9 amino acid physicochemical
954 properties. Seventeen sites in and around the flap-lid region of
955 the protein passed a 10% FDR significance threshold, including
956 the single residue change p.Ile211Val previously identified in
957 *ZmPLA1.2-PT*. Welch's two-sided t-test was used to compare
958 the optimal growth temperatures of prokaryote species with
959 the *ZmPLA1.2-B73* allele to the optimal growth temperatures of
960 prokaryote species with the *ZmPLA1.2-B73* allele at this site.

961 **Nucleotide Diversity of *ZmPLA1.2***

962 We estimated nucleotide diversity for the promoter and cod-
963 ing regions of *ZmPLA1.2* using WGS data from highland and
964 lowland accessions from México and South America obtained
965 from (34) and the R package PopGenome (96). FASTA files
966 were partitioned into 4 populations by origin supplemental
967 for accessions included) and then subset into coding and pro-
968 moter regions, which was defined as 3Kb upstream of the
969 cds. Data were imported into PopGenome using the option
970 include.unknown=FALSE in order to prevent bias by excluding
971 missing and ambiguous nucleotide data. Pi was measured sep-
972 arately within each population and averaged by the number of
973 sites in the coding and promoter regions.

974 **Teosinte mexicana introgression in highland maize**

975 To evaluate if the *ZmPLA1.2-PT* allele was the result of standing
976 variation from teosinte *parviflumis* or introgression from *mexicana*
977 we used Patterson's *D* statistic and genome-wide f_d to calculate
978 ABBA-BABA patterns. The data was obtained from (12). The
979 material used in (12) included whole genome sequence data from
980 three highland outbred individuals: two Palomero Toluqueño
981 and one Mushito de Michoacán; three lowland landraces: Nal-

Tel (RIMMA0703) and Zapalote Chico (RIMMA0733) obtained
982 from (34) and BKN022 from (55); two *mexicana* inbreds: TIL08
983 and TIL25; three *parviflumis* inbreds: TIL01, TIL05, TIL10 and
984 *Tripsacum* TDD39103 (55) as an outlier.

985 **Haplotype network analysis of *ZmPLA1.2* in Mexican maize lan- 986 draces and teosintes.**

987 We extracted SNP genotypes for *ZmPLA1.2* from the TIL teosinte
988 accessions HapMap 3 imputed data (55) and the 3700 Mexican
989 landraces in the SEEDs dataset. With the set of 1060 accessions
990 that were homozygous at all sites in this genomic region we
991 calculated an haplotype network depicting the minimal spanning
992 tree for haplotypes covering 90% of the input accessions with
993 the R package pegas (97), and haplotype frequencies for three
994 elevation classes in the landraces (0-1000,1000-2000, >3000 masl).

995 **Phylogenetic analysis of *ZmPLA1.2* in maize inbreds and 996 teosintes.**

997 Using v3 of B73 genome, *ZmPLA1.2* SNPs were obtained from
998 the 282-panel, the German inbreds, Palomero Toluqueño and
999 teosinte TIL inbreds from HapMap 3 (55). SNPs were aligned
1000 using Geneious2020.0.5 and a neighbor-joining phylogenetic tree
1001 was generated. To facilitate visualization and interpretation of
1002 the tree we condensed tree branches from lines with identical
1003 haplotypes and from similar geographic locations. The full tree
1004 is available as Supplementary file 8.

1005 **Expression analysis of candidate genes and association with 1006 flowering traits in the 282 panel**

1007 We used gene expression RNA-Seq data obtained from the 282
1008 panel at different developmental stages (54) and BLUP values of
1009 several flowering and photoperiod sensitivity traits (27) to study
1010 the correlation of *ZmPLA1.2* expression values in the 282 panel
1011 with flowering time traits.

1012 **Association of *ZmPLA1.2* with agronomic traits**

1013 We re-analyzed phenotypic data from the F1 Association Map-
1014 ping (FOAM) panel of Romero-Navarro *et al* (16) and Gates *et*
1015 *al* (17) to more fully characterize associations signatures of *Zm-*
1016 *PLA1.2*. Full descriptions of this experiment and data access are
1017 described in those references. We downloaded BLUPs for each
1018 trait and line from Germinal 3, and subseted to only those lines
1019 with GBS genotype data from México. We fit a similar model to
1020 the GWAS model used by Gates *et al* (17) to estimate the effect
1021 of the *ZmPLA1.2-PT* allele on the trait's intercept and slope on
1022 trial elevation, accounting for effects of tester ID in each field
1023 and genetic background and family effects on the trait intercept
1024 and slope using four independent random effects. We imple-
1025 mented this model in the R package *GridLMM* (56). We extracted
1026 effect sizes and covariances conditional on the REML variance
1027 component estimates and used these to calculate standard errors
1028 for the total *ZmPLA1.2-PT* effect as a function of elevation. To
1029 test whether the phenotypic effects of *ZmPLA1.2-PT* on yield
1030 components could be explained as indirect effects via flowering
1031 time, we additionally re-fit each model using Days-To-Anthesis
1032 as a covariate with an independent effect in each trial.

1033 **Effect of *Zmpla1.2CR* mutant on flowering time**

1034 Seeds of T2 CAS9-free plants were grown in isolation during
1035 the 2020 Summer in Clayton, NC. Female (Days to Silking) and
1036 male (Days to Anthesis) flowering time were calculated from
1037 the day of planting until the first silks and anther pollen shed
1038 could be observed in each individual plant. Plants homozygous
1039 *Zmpla1.2CR* and WT derived from the same T1 families were
1040 planted for this experiment. Effect size analysis was performed

1042 using the dabestR package (98).

1043 **Phosphorus analysis and phosphorus soil availability data**

1044 We analyzed phosphorus content in B73 and 5 Mexican and
1045 5 Peruvian highland landraces (Supplementary file 9) grown
1046 in control field conditions in the 2018 Puerto Vallarta, México,
1047 Winter nursery. Samples were analyzed using ICP-MS according
1048 to (99). Frequency of Andosol soils at different elevations was
1049 calculated using the soilP package (100). Phosphorus content in
1050 flag leaves of the the same *Zmpla1.2^{CR}* and WT mutants used for
1051 the flowering time experiment were analyzed using ICP-OES at
1052 the North Carolina Department of Agriculture.

1053 **Acknowledgments**

1054 This work was supported by Conacyt Young Investigator (CB-
1055 238101), Conacyt National Problems (APN-2983), UC-Mexus
1056 and NC State startup funds awarded to Rubén Rellán-Álvarez.
1057 Fieldwork and mapping population development was supported
1058 by NSF-PGR award 1546719 to Jeff Ross-Ibarra, Ruairidh Saw-
1059 ers, Daniel Runcie and Matthew Hufford. Dan Gates and Tay-
1060 lor Crow were supported by NSF-PGR award 1546719. Fausto
1061 Rodríguez-Zapata and Blöcher-Juárez were suported by Conacyt
1062 doctoral fellowships. Allison Barnes was supported by NSF-
1063 PGRP grant 201703. Andi Kur was supported as an AgBioFEWS
1064 fellow by the NSF award 1828820. We want to acknowledge
1065 the work of the Rellán-Álvarez and Sawers lab students that
1066 have helped generate the mapping populations developed in this
1067 manuscript. We specially want to thank the indigenous people
1068 of the Americas that through fascinating agricultural ingenuity
1069 domesticated and made possible the spread and adaptation of
1070 maize throughout the continent. Field experiments were carried
1071 out in Wixárika (Puerto Vallarta, México); Nahuatl, Matlatzinca,
1072 and Mazahua (Metepic, México); and Tuscarora (Clayton, NC)
1073 land.

1074 **References**

- 1075 [1] C Natarajan, et al., Predictable convergence in hemoglobin
1076 function has unpredictable molecular underpinnings. *Science* **354**, 336–339 (2016).
- 1077 [2] X Yi, et al., Sequencing of 50 human exomes reveals adap-
1078 tation to high altitude. *Science* **329**, 75–78 (2010).
- 1079 [3] A Bigham, et al., Identifying signatures of natural selection
1080 in tibetan and andean populations using dense genome
1081 scan data. *PLoS genetics* **6**, e1001116 (2010).
- 1082 [4] X Liu, et al., EPAS1 gain-of-function mutation contributes
1083 to high-altitude adaptation in tibetan horses. *Molecular
1084 biology and evolution* (2019).
- 1085 [5] J Yang, et al., Genetic signatures of high-altitude adaptation
1086 in tibetans. *Proceedings of the National Academy of Sciences of
1087 the United States of America* **114**, 4189–4194 (2017).
- 1088 [6] JP Velotta, CE Robertson, RM Schweizer, GB McClelland,
1089 ZA Cheviron, Adaptive shifts in gene regulation underlie
1090 a developmental delay in thermogenesis in High-Altitude
1091 deer mice. *Molecular biology and evolution* **37**, 2309–2321
1092 (2020).
- 1093 [7] F Cicconardi, et al., Genomic signature of shifts in selection
1094 in a subalpine ant and its physiological adaptations (2020).
- 1095 [8] Y Matsuoka, et al., A single domestication for maize shown
1096 by multilocus microsatellite genotyping. *Proceedings of the
1097 National Academy of Sciences of the United States of America*
1098 **99**, 6080–6084 (2002).
- 1099 [9] DR Piperno, AJ Ranere, I Holst, J Iriarte, R Dickau, Starch
1100 grain and phytolith evidence for early ninth millennium
- 1101 BP maize from the central balsas river valley, mexico. *Pro-
1102 ceedings of the National Academy of Sciences* **106**, 5019–5024
1103 (2009).
- 1104 [10] DR Piperno, KV Flannery, The earliest archaeological
1105 maize (*zea mays l.*) from highland mexico: new accelerator
1106 mass spectrometry dates and their implications. *Proceed-
1107 ings of the National Academy of Sciences of the United States of
1108 America* **98**, 2101–2103 (2001).
- 1109 [11] MB Hufford, et al., The genomic signature of crop-wild
1110 introgression in maize. *PLoS genetics* **9**, e1003477 (2013).
- 1111 [12] E Gonzalez-Segovia, et al., Characterization of introgres-
1112 sion from the teosinte *zea mays* ssp. *mexicana* to mexican
1113 highland maize. *PeerJ* **7**, e6815 (2019).
- 1114 [13] N Lauter, C Gustus, A Westerbergh, J Doebley, The inher-
1115 itance and evolution of leaf pigmentation and pubescence
1116 in teosinte. *Genetics* **167**, 1949–1959 (2004).
- 1117 [14] AK Hardacre, HA Eagles, Comparisons among popula-
1118 tions of maize for growth at 13°c. *Crop science* **20**, 780–784
1119 (1980).
- 1120 [15] T Crow, et al., Gene regulatory effects of a large chromoso-
1121 mal inversion in highland maize. *bioRxiv* (2020).
- 1122 [16] JA Romero Navarro, et al., A study of allelic diversity
1123 underlying flowering-time adaptation in maize landraces.
1124 *Nature genetics* **49**, 476 (2017).
- 1125 [17] DJ Gates, D Runcie, GM Janzen, AR Navarro, others,
1126 Single-gene resolution of locally adaptive genetic varia-
1127 tion in mexican maize. *bioRxiv* (2019).
- 1128 [18] JA Aguirre-Liguori, et al., Divergence with gene flow is
1129 driven by local adaptation to temperature and soil phos-
1130 phorus concentration in teosinte subspecies (*zea mays*
1131 *parviglumis* and *zea mays mexicana*). *Molecular ecology*
1132 **24**, 2663 (2019).
- 1133 [19] MA Fustier, et al., Signatures of local adaptation in lowland
1134 and highland teosintes from whole genome sequencing of
1135 pooled samples. *Molecular ecology* (2017).
- 1136 [20] P Krasilnikov, et al., *The Soils of Mexico*. (Springer Science),
1137 (2013).
- 1138 [21] J Stephen Athens, et al., Early prehistoric maize in northern
1139 highland ecuador. *Latin American Antiquity* **27**, 3–21 (2016).
- 1140 [22] A Grobman, et al., Preceramic maize from paredones and
1141 huaca prieta, peru. *Proceedings of the National Academy of
1142 Sciences of the United States of America* **109**, 1755–1759 (2012).
- 1143 [23] S Takuno, et al., Independent Molecular Basis of Con-
1144 vergent Highland Adaptation in Maize. *Genetics, genetics*.
1145 115.178327 (2015).
- 1146 [24] L Wang, et al., Molecular parallelism underlies convergent
1147 highland adaptation of maize landraces. *bioRxiv* (2020).
- 1148 [25] RR da Fonseca, et al., The origin and evolution of maize
1149 in the southwestern united states. *Nature plants* **1**, 14003
1150 (2015).
- 1151 [26] K Swarts, et al., Genomic estimation of complex traits re-
1152 veals ancient maize adaptation to temperate north america.
1153 *Science* **357**, 512–515 (2017).
- 1154 [27] HY Hung, et al., ZmCCT and the genetic basis of
1155 day-length adaptation underlying the postdomestication
1156 spread of maize (2012).
- 1157 [28] Z Liang, et al., Conventional and hyperspectral time-series
1158 imaging of maize lines widely used in field trials. *Giga-
1159 Science* **7**, 1–11 (2018).
- 1160 [29] L Guo, et al., Stepwise cis-regulatory changes in ZCN8
1161 contribute to maize Flowering-Time adaptation. *Current
1162 biology: CB* **28**, 3005–3015.e4 (2018).
- 1163

- [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 ZmCCT 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] L Wang, et al., The interplay of demography and selection during maize domestication and expansion. *Genome biology* **18**, 215 (2017).
- [35] JT Brandenburg, et al., Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS genetics* **13**, e1006666 (2017).
- [36] 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).
- [37] 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).
- [38] 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).
- [39] H Lambers, et al., Proteaceae from severely phosphorus-impoverished 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).
- [40] Y Gu, et al., Biochemical and transcriptional regulation of membrane lipid metabolism in maize leaves under low temperature. *Front. Plant Sci.* **8**, 2053 (2017).
- [41] RP Poincelot, Lipid and fatty acid composition of chloroplast envelope membranes from species with differing net photosynthesis. *Plant Physiol.* **58**, 595–598 (1976).
- [42] JC Hawke, MG Rumsby, RM Leech, Lipid biosynthesis in green leaves of developing maize. *Plant Physiol.* **53**, 555–561 (1974).
- [43] Y Nakamura, et al., Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nature communications* **5**, 3553 (2014).
- [44] 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).
- [45] JE Pool, DT Braun, JB Lack, Parallel evolution of cold tolerance within *Drosophila melanogaster*. *Molecular biology and evolution* **34**, 349–360 (2017).
- [46] 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.
- [47] 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).
- [48] 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).
- [49] SB Ryu, Phospholipid-derived signaling mediated by phospholipase a in plants. *Trends in plant science* **9**, 229–235 (2004).
- [50] O Emanuelsson, H Nielsen, G von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978–984 (1999).
- [51] SC Potter, et al., HHMMER web server: 2018 update (2018).
- [52] S El-Gebali, et al., The pfam protein families database in 2019. *Nucleic acids research* **47**, D427–D432 (2019).
- [53] E Cimen, SE Jensen, ES Buckler, Building a tRNA thermometer to estimate microbial adaptation to temperature. *Nucleic Acids Research* (2020).
- [54] KAG Kremling, et al., Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).
- [55] R Bukowski, et al., Construction of the third generation zea mays haplotype map. *GigaScience* (2017).
- [56] DE Runcie, L Crawford, Fast and flexible linear mixed models for genome-wide genetics. *PLOS Genetics* **15**, 1–24 (2019).
- [57] Y Nakamura, Plant phospholipid diversity: Emerging functions in metabolism and Protein–Lipid interactions. *Trends in plant science* **22**, 1027–1040 (2017).
- [58] EJ Veneklaas, et al., Opportunities for improving phosphorus-use efficiency in crop plants: Tansley review. *The New phytologist* **195**, 306–320 (2012).
- [59] A Cruz-Ramirez, et al., The xipotl mutant of *Arabidopsis* reveals a critical role for phospholipid metabolism in root system development and epidermal cell integrity. *The Plant cell* **16**, 2020–2034 (2004).
- [60] SR Marla, et al., Comparative transcriptome and lipidome analyses reveal molecular chilling responses in Chilling-Tolerant sorghums. *Plant Genome* **10** (2017).
- [61] L Yan, et al., Parallels between natural selection in the cold-adapted crop-wild relative *Tripsacum dactyloides* and artificial selection in temperate adapted maize. *The Plant journal: for cell and molecular biology* (2019).
- [62] M Kisko, et al., LPCAT1 controls phosphate homeostasis in a zinc-dependent manner. *eLife* **7** (2018).
- [63] C Riedelsheimer, et al., Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics* **44**, 217–220 (2012).
- [64] C Chen, et al., PICARA, an analytical pipeline providing probabilistic inference about a priori candidates genes underlying genome-wide association QTL in plants. *PloS one* **7**, e46596 (2012).
- [65] SC Stelpflug, et al., An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. *The plant genome* **9**, 0 (2016).
- [66] AJ Waters, et al., Natural variation for gene expression responses to abiotic stress in maize. *The Plant Journal* **89**, 706–717 (2017).

- [67] AK Hottes, et al., Bacterial adaptation through loss of function. *PLoS genetics* **9**, e1003617 (2013).
- [68] FI Khan, et al., The lid domain in lipases: Structural and functional determinant of enzymatic properties. *Frontiers in bioengineering and biotechnology* **5**, 16 (2017).
- [69] S Khan, SC Rowe, FG Harmon, Coordination of the maize transcriptome by a conserved circadian clock. *BMC Plant Biol.* **10**, 126 (2010).
- [70] S Maatta, et al., Levels of arabidopsis thaliana leaf phosphatidic acids, phosphatidylserines, and most Trienoate-Containing polar lipid molecular species increase during the dark period of the diurnal cycle. *Front. Plant Sci.* **3**, 49 (2012).
- [71] CM Lazakis, V Coneva, J Colasanti, ZCN8 encodes a potential orthologue of arabidopsis FT florigen that integrates both endogenous and photoperiod flowering signals in maize. *J. Exp. Bot.* **62**, 4833–4842 (2011).
- [72] T Guo, et al., Optimal designs for genomic selection in hybrid crops. *Molecular plant* **12**, 390–401 (2019).
- [73] T Pyhäjärvi, MB Hufford, S Mezmouk, J Ross-Ibarra, Complex Patterns of Local Adaptation in Teosinte. *Genome Biology and Evolution* **5**, 1594–1609 (2013).
- [74] V Matyash, G Liebisch, TV Kurzchalia, A Shevchenko, D Schwudke, Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of lipid research* **49**, 1137–1146 (2008).
- [75] H Tsugawa, et al., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* **12**, 523–526 (2015).
- [76] 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.
- [77] 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.
- [78] M Kanehisa, Y Sato, M Furumichi, K Morishima, M Tamabe, New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**, D590–D595 (2019) Publisher: Oxford Academic.
- [79] JL Portwood, et al., MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research* **47**, D1146–D1154 (2019) Publisher: Oxford Academic.
- [80] JR Walsh, et al., The quality of metabolic pathway resources depends on initial enzymatic function assignments: a case for maize. *BMC Systems Biology* **10**, 129 (2016).
- [81] FR Zapata, sawers-rellan-labs/pglipid: Glycerophospholipid Pathway analysis (2020).
- [82] 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.
- [83] MC Whitlock, Evolutionary inference from qst. *Molecular ecology* **17**, 1885–1896 (2008).
- [84] D Bates, M Maechler, B Bolker, S Walker, lme4: Linear mixed-effects models using eigen and s4. r package version 1.1-7 (2014).
- [85] 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).
- [86] N Duforet-Frebourg, 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).
- [87] 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).
- [88] SN Char, et al., An agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* **15**, 257–268 (2017).
- [89] VA Brazelton, Jr, et al., A quick guide to CRISPR sgRNA design tools. *GM Crops Food* **6**, 266–276 (2015).
- [90] BT Townsley, MF Covington, Y Ichihashi, K Zumstein, NR Sinha, Brad-seq: Breath adapter directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mrna library construction. *Frontiers in plant science* **6** (2015).
- [91] AM Bolger, M Lohse, B Usadel, Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- [92] NL Bray, H Pimentel, P Melsted, L Pachter, Near-optimal probabilistic rna-seq quantification. *Nature biotechnology* **34**, 525 (2016).
- [93] MD Robinson, DJ McCarthy, GK Smyth, edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- [94] SR Eddy, Accelerated profile HMM searches. *PLoS computational biology* **7**, e1002195 (2011).
- [95] Z Li, J Tang, F Guo, Identification of 14-3-3 proteins Phosphopeptide-Binding specificity using an Affinity-Based computational approach. *PloS one* **11**, e0147467 (2016).
- [96] B Pfeifer, U Wittelsbürger, SE Ramos-Onsins, MJ Lercher, PopGenome: an efficient swiss army knife for population genomic analyses in R. *Molecular biology and evolution* **31**, 1929–1936 (2014).
- [97] E Paradis, pegas: an R package for population genetics with an integrated modular approach. *Bioinformatics* **26**, 419–420 (2010).
- [98] J Ho, T Tumkaya, S Aryal, H Choi, A Claridge-Chang, Moving beyond P values: data analysis with estimation graphics. *Nat. Methods* **16**, 565–566 (2019).
- [99] IR Baxter, et al., Single-kernel ionomic profiles are highly heritable indicators of genetic and environmental influences on elemental accumulation in maize grain (*zea mays*). *PloS one* **9**, e87628 (2014).
- [100] F Rodríguez-Zapata, R Sawers, R Rellán-Álvarez, SoilP: A package to extract phosphorus retention data from soil maps and study local adaptation to phosphorus availability (2018).

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.

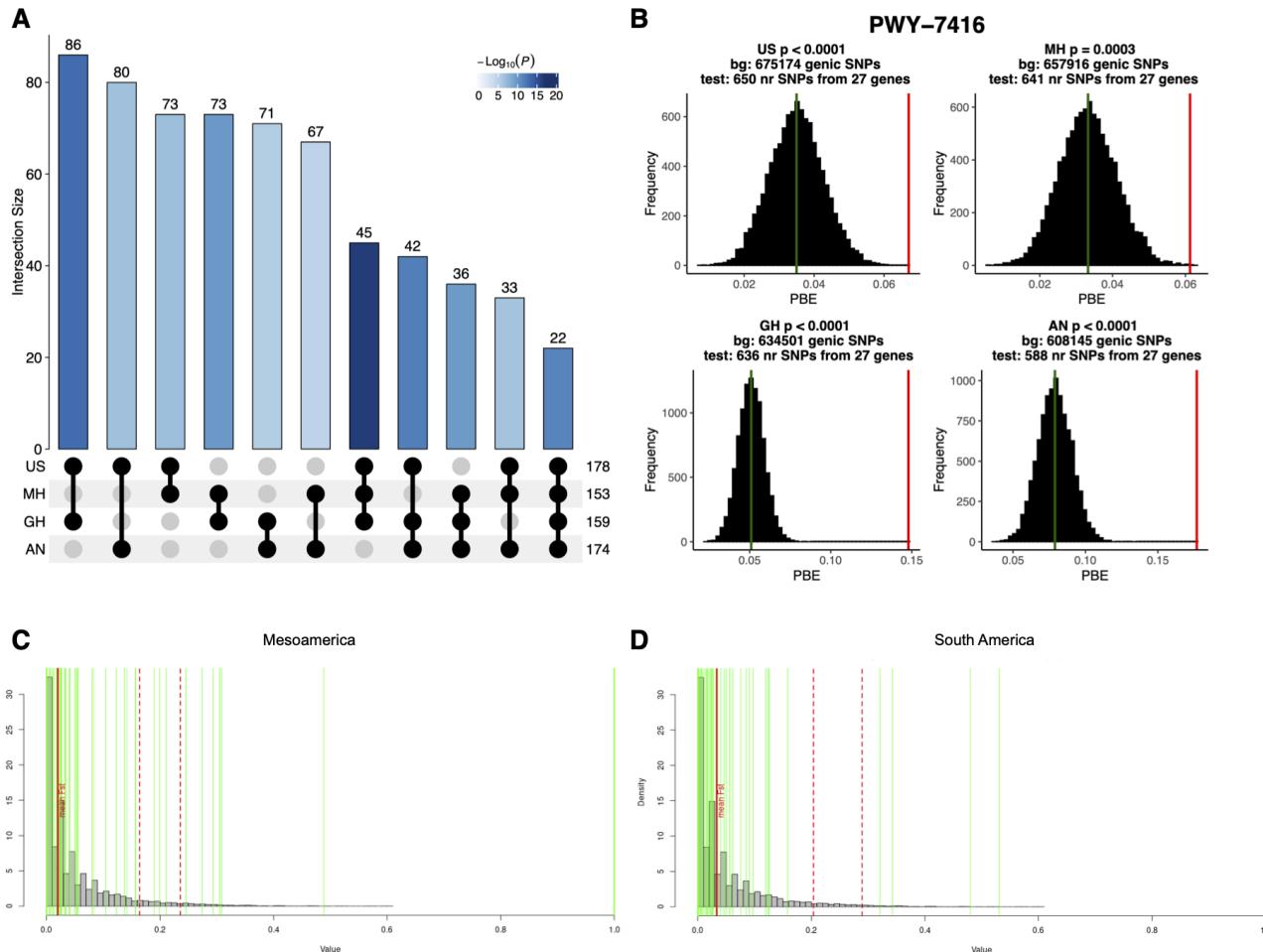


Figure S1 A) Population Branch Excess Analysis of glycerolipid pathway genes. A)From the initial set of 211 genes we used 186 genes with 6219 non redundant SNPs and 683162 non redundant SNPs for the SW US group; 186 genes with 6106 non redundant SNPs and 664555 non redundant SNPs for the Mexican Highland 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. B) Example of pathway 7416 (phospholipid remodeling) PBE values (red lines) for each highland population and genome wide PBE distributions (black histograms). $Q_{ST}-F_{ST}$ analysis of glycerolipid compounds between highland and lowland landraces from Mesoamerica (C) and South America (D). Samples for Dart-Seq genotyping were taken from the same plants that were grown in the common garden experiment shown in Figure 1D-F and that were used for glycerolipid analysis.

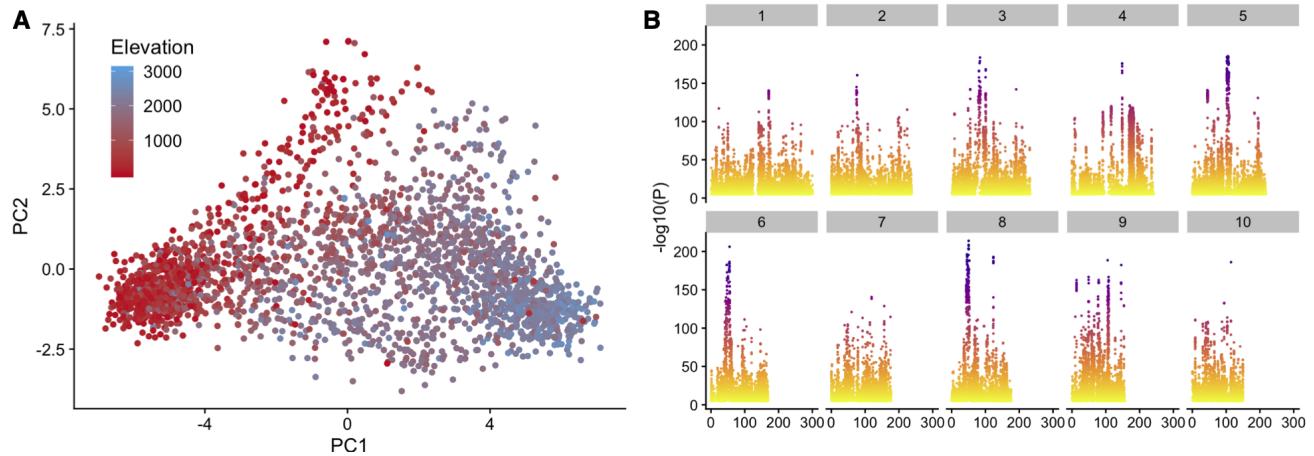


Figure S2 *Pcadapt* analysis of Mexican landraces. We used GBS data from the Mexican landraces of the SEEDs dataset (16) and run a *pcaadapt* analysis (48) that identified (A) elevation as the major driver of population differentiation polarizing PC1. B) Genome wide analysis of *Pcadapt* PC1 outliers.

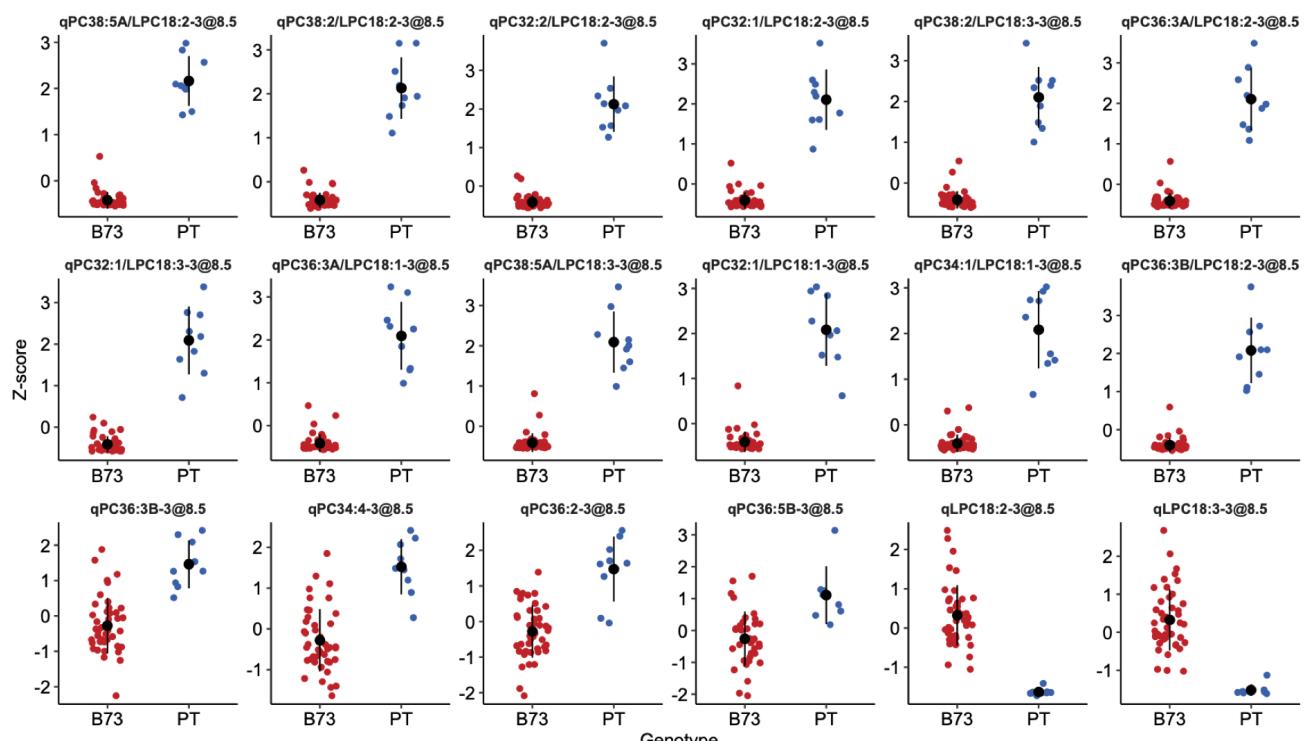


Figure S3 A) Effect sizes of several individual PC, LPC, and PC/LPC QTL peaks.

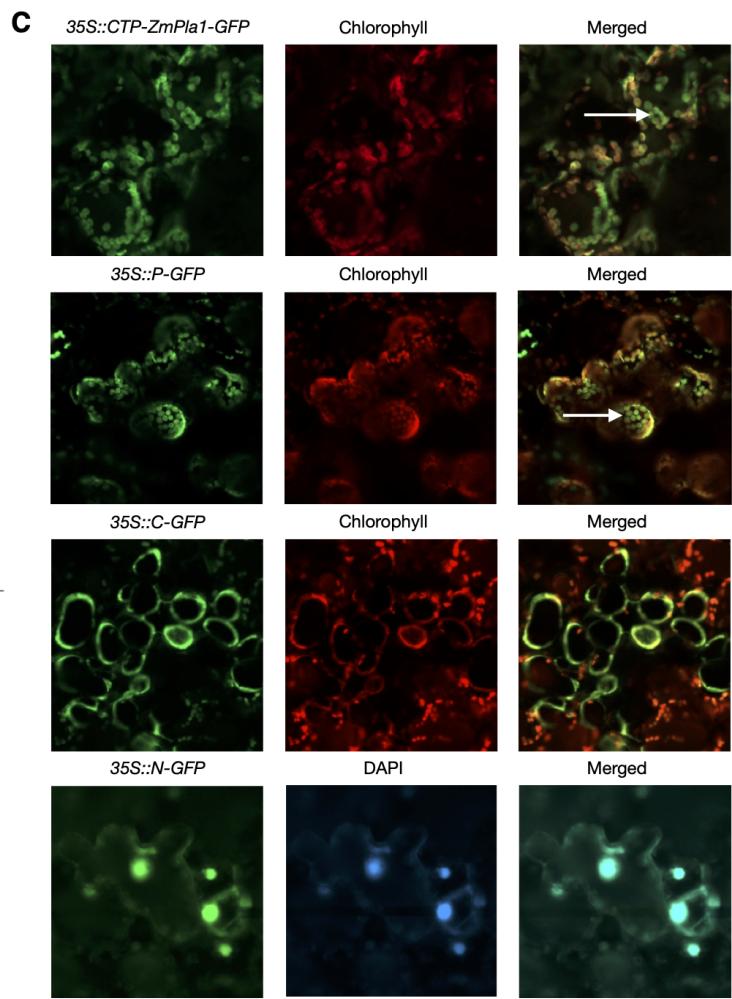
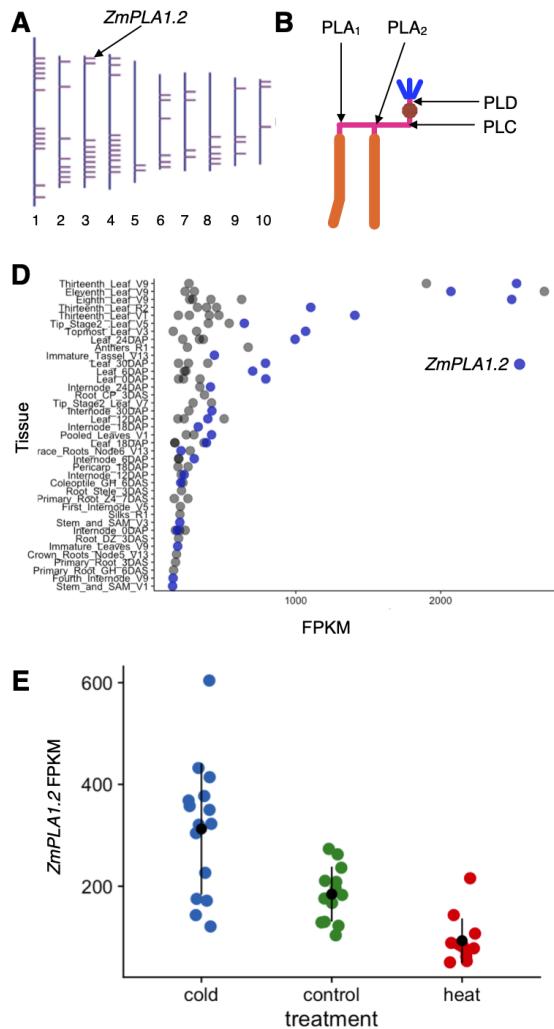


Figure S4 A) Genomic Location of genes coding for enzymes predicted Phospholipase A1 activity. B) Site of action of the different types of phospholipases. C) Subcellular localization of *ZmPLA1.2*. D) B73 expression levels of genes coding for enzymes with predicted Phospholipase A1 activity across different 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 (66).

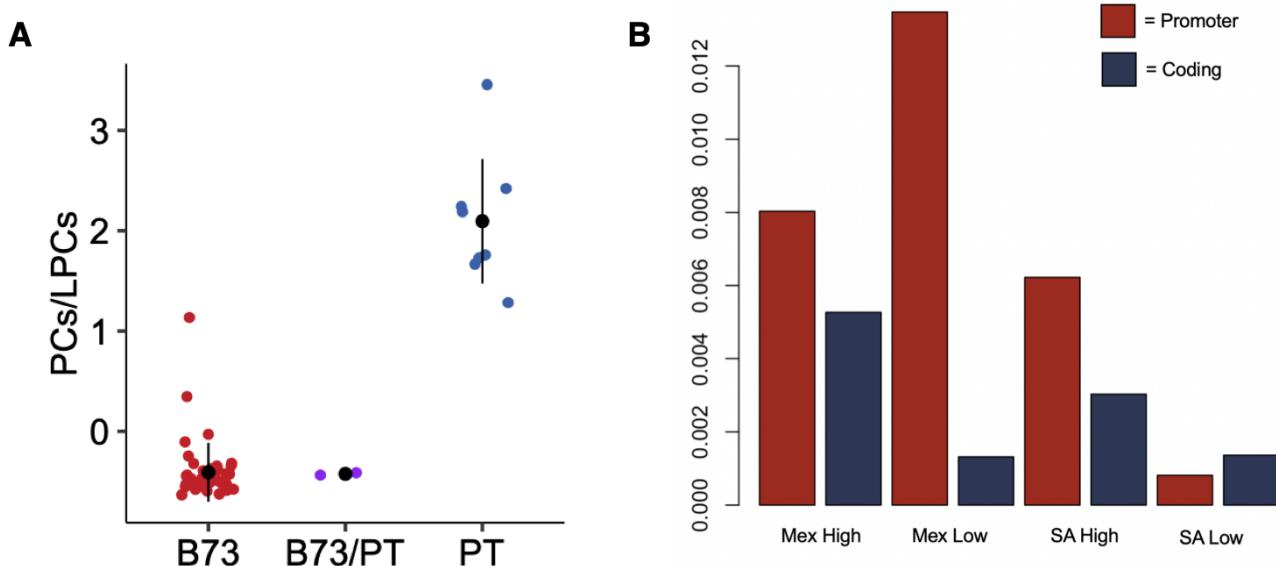


Figure S5 A)Effect sizes of PC/LPC levels at BILs homozygous B73, PT and heterozygous qPC/LPC3@8.5. B) Nucleotid diversity analysis of the promoter and CDS region of *ZmPLA1.2* using whole genome sequencing data of highland and lowland landraces México and South American obtained from (34)

Figure S6 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.

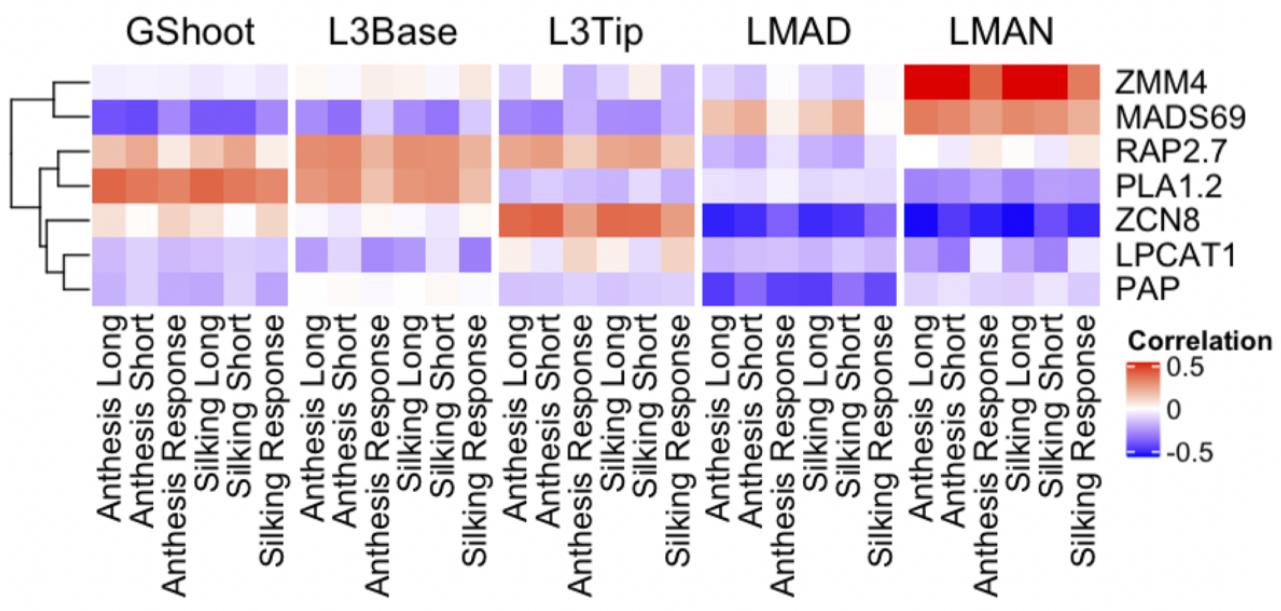
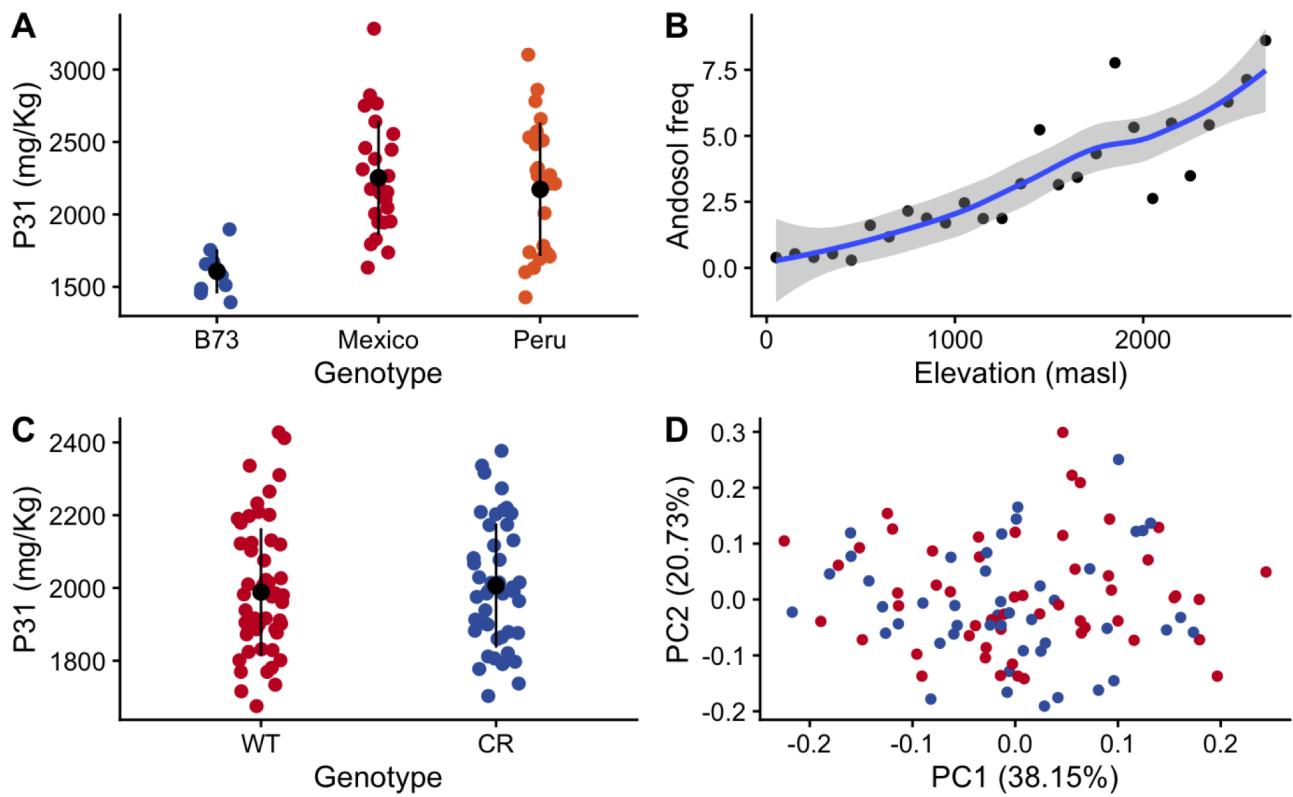


Figure S7 Flowering time and phospholipid related gene expression correlation with flowering time traits in aerial tissues in the 282 panel. Data obtained from (54)



Supplementary Figure 8

Figure S8 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 (100). 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