bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time

Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time

Fausto Rodríguez-Zapata[a,b,1], Allison C Barnes[a,1], Karla A Blöcher-Juárez[b,1], Dan Gates[c], Andi Kur[a], Li Wang[d], Garrett M Janzen[d], Sarah Jensen[e], Juan M Estévez-Palmas[b], Taylor Crow[f], Rocío Aguilar-Rangel[b], Edgar Demesa-Arevalo[g], Tara Skopelitis[g], Sergio Pérez-Limón[b], Whitney L Stutts[a, h], Peter Thompson[a, h], Yu-Chun Chiu[h], David Jackson[g], Oliver Fiehn[i], Daniel Runcie[f], Edward S Buckler[e], Jeffrey Ross-Ibarra[c], Matthew B. Hufford[d], Ruairidh JH Sawers[b,j] and Rubén Rellán-Álvarez[a, b, *]

[a]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, [b]National Laboratory of Genomics for Biodiversity, Irapuato, México, [c]Department of Evolution and Ecology, Center for Population Biology and Genome Center, University of California, Davis, CA, [e]US Department of Agriculture–Agricultural Research Service, Cornell University, Ithaca, NY, [f]Department of Plant Sciences, University of California, Davis, CA, [d]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, USA, [g]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, [h]Molecular Education, Technology and Research Innovation Center, North Carolina State University, Raleigh, NC, [i]West Coast Metabolomics Center, University of California, Davis, CA, USA, [j]Department of Plant Science, The Pennsylvania State University, PA, USA

Abstract

After domestication from lowland teosinte *parviglumis* (*Zea mays* ssp.*parviglumis*) in the warm Mexican southwest, maize (*Zea mays* ssp. *mays*) colonized the highlands of México and South America. In the highlands, maize was exposed to lower temperatures that imposed strong selection on flowering time. Phospholipids are important metabolites in plant responses to low-temperature, low phosphorus availability and have also been suggested to influence flowering time.

Here, we combined linkage mapping analysis with genome scans to identify *High PhosphatidylCholine 1* (*HPC1*), a gene which encodes a phospholipase A1 enzyme, as a major driver of phospholipid variation in highland maize. Common garden experiments demonstrated strong genotype-by-environment interactions associated with variation at *HPC1*, with the highland *HPC1* allele leading to higher fitness in highlands, possibly by hastening flowering.

The highland maize *HPC1* variant results in impaired function of the encoded protein due to a polymorphism in a highly conserved sequence. A meta-analysis indicated a strong association between the identity of the amino acid at this position and optimal growth in prokaryotes. Mutagenesis of *HPC1* via genome editing validated its role in regulating phospholipid metabolism.

Finally, we showed that the highland *HPC1* allele entered cultivated maize by introgression from the wild highland teosinte *Zea mays* ssp. *mexicana* and has been maintained in maize breeding lines from Northern US, Canada and Europe.

Thus, *HPC1* introgressed from teosinte *mexicana* underlies a large metabolic QTL that modulates phosphatidylcholine levels and has an adaptive effect at least in part via induction of early flowering time..

phospholipid metabolism; maize genetics; highland adaptation; convergent evolution

Introduction

Elevation gradients are associated with changes in environmental factors that impose significant physiological constraints on an organism. Adaptation to highland environments is achieved via the selection of genetic variants that improve their ability to withstand reduced oxygen availability (1; 2; 3; 4), increased UV radiation (5) and lower temperatures (6; 7). In particular, low temperatures can significantly reduce the thermal time accumulation to flowering (measured in growing degree days (GDDs) (8) and select for accelerated development and maturity as a compensatory mechanism (9).

Following domestication from teosinte *parviglumis* (*Zea mays* ssp. *parviglumis*) (10; 11) in the lowland, subtropical environment of the Balsas River basin (Guerrero, México), cultivated maize (*Zea mays* ssp. *mays*) expanded throughout México and reached the highland valleys of central México around 6,500 years before present (12).

In México, highland adaptation of maize was aided by substantial adaptive introgression from a second subspecies of teosinte, teosinte *mexicana* (*Zea mays* ssp. *mexicana*) that had already adapted to the highlands of México thousands of years after its divergence from teosinte *parviglumis* (13; 14). Adaptation to low temperature and soils with low phosphorus in highland

[1]These authors contributed equally to this work.

[*]Department of Molecular and Structural Biochemistry, North Carolina State University, 27607, Raleigh, NC. Email: rrellan@ncsu.edu

environments drove *parviglumis* and *mexicana* genetic divergence (15). Phenotypically, the most evident signs of *mexicana* introgression into maize are the high levels of stem pigmentation and pubescence (16) that are thought to protect against high UV radiation and low temperatures. The ability to withstand low temperatures and efficiently photosynthesize during the early stages of seedling development are key factors in maize highland adaptation (17). Indeed, recent transcriptome deep sequencing (RNA-seq) analysis showed that the inversion *Inv4m*, introgressed from *mexicana*, strongly affects the expression of genes involved in chloroplast physiology and photosynthesis (18). *Inv4m* is also associated with earlier flowering times in highland maize (19; 20). Given the low accumulation of growing degree days in typical highland environments, selection has favored shorter generation times in highland adapted maize (20). By the time maize reached the Mexican highlands, its range had already expanded far to the South, including colonization of highland environments in the Andes (21; 22). Andean maize adaptation occurred without *mexicana* introgression, as there is no known wild teosinte relative in South America. These multiple events of maize adaptation to highland environments make maize a good system to study the evolutionary and physiological mechanisms of convergent adaptation (23; 24).

In comparison to its southward expansion, the northward migration of maize into modern-day United States, where summer daylengths are longer, occurred at a much slower pace (25; 26) due to delayed flowering of photoperiod-sensitive tropical maize lines (27). Indeed, a host of evidence suggests that maize cultivation in northern latitudes was enabled by the selection of allelic variants that led to a reduction in photoperiod sensitivity to allow flowering under longer photoperiods (28; 29; 30; 31; 32; 33; 27). Some of the early flowering alleles that conferred an adaptive advantage in highland environments are the result of *mexicana* introgressions into highland maize (29) that were further selected at northern latitudes. Photoperiod-insensitive maize from the Northern US and Canada thrived when introduced into Northern Europe, as it was already pre-adapted to northern latitudes and lower temperatures (34). Introgression of highland-adapted teosinte *mexicana* into highland maize can be a critical source of adaptive genetic variation for modern maize adapted to high latitudes. However, genetic, physiological and phenotypic evidence of such events is limited.

Plant phospholipids, as well as other glycerolipids such as sulfolipids, galactolipids and less polar lipids such as triacylglycerols, are involved in plant responses to low temperature. Phospholipid levels are increased in plants exposed to low temperatures (35) and levels of unsaturated fatty acids in glycerolipids are reduced (36; 37), which may help maintain the fluidity of cell membranes. Under stressful conditions, the proportions of differently shaped lipids in plants are modulated to maintain membrane flexibility while preventing membrane leakage. For instance, phosphatidylcholines (PCs) are rectangular polar lipids that are well suited for the formation of bilayer membranes because the size of their glycerol backbone, choline headgroup and fatty acid tails are similar. By contrast, lyso-phosphatidylcholine (LPC) is a triangular PC with a single acyl group that cannot form a bilayer because its headgroup is much larger than its fatty acid (38). LPCs do allow for some membrane movement, but high LPC concentrations act as a detergent (39) and can facilitate cell leakage and damage at low temperatures, effects that would be prevented by higher PC levels. In cold-adapted maize temperate lines and *Tripsacum* species (a distant maize relative), genes involved in phospholipid biosynthesis show accelerated rates of protein sequence evolution, further supporting an important role for phospholipid metabolism across several species during cold adaptation (40). Finally, certain species of PC, the most abundant phospholipid in plant cells (41), can bind to *Arabidopsis thaliana* FLOWERING LOCUS T (FT) and accelerate flowering through an unknown mechanism (42). Consistent with this observation, glycerolipid levels in maize have predictive power for flowering time (43).

Here, we identified an adaptive teosinte *mexicana* introgression that alters highland maize phospholipid metabolism and leads to early flowering. Using genome scans and linkage mapping, we identified *High PhosphatidylCholine1* (*HPC1*), encoding a phospholipase A1, as a driver of high PC levels in highland maize. Data from thousands of genotyped landrace test-crosses grown in common garden experiments at different elevations in México showed a strong genotype-by-environment effect at the *HPC1* locus, where the highland allele leads to higher fitness in the highlands and reduced fitness at lower elevations. Furthermore, we determined that the highland *HPC1* allele that was introgressed from teosinte *mexicana* was carried northward and is now present in maize cultivars grown in the Northern US and European Flint lines. These results suggest that the *HPC1* highland allele has a beneficial effect in cold, high-latitude environments, where early flowering is advantageous.

Results

Highland Mesoamerican landraces show high PC/LPC ratios and selection of highly unsaturated PCs

We grew a diversity panel composed of 120 highland and lowland landraces from Mesoamerica and South America (Figure 1A, Sup. File 1), hereafter referred to as the HiLo diversity panel, in highland and lowland Mexican common gardens and quantified glycerolipid levels. We observed that Mesoamerican highland landraces have high PC/LPC ratios, particularly when grown in the highlands (Figure 1B). The differences observed in phospholipid levels between highland and lowland maize may be the result of adaptive natural selection or random genetic drift during the process of maize colonization of highland environments. To distinguish between these two competing possibilities, we compared the variance of each phenotype across the population with the genetic variance of neutral markers, using a Q_{ST}-F_{ST} comparison (44). We calculated Q_{ST}-F_{ST} using DartSeq genotypic data from the same plants that were used to analyze glycerolipid levels, and calculated the Q_{ST}-F_{ST} values for each glycerolipid species for highland/lowland populations from each continent. Mean Q_{ST} was greater than mean F_{ST} in Mesoamerican and South American comparisons, although only the Mesoamerican comparison was significant (two-tailed *t*-test, $p = 0.00073$; South American comparison $p = 0.12$). However, we noticed several PC and LPC species with higher Q_{ST} values than the neutral F_{ST} in both continents (Figure 1C-D). The species with the highest Q_{ST} value in both continents was PC-36:5. In summary, highland Mesoamerican landraces grown in highland conditions had particularly high PC/LPC ratios and levels of PC species, especially of species with high levels of unsaturation (e.g., PC-36:5).

Genes involved in PC/LPC conversion show strong highland selection signals

We used genotyping by sequencing (GBS) data from 2,700 Mexican maize landraces, generated by the SeeD project (19; 20), to run a principal component-based (*pcadapt*) analysis to determine

how loci might contribute to observed patterns of differentiation between major principal components of genetic variation (45). The first principal component of *pcadapt* separated Mexican landraces based on the elevation of their geographical origin (Supplementary Figure 1A). Using this first principal component, we identified outlier SNPs across the genome that are significantly associated with landrace elevation and potentially involved in local adaptation (Supplementary Figure 1B). We independently catalogued ≈ 600 maize genes known to be involved in glycerolipid metabolism (See Materials and Methods for details). From those, 85 contained single nucleotide polymorphisms (SNPs) that were *pcadapt* PC1 outliers (top 5% -log(P)) (Sup. File 2).

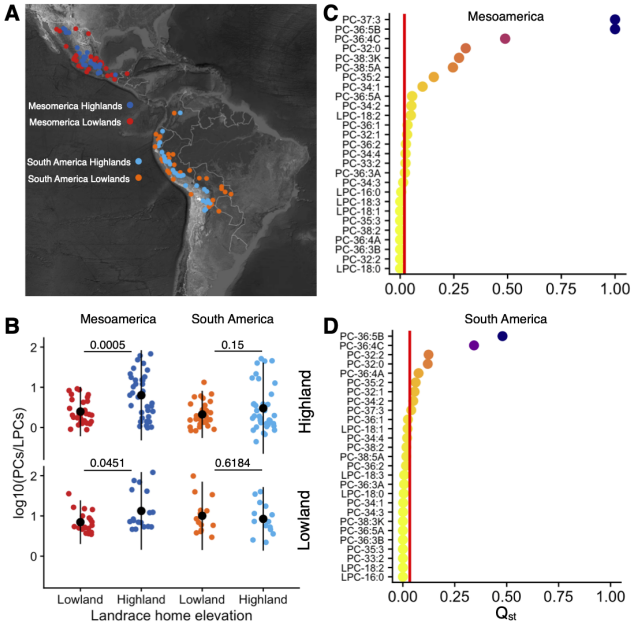

Figure 1 Phospholipid selection in highland maize. A) Map showing the geographical origin of the 120 maize accessions from the HiLo diversity panel used in the common garden experiment to quantify glycerolipid levels. **B)** PC/LPC ratio, in $\log_{10}$ scale, of highland and lowland landraces from Mesomerica and South America; adjusted pairwise comparison *t*-test *p*-values are shown. Q_{ST}-F_{ST} analysis of phospholipid compounds between highland and lowland landraces from Mesoamerica **C)** and South America **D)**.

The genes *High PhosphatidylCholine 1* (*HPC1*, *Zm00001d039542*) and *Lyso-Phosphatidylcholine Acyl Transferase 1* (*ZmLPCAT1*, *Zm00001d017584*), which are involved in phospholipid remodeling, had SNPs within their coding sequence with significantly high -log(P) values (Figure 2A-B), indicating strong elevation-dependent allele frequency changes (Figure 2C-D). We then used the same set of 600 glycerolipid metabolism-related genes to identify selection signals using the Population Branch Excess (PBE) (46) statistic across four highland populations: Southwestern US (SWUS), Mexican Highlands (MH), Guatemalan Highlands (GH) and Andes (24). We established that, of the 153 glycerolipid-related genes that were PBE outliers in the Mexican Highlands, 38 were also *pcadapt* PC1 outliers (Sup. File 2). There are two possible explanations for the extent of convergent selection in highland populations (see (24; 47)). Adaptation may be conferred by a small number of genes, thereby imposing a *physiological* constraint on the sources of adaptation leading to convergence. Alternatively, adaptation may be the result of many genes, but deleterious *pleiotropic* effects restrict the number of genes that can be targeted by selection, also leading to convergence. Using Yeaman's *et al.* C_{hyper} statistic (47), which quantifies these two modes of convergent adaptation, we determined that the overlap among putative adaptive genes in the four highland populations cannot be explained merely by physiological constraints ($C_{hyper} = 3.96$). A certain degree of pleiotropic constraint is therefore likely. Overlap between adaptation candidates was higher for the SWUS MH and GH population pairs ($C_{hyper} = 4.79$) than between the Andean and SWUS MH and GH pairs ($C_{hyper} = 3.14$). We identified a significant excess of genes that are targets of selection in more than two populations ($P < 3 \times 10^{-5}$, Supplementary Figure 1C). The most over-represented intersection of selected glycerolipid genes was in the SWUS, MH and GH populations ($p = 1 \times 10^{-15}$, Supplementary Figure 1C), perhaps indicating a set of genes specifically selected in this geographical region relative to the Andean material and/or closer kinship between those populations and thus weaker statistical independence. From the glycerolipid pathways, 22 genes were PBE outliers consistently across all four populations ($p =< 1\times^{-10}$, Supplementary Figure 1C). We then performed an independent analysis for each of the 30 glycerolipid metabolism pathways and compared their average PBE value (using a 10-kb window around each constituent gene) with a genome-wide genic random sampling distribution of PBE values. The pathways related to 'phospholipid remodeling' and 'PC acyl editing' had significantly higher PBE values in all four populations, indicating a possible role for phospholipid remodeling in maize highland adaptation (Supplementary Figure 1D and Sup. File 2). For example, *ZmLPCAT1* is a selected gene in the 'phospholipid remodelling' and 'PC acyl editing' pathways with shared outlier SNPs in the coding sequence region across all four populations (Figure 2E, Sup. File 2). HPC1 performs the reverse reaction of ZmLPCAT1 and is part of the 'PC acyl editing', 'triacylglycerol degradation' and 'phospholipases' pathways. *HPC1* displayed particularly high PBE values in the Mexican Highland population and contained SNPs that are unique to each population and others that are shared across populations (Figure 2E). Together, we showed here using two independent datasets that pathways involved in phospholipid remodelling and in particular two genes (*HPC1* and *ZmLPCAT1*) controlling the PC/LPC ratio show strong selection signals in highland maize.

A major QTL explaining PC to LPC conversion overlaps with HPC1

To break population structure and identify loci involved in phospholipid biosynthesis in highland maize, we developed a recombinant inbred line (RIL) BC1S5 population derived from a cross between the temperate inbred line B73 and the Mexican highland landrace Palomero Toluqueño (PT), using B73 as the recurrent parent (75% B73, 25% PT).

The parental PT accession is a popcorn landrace (palomero means popcorn in Spanish) from the Toluca Valley in México (*Mexi5* CIMMYTMA 2233) (Figure 3A). We grew the Hilo diversity panel and the B73 x PT BC1S5 mapping population in the same highland and lowland common gardens and collected samples for glycerolipid analysis. The locally adapted PT landrace showed higher fitness than B73 in the highland field (Figure 3A), probably due to adaptation to low temperatures in this highland environment. In the Mexican highlands, values of around 5 GDD are typical, while 15-20 GDD are common in lowland environments. We detected major quantitative trait loci (QTLs) for the sum of LPC species levels, PC species levels and the

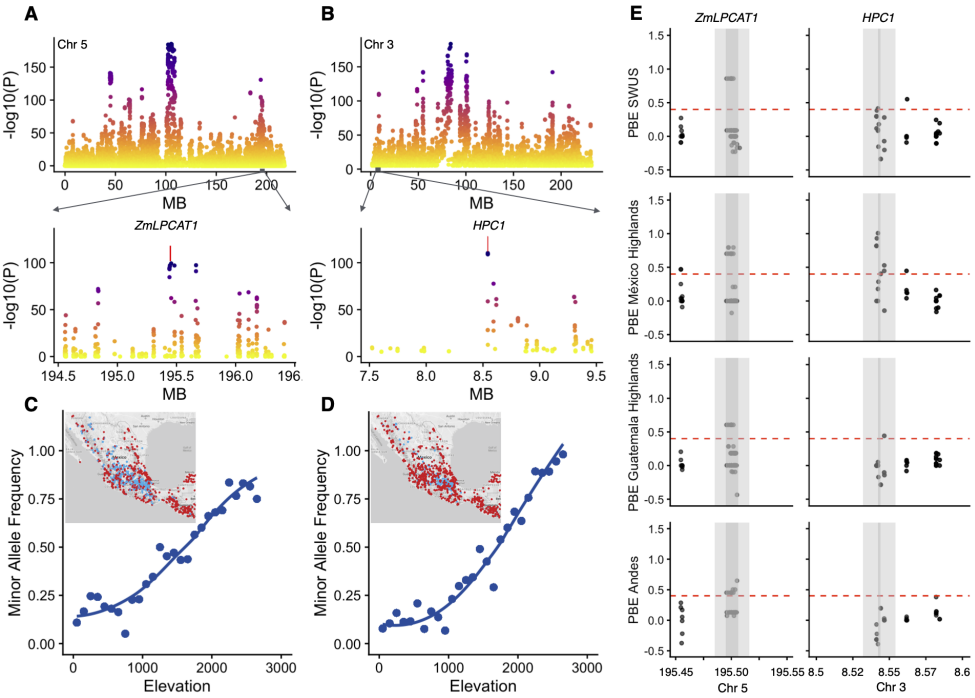

Figure 2 Evidence of highland selection in genes determining PC/LPC ratios. Manhattan plots of minus log10(p-values) *pcadapt* outliers. **A)** and **B)** *pcadapt* PC1 outlier plots of chromosomes 5 and 3, respectively. Lower panels zoom in on areas of outlier SNPs that co-localize with the physical position of the coding sequences (marked with a red vertical line) of *HPC1* and *ZmLPCAT1*. **C)** and **D)** the geographic- and elevation-dependent minor allele frequencies of the highland (blue) and lowland (red) alleles of one of the outlier SNPs in the coding sequence of *HPC1* and *ZmLPCAT1*. **E)** PBE values of SNPs in *ZmLPCAT1* and *HPC1*.

PC/LPC ratio that all mapped to the same locus on chromosome 3, at around 8.5 Mb (Figure 3B). The PC/LPC ratio QTL had the highest logarithm of the odds (LOD) score, i.e., 24.5, and explained the most phenotypic variance (87%). We searched for evidence of epistatic interactions for LPC levels, PC levels, and the PC/LPC ratio through a combination of R/qtl `scantwo` and `stepwise` functions (49), but identified none. The three QTLs *qLPCs3@8.5*, *qPCs3@8.5* and *qPCs/LPCs3@8.5* were robust against environmental effects and were detected in both the highland and lowland environments. The additive effect of the PT allele at these QTLs was associated with high levels of PCs, low levels of LPCs and consequently high PC/LPC ratios, while the B73 allele had the opposite effect (Figure 3C, top panel). Individual QTLs for PC and LPC species at this locus showed the same additive effect for the PT allele as the sum of each class of species (from the QTLs *qPCs@8.5* and *LPCs3@8.5*) (Figure 3C, top panel and Supplementary Figure 2). All individual LPC QTLs at the *qLPCs3* locus corresponded to LPCs that contain at least one double bond in their fatty acid (Supplementary Figure 2, Sup. file 3). *qPCs3@8.5* was driven mainly by PC species with more than two fatty acid double bonds, such as PC 36:5 (Figure 3D and Supplementary Figure 2 bottom panel).

We then sought to identify candidate genes underlying the QTLs on chromosome 3. The QTL contained 72 genes within its 1.5 LOD drop confidence interval (7.9-10 Mb) We hypothesized that the metabolic phenotypes we observed might be due to a gene involved in the process of PC-LPC conversion. The maize genome encodes 75 putative phospholipases (Supplementary Figure 3A), of which half are predicted to be phospholipase A1-type (PLA1) (Supplementary Figure 3A). We identified *HPC1* (position on chromosome 3: 8,542,107 bp to 8,544,078 bp), within the QTL peak, as the most likely candidate (Figure 3B). HPC1 has a predicted phospholipase A1-Igamma1 activity and can be classified, based on its two closest Arabidopsis orthologs (encoded by At1g06800 and At2g30550), as a PC-hydrolyzing PLA1 Class I phospholipase (50). PLA1-type phospholipases hydrolyze phospholipids in the sn-1 position and produce a lyso-phospholipid and a free fatty acid (Supplementary Figure 3B). Class I phospholipases are targeted to the chloroplast; indeed, we identified a chloroplast transit peptide (CTP) at the beginning of the predicted HPC1 sequence using the online tool ChloroP (51). We validated the chloroplast localization of HPC1 by transiently expressing the HPC1 CTP fragment fused to green fluorescent protein (GFP) in *Nicotiana benthamiana* leaves (Supplementary Figure 4). In B73, *HPC1* was one of the most highly expressed phospholipases, with expression almost exclusively restricted to vegetative leaves (V4-V9) (Supplementary Figure 5A) (52), which is the biological material we sampled for glycerolipid analysis. In B73 leaves, *HPC1* was also the most highly expressed gene in the QTL interval (Supplementary Figure 5B (52)).

The effect of *HPC1* on the PC/LPC levels may be caused by mis-regulation of *HPC1* expression in highland landraces and/or by a mutation affecting the enzymatic activity of *HPC1*. To distinguish between these two possibilities, we analyzed *HPC1* expression in B73, PT and the corresponding F_1 hybrid plants grown at high or low temperatures to simulate highland and lowland conditions, respectively (Figure 3E). Under cold conditions, *HPC1-B73* was up-regulated, but *HPC1-PT* was not (Figure 3E). The lack of up-regulation in cold conditions of *HPC1* may explain the high PC/LPC levels in PT. However, the gene is expressed in PT at the same levels as in B73 under control conditions. In the F_1 hybrids, *HPC1* expression was consistent with a dominant B73 effect. We observed the same phenomenon for the PC/LPC ratio in the few B73 x PT BC1S5 RILs that are heterozygous at

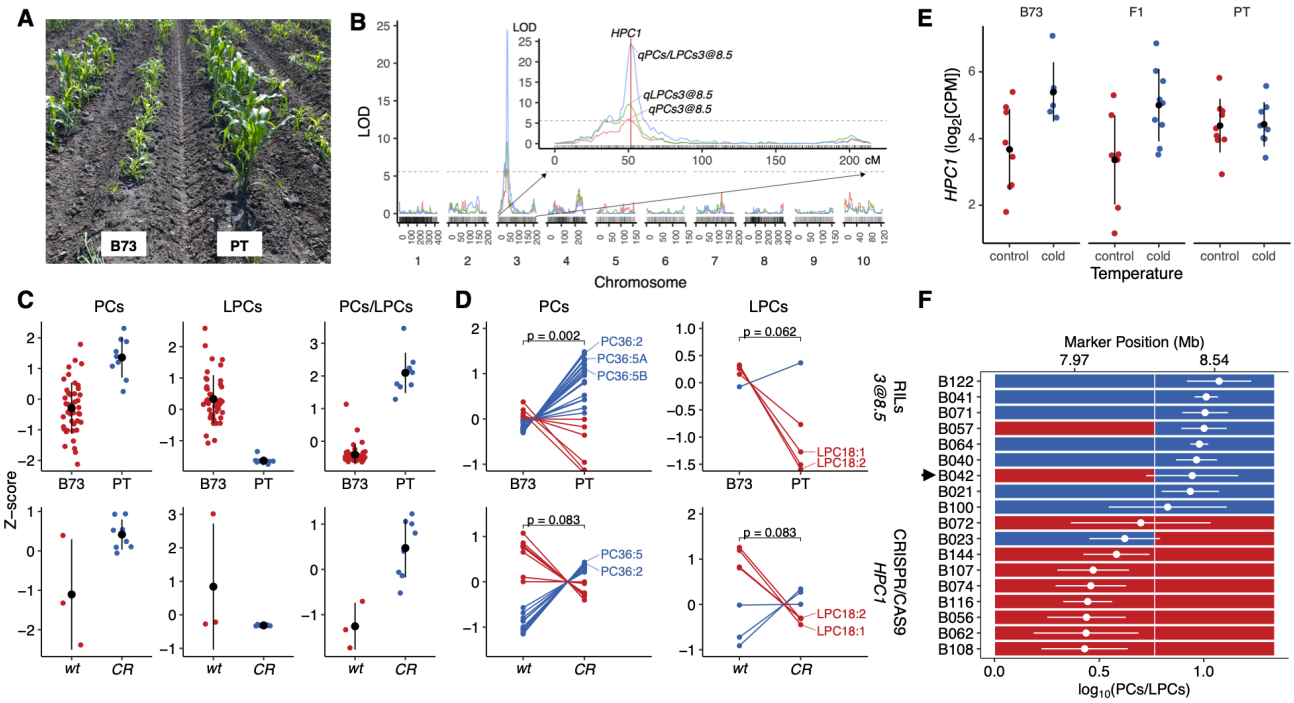

Figure 3 *HPC1* defines a major QTL explaining PC/LPC conversion. A) PT and B73 plants growing in the highland Metepec field. **B)** QTL analysis using data collected from plants growing in the highland and lowland fields for PC and LPC levels and the PC/LPC ratio identified overlapping major QTLs at 8.5 Mb on chromosome 3. The QTL peaks coincide with the physical location of *HPC1*. **C)** Effect sizes for PCs, LPCs and PC/LPCs (z-score normalized) in RILs at 8.5 Mb on chromosome 3 that are either homozygous for B73 or PT (top row) and CRISPR-Cas9 $hpc1^{CR}$ mutant and wild-type plants (bottom row). **D)** Effect sizes for individual PC and LPC species (z-score normalized) in RILs at 8.5 Mb on chromosome 3 (top row) and CRISPR/Cas9 $hpc1^{CR}$ mutant. **E)** *HPC1* expression analysis in B73, PT and their F_1 hybrids grown in standard and cold temperatures in a growth chamber. **F)**PC/LPC ratio for several RILs. RIL B042 (indicated by the black arrow) bears a recombination event 500 bp upstream of the *HPC1* translation start codon. In panels C-F, phenotypes associated with the B73 haplotype are in red; the equivalent values for the PT haplotype are in blue.

the *HPC1* locus (Supplementary Figure 3C). Variation in *HPC1* may also (or in addition to any effects on expression levels) affect enzymatic activity of the HPC1-PT variant. In agreement with this notion, several PBE and *pcadapt* outlier SNPs mapped to the *HPC1* coding sequence rather than to regulatory regions. We used Sanger sequencing to analyze three B73 x PT RILs (B021, B042, B122) that are homozygous for the *HPC1-PT* allele. We discovered a recombination point 500 bp upstream of the *HPC1* translation start codon (Figure 3F, Supplementary Figure 6) in RIL B042, resulting in a chimeric gene with the coding region from PT, and a promoter segment from B73. PC/LPC levels in RIL B042 were similar to other RILs that are homozygous for the PT haplotype at the 8.54 Mb marker at the QTL peak (Figure 3F). This result supports the hypothesis that the metabolic effect we see is likely due to an impaired function of the HPC1-PT enzyme rather than to changes in the *HPC1-PT* regulatory region. If *HPC1* is the underlying causal gene of this QTL, the observed metabolic phenotypes would be consistent with reduced or loss of function of the HPC1-PT enzyme leading to higher levels of PCs and lower levels of LPCs in the PT background. Next, we generated a genome-edited $hpc1^{CR}$ mutant via clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated nuclease Cas9 (Sup. File 4) in the B104 background, a temperate inbred derived from B73. We then measured PC and LPC species in wild-type and mutant plants grown under standard greenhouse conditions. $hpc1^{CR}$ plants recapitulated the effects seen for the *PT* allele in the RILs (Figure 3 C-D bottom panels), further confirming that the *HPC1-PT* allele impairs HPC1 function and thus underlies the QTL on chromosome 3 at around 8.5 Mb.

Together, these data show that *HPC1* underlies a major metabolic QTL explaining PC/LPC ratio. A mutation affecting the enzymatic activity of its encoded protein probably leads to impaired function in the highland PT allele.

HPC1 shows strong elevation-dependent antagonistic pleiotropy in Mexican landrace fitness phenotypes

We re-analyzed phenotypic data from a previously reported F1 association mapping panel (19) (20), by fitting a model to estimate the effect stemming from variation at *HPC1-PT* on the relationship between fitness trait and elevation (48). We determined that variation at *HPC1* significantly affected G x E interactions for several fitness traits (Figure 4). The effect of *HPC1* on flowering time showed antagonistic pleiotropy between highland and lowland environments (Figure 4). The highland *HPC1-PT* allele was associated with delayed flowering in lowland environments (an increase of around one day in days to anthesis (DTA) and 0.25 day for the anthesis to silking interval (ASI)), but accelerated flowering at high elevation (with a decrease of DTA and ASI of one and 0.25 day, respectively) (Figure 4). Variation at *HPC1* showed conditional neutrality on the fresh ear weight and grain weight per hectare traits: the highland allele had no effect in lowland environments, but was associated with greater values

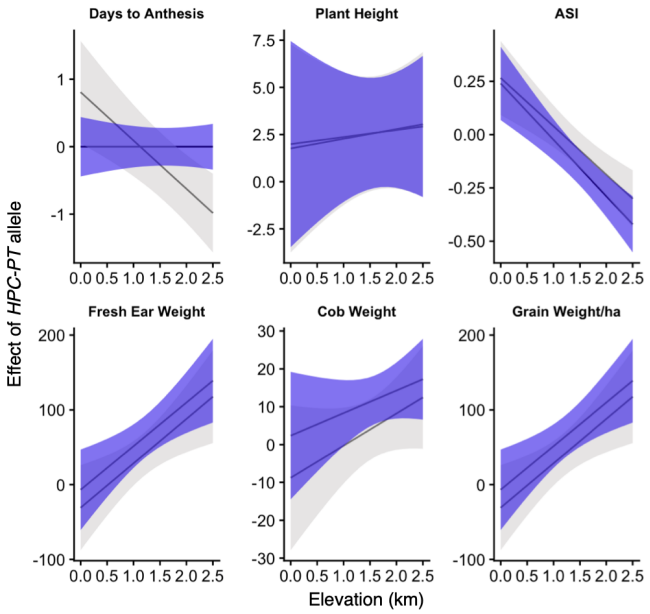

Figure 4 Fitness effects of the *HPC1-PT* allele We used best linear unbiased predictions (BLUPs) and GBS data from 2,700 landraces from (20), evaluated in 23 common gardens at different elevations in México. We modeled each trait as a function of the *HPC1-PT* genotype, trial elevation, and tester line, 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 *HPC1-PT* as a function of common garden elevation ± 2 standard errors of the mean, using the *GridLMM* package (48). Purple lines show estimates of the *HPC1-PT* effect in a model that also includes effects of days to anthesis. ASI, anthesis to silking interval.

in highland environments (Figure 4).

Identification of the putative causal SNP in HPC1.

We sequenced the *HPC1* locus from several RILs harboring the PT haplotype at *HPC1* and identified several of non-synonymous SNPs within the coding sequence that might influence HPC1 function (Supplementary Figure 7) We focused our attention on SNP 631 in the flap lid domain, which leads to a conservative replacement of isoleucine for valine (V211I, Figure 5A). The flap lid domain is important for phospholipase activity and is located in a lipase class 3 domain (PFAM domain PF01764) that is highly conserved across the tree of life. We recovered 982 observations of the lipase class 3 PFAM domain from 719 prokaryotic species using PfamScan (53; 54), and estimated optimal growth temperatures from their tRNA sequences (55). We then tested whether genetic variation in the lipase class 3 domain was significantly associated with optimal growth temperature in bacteria. We identified several significant associations, all of which were located in the flap lid region (Figure 5B, bold letters). Notably, the presence of a valine at residue 211, as observed in the PT allele, was associated with lower bacterial optimal growth temperatures relative to bacterial lipases that harbor the isoleucine residue observed in B73 (Figure 5B), suggesting that the PT allele may be better adapted to the low temperatures to which highland maize is exposed. Figure 5 C shows the top I-TASSER model (plum) for the putative phospholipase encoded by *HPC1-B73*, overlaid on chain B of the crystal structure of phospholipase A1 from Arabidopsis thaliana (PDB 2YIJ, tan). Residues that differ between the PT and B73 HPC1 are shown in yellow. The catalytic triad and H400 identified from CDD conservation analysis are also labeled. However, our models put H416 in the catalytic triad instead of H400. Of the residues that differ between PT and B73 HPC1, residues 211, 448, and 449 are the closest to the catalytic triad. I211 is positioned on the N-terminal end of the flap lid domain and is well suited to stabilize, through hydrophobic interactions, binding of a lipid substrate. Replacement of I211 with V211 results in the loss of a methyl group and may influence the strength of these hydrophobic interactions, influencing substrate binding and/or the dynamics of the flap lid domain.

HPC1-PT was introgressed from teosinte mexicana and is conserved in Flint inbred lines

We next explored the segregation of the SNP causing the V211I change among other highland maize varieties. We detected the PT allele at high frequencies in highland landraces from México and Guatemala. In addition, the PT allele segregated in Southwestern US landraces. The B73 allele was fixed in lowland Mexican, South American and Andean landraces (Figure 6A). These results are consistent with our previous PBE results (Figure 2E). The PT allele was also present in teosinte, in one fourth of all teosinte *parviglumis* accessions tested and in both *mexicana* accessions reported in Hapmap 3 (56) (Figure 6A). This observation prompted us to examine whether the PT allele was the result of post-domestication introgression from teosinte *mexicana* during maize highland colonization, or whether it was selected from *parviglumis* standing variation.

To test for introgression from *mexicana*, we used f_d data from (14) and established that the genomic region containing *HPC1* shows signatures of introgression from *mexicana* into highland maize (Figure 6B). We then performed a haplotype network analysis using SNP data from the *HPC1* coding region of 1,160 Mexican accessions from the SeeD Dataset (19) that were homozygous for all SNPS across the coding region and the teosinte inbred lines (TIL) from Hapmap 3 (56). We thus identified nine haplotype groups that cluster mainly based on elevation (Figure 6C). The two major groups, II and VI, contained mainly lowland and highland landraces, respectively. The two *mexicana* teosinte inbred lines (TIL08 and 25) belonged to group IV (Figure 6C) together with highland landraces primarily collected in the Trans-Mexican Volcanic Belt (30/36 from the highlands of Jalisco, Michoacán, México, Puebla and Veracruz). We then examined whether this *mexicana* haplotype (denoted *ZxHPC1*) that is introgressed into Mesoamerican highland maize was also present in modern maize inbred lines. We performed a neighbor-joining cluster analysis 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 *HPC1-PT* haplotype and the other containing the *HPC1-B73* haplotype. PT and the teosinte *mexicana* inbred lines TIL-08 and TIL-25 clustered together with Northern European Flint inbred lines such as EP1, UH008, and UH009 (Figure 6D). Other Northern US flints, e.g., CM7, are also closely related to the *mexicana ZxHPC1* haplotype. These data suggest that after introgression into highland maize, the *ZxHPC1* haplotype was maintained in Flint materials adapted to cold environments in the Northern US, Canada and Europe.

Building on previous reports for a role of highly unsaturated PC species in determining flowering time (42; 43) and considering the significant accumulation of those PC species induced by the *HPC1-PT* allele, we asked if variation in *HPC1* might be associated with differences in flowering time in modern maize. We used a large gene expression dataset obtained from the 282 maize

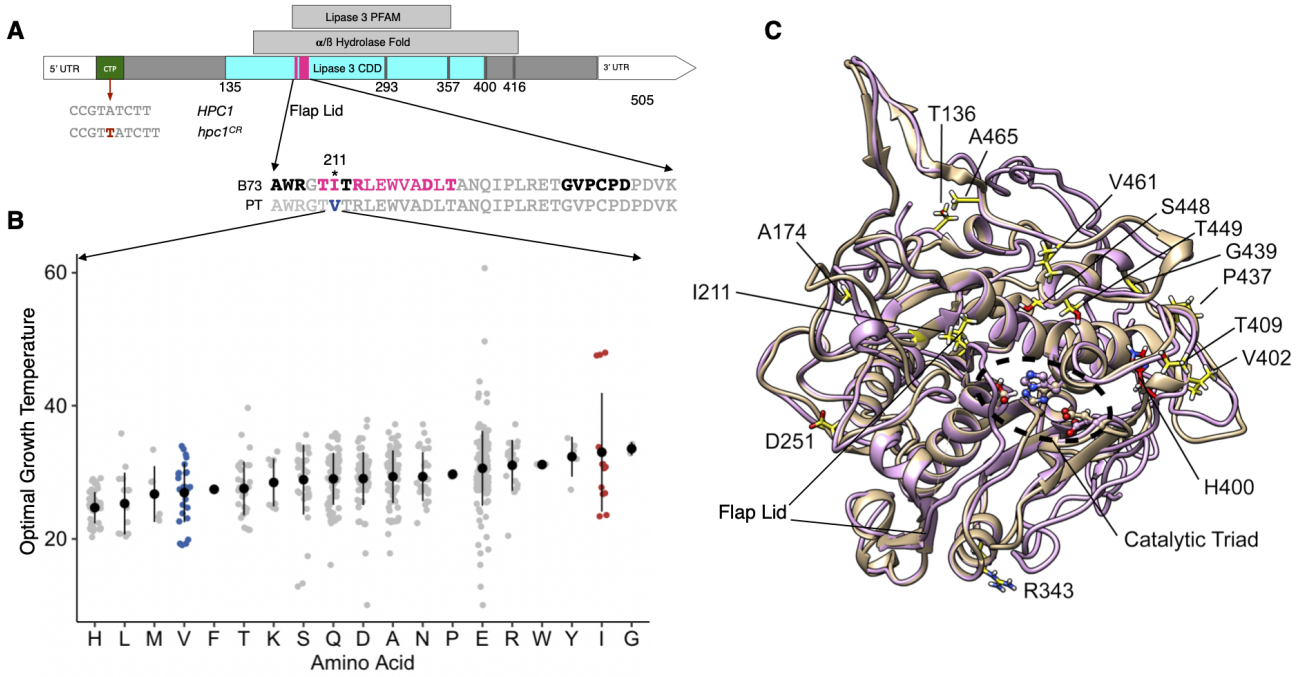

Figure 5 Fitness effects of the *HPC1-PT* causal SNP and the *hpc1*^{CR} mutant. A) *HPC1* coding sequence showing different features. CTP represents the chloroplast transit peptide. The arrow within CTP indicates the nucleotide insertion introduced by CRISPR/Cas9-mediated genome editing that leads to a frameshift. Magenta indicates the coding sequence fragment that encodes the flap lid domain and the corresponding amino acids. Residues in bold showed significant associations with prokaryote optimal growth temperature. **B)** Prokaryote optimal growth temperature (OGT) test of different amino acids at the polymorphic residue 211 of HPC1. **C)** Top I-TASSER model (plum) for the putative phospholipase encoded by HPC1-B73, overlaid on chain B of the crystal structure of phospholipase A1 from Arabidopsis (PDB 2YIJ, tan). Residues that differ between the PT and B73 phospholipase are shown in yellow and some are labeled

diversity panel that was sampled across several developmental stages (57), and phenotypic datasets collected from the same panel grown in long days and short days. We discovered that *HPC1* and *ZmLPCAT1* expression levels are inversely correlated in most tissues (Supplementary Figure 8), further supporting the idea that these two enzymes are coordinately regulated. There were significant associations between *HPC1* expression levels in aerial tissues and several flowering time traits. The magnitude of these associations was similar to those seen for other well-characterized flowering time genes (Supplementary Figure 8) such as *ZEA CENTRORADIALIS8* (*ZmZCN8*) and the AP2/ERF transcription factor gene *ZmRAP2.7*. Furthermore, in both long days and short days, lines carrying the *HPC1-PT* allele were characterized by lower *HPC1* expression and earlier flowering times relative to lines carrying the *HPC1-B73* allele (Figure 6E).

Taken together, these data show that *HPC1* was introgressed from teosinte *mexicana* into highland maize and that this introgression was carried over into high latitude adapted modern inbred lines that show low *HPC1* expression in association with early flowering.

Discussion

Understanding the genetic, molecular and physiological basis of crop adaptation to different environments and the role that wild relatives have played in these processes is relevant for the identification of 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. Recent studies (24; 23; 18) have helped expand our understanding of the genetic basis underlying maize highland adaptation. However, the molecular, physiological and genetic mechanisms underlying maize highland adaptation and the possible role of highland maize traits in modern, commercial varieties remain largely unknown. Phospholipids are key structural components of plant membranes that also function as signaling molecules during adaptation to stresses that would be prevalent in highland environments (50; 58) such as low phosphorus availability (59; 60; 61) and low temperatures (35; 36; 62). Additionally, in Arabidopsis, accumulation of certain highly unsaturated PC species can accelerate flowering time (42), which is a major driver of maize adaptation to highland environments (19; 20; 63).

We used genomic scans and linkage mapping analysis to identify and clone *High PhosphatidylCholine 1* (*HPC1*), which encodes a phospholipase A1 enzyme with a major role in determining phosphatidylcholine levels in highland maize landraces. We described how genetic variation at this locus shows strong antagonist pleiotropic G X E effects on several fitness traits. We then showed that the highland allele is the result of an introgression from teosinte *mexicana* that has been maintained in modern maize inbred lines adapted to high latitudes.

Genes involved in the biosynthesis and degradation of phospholipids were repeatedly selected in several highland maize populations of North America, Central America and South America (Figures 1 and 2).

ZmLPCAT1 and *HPC1* were two of the genes that showed the strongest, repeated signals of selection, as measured by PBE and *pcadapt* in highland populations (Figure 2). In a previous

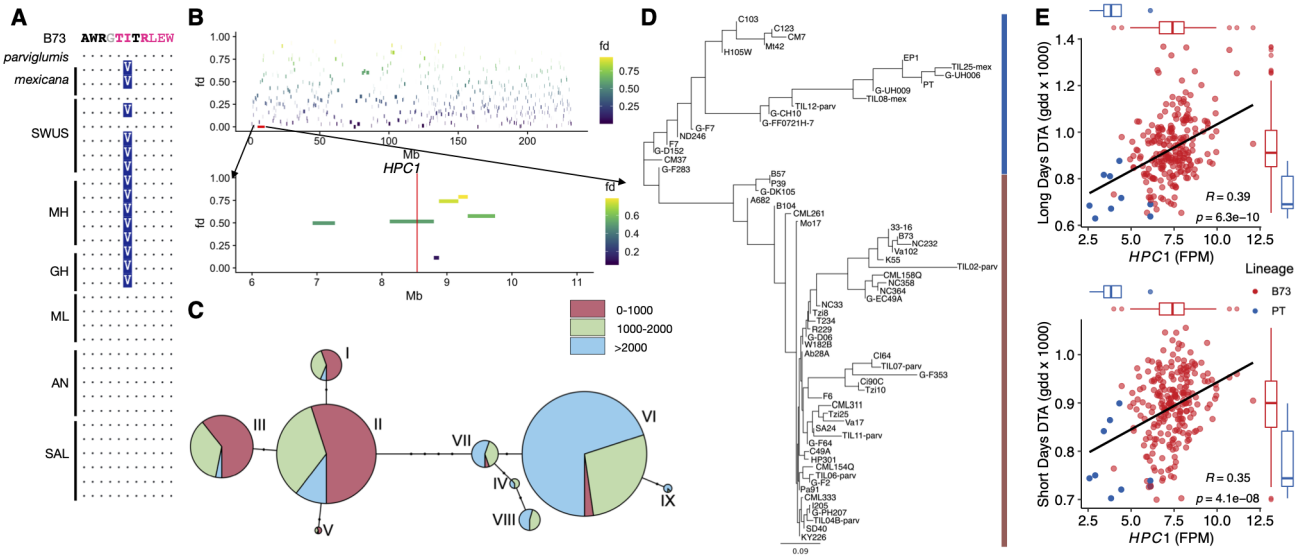

Figure 6 Introgression of teosinte *mexicana* into maize *HPC1*. A) Alignments around the V211I polymorphism in the flap-lid domain of HPC1 in B73, *mexicana* and *parviglumis* and landraces of the Southwestern US (SWUS), Mexican Highland (MH) and Lowlands (ML), Guatemalan Highlands (GH), Andes (AN) and South America Lowlands (SAL). **B)** f_d analysis of the *mexicana* introgression. Data were obtained from (14). **C)** Haplotype network analysis of SNPs within the *HPC1* coding region, using 1,060 Mexican homozygous individuals from the SeeD dataset. **D)** Cluster analysis of the *HPC1* coding region using a sample of Hapmap3 inbred lines and the PT landrace. **E)** Correlation between *HPC1-PT* expression and days to anthesis (DTA) in plants grown in short or long days. Inbred lines from the PT lineage shown in panel C are indicated in blue and inbred lines from the B73 lineage are in red; data from (57).

study, *ZmLPCAT1* exhibited high *Fst* values when highland and lowland landraces were compared (23). The predicted function of ZmLPCAT1 is to synthesize PC species via the acylation of LPC species, while HPC1 is predicted to hydrolyze PCs into LPCs and free fatty acids by cleaving the sn1-position. Selection at these two loci is likely to have driven the high PC/LPC ratio we observed in highland Mexican landraces (Figure 1). Natural variation at maize *HPC1* is associated with lipid levels (64) and flowering time (65; 27). In B73, *HPC1* is one of the most highly expressed phospholipase-encoding genes and its expression pattern is almost entirely restricted to vegetative leaves (V4-V9) (Supplementary Figure 3C) (52), which we sampled for our glycerolipid analysis. In B73 leaves, *HPC1* is the most highly expressed gene within the QTL interval (7.9 Mb to 10 Mb) (Supplementary Figure 5B (52)). In addition, *HPC1* is highly expressed in B73 and other temperate inbred lines grown in low temperature conditions and is down-regulated by heat stress (Supplementary Figure 3D) (66). In our experiments, *HPC1* was up-regulated in B73 and the B73 x PT F_1 hybrids after cold exposure. By contrast, the *HPC1-PT* allele was expressed at similar levels as *HPC1-B73* in control conditions, and was not up-regulated in cold conditions (Figure 3E). In Arabidopsis, *LPCAT1* is involved in modulating phosphorus levels in leaves under low zinc conditions (67). In maize, a positive signal from a genome-wide association study (GWAS) 45 kb upstream of *ZmLPCAT1* was associated with flowering time under low phosphorus conditions, further lending support to a possible role for natural variation at *ZmLPCAT1* in plant adaptation to low phosphorus availability.

Our QTL analysis of the PC/LPC ratios in a B73 x PT mapping population and in the *hpc1*^{CR} mutant demonstrates that the highland *HPC1-PT* allele results in an enzyme with impaired function that alters highland Mexican maize PC metabolism, leading to higher PC/LPC ratios (Figure 3). Using the PC/LPC ratio as a phenotype, we obtained much more significant LOD QTL signals than when using PC or LPC levels alone, possibly because the ratio can much more precisely capture genetically-induced changes in enzymatic activity (68). Adaptive loss-of-function mutations can be an effective way to gain new metabolic functions in new environmental conditions (69). Our data illustrate how an enzymatically impaired enzyme, due to a single conservative amino acid replacement located in the flap lid domain of HPC1, can affect substrate accessibility and/or substrate binding (Figure 5A). Indeed, the flap lid domain has been the target of biotechnological modification of these types of enzymes (70).

Why were the metabolic changes induced by HPC1 variation selected for in highland maize? PC metabolism is intimately connected to multiple stress responses and developmental pathways; alterations in PC amounts and PC/LPC ratios affect overall plant fitness. The *qPC/LPC3@8.5* QTL is driven by individual QTLs of PC and LPC species with high levels of unsaturated fatty acids (Supplementary Figure 2). Several of these species, like PC 36:5 and LPC 18:1 (Figure 3D, Sup. File 3), have been shown to display similar patterns during Arabidopsis cold acclimation (36) and sorghum (*Sorghum bicolor*) low temperature responses (62). PC 36:5 also showed high Q_{ST} values when comparing highland and lowland landraces from both Mesoamerica and South America (Figure 1C-D, Sup. File 5). In maize, *HPC1* expression is under the control of the circadian clock (71) with a peak at the end of the day. In Arabidopsis, highly unsaturated PC (34:3, 34:4, 36:5, 36:6) species increase in the dark (72). This peak of concentration coincides in maize with low expression levels of *HPC1* during the same dark hours (71). Furthermore, Yuki Nakamura and colleagues elegantly demonstrated that PC 36:5 and 36:6 species accumulate during the night and can bind to Arabidopsis FLOWERING LOCUS T (FT) to hasten flowering (42) by unknown cellular mechanisms. The Arabidopsis *FT* ortholog in maize, *ZCN8* (73), underlies a major flowering time

and photoperiod sensitivity QTL (27). Additive mutations in the regulatory region of *ZCN8*, including a teosinte *mexicana* introgression, lead to higher expression of *ZCN8*, which contributes to maize adaptation to long days in temperate conditions (74). Our results show that HPC1 drives the accumulation of the same PC species (such as PC 36:5) that accumulate in Arabidopsis at the end of the day, at which time they might bind to FT and activate the transition from vegetative to reproductive development. We hypothesize that the accumulation of these PC species in highland maize may drive early flowering in a manner similar to that in Arabidopsis.

The effects of genetic variation at *HPC1* on flowering time in Mexican landraces indicated a strong GxE interaction, whereby the highland allele leads to an acceleration of flowering and ASI in highland environments (Figure 4), similar to the effects observed with the well-known teosinte *mexicana* introgression *inv4m* (18). Other flowering time loci analyzed did not show this clear G x E effect (20). Interestingly, the effect of the highland allele exhibited typical conditional neutrality in yield-related traits, with increased fitness conferred by the *HPC1-PT* allele in highlands (Figure 4).

The mutation in the flap-lid of HPC1-PT at residue 211 is conserved in highland Mexican and Guatemalan landraces and in teosinte *mexicana* (Figure 6A) but is still segregating in teosinte *parviglumis*. We showed that *HPC1-PT* is indeed an introgression from teosinte *mexicana* (Figure 6B) and that this introgression has been maintained in modern inbred lines, particularly in the Northern US, Canadian and European Flint lines (Figure 6D). *HPC1* in inbred lines carrying the *ZxHPC1 mexicana* haplotype was expressed at low levels and resulted in earlier flowering (Figure 6E) (57). Other grasses such as eastern gamagrass (*Tripsacum dactyloides*) are adapted to temperate latitudes and also show accelerated rates of evolution in genes involved in PC metabolism (40). Moreover, genetic variation in the regulatory region of *HPC1* was significantly associated with photoperiod sensitivity in the maize Nested Association Mapping population (27).

In summary, we used here a combination of genomic scans, linkage mapping, lipidomics and reverse genetics to identify and clone the adaptive gene *HPC1*, introgressed from teosinte *mexicana*, in highland maize landraces. HPC1 leads to a major reorganization of phosphatidylcholine metabolism. We showed that the fitness advantage conferred by the *HPC1* highland *mexicana* allele is at least in part due to its effect on accelerating flowering and that this effect may have contributed to adaptation of maize to cold, high latitudes. This study highlights the largely underappreciated role of highland maize and highland teosinte *mexicana* in modern maize.

Materials and Methods

Populations used in the analysis. Highland and lowland populations used for Population Branch Excess analysis consisted of three to six accessions from each of the highland and lowland populations and have been previously described in (24; 75). The 120 landraces from the HiLo diversity panel were selected and ordered from the CIMMYT germplasm bank to maximize a good latitudinal gradient sampling across Mesoamerica and South America. For each highland landrace (>2,000 meters above sea level [masl]) a lowland landrace (<1,000 masl) was selected at the same latitude (<0.5°) to form 60 highland/lowland pairs, with 30 from each continent. The list of the accessions used is provided in Sup. file 1. B73 x Palomero Toluqueño (PT) recombinant inbred lines (RILs) were developed by crossing B73 with a single PT plant (Mexi5 accession, CIMMYTMA 2233) that was then backcrossed with B73 once and selfed five times (BC1S5). We used individual landrace accession genotype and fitness data from the CIMMYT Seeds of Discovery project (SeeD) (20) to calculate *pcadapt* (45) values and GxE effects of *HPC1-PT*.

Field Experimental Conditions and sampling. Two replicates of the HiLo diversity panel accessions and three replicates of the B73 x PT RILs were planted in a highland and lowland common gardens. The highland common garden was located in Metepec, Edo de México, (19°13'28.7"N 99°32'51.6"W) in the Trans-Mexican volcanic belt. The field is at 2,610 masl, and the range of average monthly temperatures along the year vary from 5°C to 21.5°C. The lowland common garden was located in Valle de Banderas, Nayarit, (20°47'01.2"N 105°14'47.0"W) in the Pacific Coast. The field is at 50 masl, and the range of average monthly temperatures along the year vary from 20°C to 29°C. Between growth stages V4 and V6, we used a leaf puncher to collect 50 mg of fresh tissue (10 discs) from the tip of the second leaf above the last leaf with a fully developed collar. Tissue discs were immediately flash-frozen in liquid nitrogen. We collected all samples from a field in a single day between 10:00 am and 12:00 pm, approximately 3 h after sunrise. Samples were transported in dry ice to the lab and stored at –80°C until extraction.

Glycerolipid analysis We crushed frozen samples in a Retsch tissue grinder (Haan, Germany) for 40 seconds at a frequency of 30 pulses per second. We performed lipid extraction following Matyash and collaborators (76). First, we added 225 µL of cold methanol (MeOH) to each sample. For the blanks, we prepared MeOH containing a Quality Control (QC) mix (Sup. File 6). We vortexed each sample for 10 sec, keeping the rest of the material on ice. Then, we added 750 µL of cold methyl tert-butyl ether (MTBE). For the blanks, we used MTBE containing 22:1 cholesterol ester as internal standard (Sup. File 6). We vortexed each sample for 10 sec, followed by 6 min of shaking at 4°C in the orbital mixer. We next added 188 µL of LC/MS grade water at room temperature (RT), and vortexed samples for 20 sec. We centrifuged the samples for 2 min at 14,000 rcf and recovered 700 µL of supernatant from the upper organic phase. We then split the supernatant into two aliquots of 350 µL each, one aliquot was used for the generation of the lipid profile and the other aliquot was used to prepare pools that were used along the lipid profiling. Finally, we dried the samples using a speed vacuum concentration system. We resuspended dried samples in 110 µL of MeOH-Toluene 90:10 (with the internal standard CUDA, 50 ng/mL). We vortexed samples at low speed for 20 sec and then sonicated them at RT for 5 min. We then transferred aliquots of 50 µL per sample into an insert within an amber glass vial. The UHPLC-QTOF MS/MS utilized were Agilent 1290 and Agilent 6530, respectively. We used a Waters Acquity charged surface hybrid (CSH) C18 2.1x100 mm 1.7 µm column which we initially purged for 5 min. We coupled the UHPLC column with a VanGuard pre-column (Waters Acquity CSH C18 1.7µm). We injected six "no sample injections" at the beginning of each run to condition the column, followed by ten samples, one pool (made out of the mix of the second aliquot of all the samples contained per UHPLC plate) and one blank. We injected 1.67 µL per sample into UHPLC-QTOF MS/MS ESI (+); the running time per sample was 15 min. Mobile phase "A" consisted of 60:40 acetonitrile:water, 10 mM of ammonium formate and 0.1% formic acid. Mobile phase "B" consisted of 90:10 isopropanol:acetonitrile, 10 mM ammonium formate and 0.1% of formic acid. The flow rate was 0.6

mL/min and the column compartment was maintained at 65° C. Initial conditions were 15% B; the gradient uniformly increased until reaching 100%. At 12.10 min the mobile phase composition returned to initial conditions. We used the mass spectrometer (Q-TOF MS/MS) in positive electrospray ionization mode (ESI). The source parameters were: ESI gas temperature 325 °C, nebulizer pressure 35 psig, gas flow 11L/min, capillary voltage 3500 V, nozzle voltage 1000 V, and MS TOF fragmentor and skimmer 120 and 65 V, respectively. We used a mass range between 60 and 1,700 m/z, under acquisition parameters. As for reference mass parameters, we used a detection window of 100 ppm and a minimum height of 1,000 counts. 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 we used as the current retention time of the experiment. We used the method retention time for internal standards and the current rt and we fitted a polynomial regression to calculate new retention times using retention times from 501 lipids of a MS1 m/z-rt library (See Sup. File 7). In MSDIAL (77), 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 selectable 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 (77) version 3.40. To use MSDIAL, we converted the raw data from .d to .abf format with Reifycs Abf converter (https://www.reifycs.com/AbfConverter/). Then, we filtered the MSDIAL alignment results based on whether compounds intensity was ten times above blank intensity. Next, filtered data were normalized using Systematic Error Removal using Random Forest (SERRF) (78). This normalization is based on the QC pool samples. We filtered out, now the normalized features, 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 (79). We also normalized the curated data using the sum of all known metabolite signal (mTIC). After data processing and normalization, we used lipid intensities for further analysis.

CRISPR-CAS9 Glycerolipid analysis Dry samples were resuspended in 110 µL of 100% MeOH (with the internal standard CUDA, 50 ng/mL), vortexed at low speed for 20 s and then sonicated at RT for 5 min. We transferred the samples into amber glass vials with inserts prior to analysis. Lipid profiling was performed using a Thermo Scientific Orbitrap Exploris 480 mass spectrometer coupled to a Thermo Vanquish Horizon UHPLC. Chromatographic separation was achieved using the same guard and analytical column type, mobile phase composition and gradient conditions described for UHPLC-QTOF analyses. Avanti Splash Lipidomix Mass Spec Standard (Avanti Polar Lipids, Alabaster, Alabama) was used to check system suitability and a Lipid QC Standard Mixture (Sup. File 6) and pooled QC samples were used to monitor instrument performance. Methanol solvent blanks were injected at the beginning of the sequence and throughout the batch between approximately 10 injections and pooled QC samples were injected between every 6–8 sample injections. For ion source parameters the spray voltage was set at 3500 V, sheath gas 60 (arb), aux gas 15 (arb), sweep gas 2 (arb) and ion transfer tube and vaporizer temp at 350 °C. Full scan and data-dependent MS/MS (ddMS^2) scans were collected in positive ESI mode following 2 μL sample injections. Full scan spectra were obtained from *m/z* 220–1700 using an Orbitrap resolution setting of 120,000, RF lens set at 25%, normalized AGC target at 25% and maximum injection time set to auto. ddMS^2 scans were performed using a cycle time of 0.6 sec, intensity threshold of 5.0e4, and a 3 sec dynamic exclusion using a 10 ppm mass tolerance. Additional parameters for ddMS^2 experiments were as follows: Isolation window equal to 1 Da, stepped collision energy at 20 and 30%, Orbitrap resolution set to 15,000, scan range set to auto, normalized AGC target of 100% and a maximum ion injection time of 55 msec. Data files were uploaded to LipidSearch 4.2.2 (Thermo Scientific, Tokyo, Japan) for identification of PC and LPC species. Skyline-daily (80) was then used to obtain normalized peak areas for the identified PCs and LPCs using the internal standards PC(12:0/13:0) and LPC(17:0), respectively.

$Q_{ST} - F_{ST}$ analysis of glycerolipid data. Quantitative trait differentiation (Q_{ST}) was contrasted to the distribution of F_{ST} for neutral genetic markers (81). Highland/Lowland contrasts were considered separately for Mesoamerica and South America. A linear mixed effects model (R package lmer, function lmer) was used to partition phenotypic variance between population pairs (Mesoamerica/South America, all highland/all lowland, Mesoamerican highland/Mesoamerican lowland, South American highland/South American lowland).

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

Within-population and between-population variances were calculated with the R function VarCorr (R package `lme4`, 82), and were used to calculate Q_{ST} following the equation below:

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

in which σ^2_{GB} and σ^2_{GB} are the between- and within-population genetic variance components, respectively (44). Pairwise F_{ST} was calculated with the R function fst.each.snp.hudson (R package *dartR*, 83).

***pcadapt* analysis of biological adaptation in Mexican landraces.** In order to conduct genome scans for signatures of adaptation we used the *pcadapt* (45) package. *pcadapt* identifies adaptive loci by measuring how strongly loci are contributing to patterns of differentiation between major axes of genetic variation. Under simple models *pcadapt* captures major patterns of F_{ST} but is conducted in a way that does not require population delimitation (84). As the genome scan comparison requires a focal SNP to be compared to the first K principal components of the genotype data, it can be biased by large regions of low recombination that drive the major axes of variation in the principal components. Thus, when SNPs from these low recombination regions are compared against principal components driven by linked loci spurious signals may arise. To prevent this bias from occurring, we used custom scripts to calculate the principal component step separately based upon all the chromosomes except for the chromosome of the focal SNPs being tested. The genotype data we used for this analysis was GBS data from roughly 2,000 landraces of Mexican origin collected by CIMMYT (www.cimmyt.org) as part of the SeeDs of discovery initiative (https://www.cimmyt.org/projects/seeds-of-discovery-seed/). From this, we calculated the strength of association between each SNP and the first five principal com-

ponents (excluding the chromosome of the focal SNP) using the communality statistic as implemented in *pcadapt* version 3.0.4.

Glycerolipid pathways selection. We compiled a list of genes pertinent to glycerolipid metabolism starting with a search of all genes belonging to the *Zea mays* 'Glycerophospholipid metabolism' and 'Glycerolipid metabolism' KEGG pathways (85) (map identifiers: zma00564 and zma00561). With the NCBI Entrez gene identifiers in KEGG we retrieved the AGPv4 transcript identifiers used in Corncyc 8.0 (86; 87) from an id cross reference file found in MaizeGDB (86). 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 these, 37 genes were found to have an enzymatic activity related to phospholipid metabolism but not placed into any particular pathway, i.e orphan enzymes, consisting mostly of alcohol dehydrogenases. Sixteen additional genes found in KEGG were not annotated at all in Corncyc probably due to differences between AGPv4 and RefSeq pseudogene annotation of the maize genome. The list of all possible candidates coming either from KEGG or Corncyc that were orphan enzymes or were unannotated in Corncyc amounted to 594 genes (Sup. File 2). This process is documented in the `0_get_glycerolipid_genes.R` script of the `pgplipid` R package accompanying this paper (88).

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. We used calculated PBE SNP values for the four populations (described in detail in (24)) and we tested for selection outliers SNPs in the 594 phosphoglycerolipid-related candidate genes and the 30 Corncyc pathways (556 genes). We first defined PBE outlier SNPs as the top 5% of the PBE score distribution; this fraction corresponds to approximately 50000 out of 1 million genotyped SNPs in each population. Following (24), we defined a gene as a PBE outlier if it contained an outlier SNP within the coding sequence or 10 kb upstream/downstream. Then we tested for over-representation of genes selected in particular subsets of populations using Fisher's exact test with the 32,283 protein-coding genes from the maize genome as background. For each pathway, we first selected all SNPs within coding regions and 10 kb upstream and downstream of genes in the pathway and we calculated the mean pathway PBE score. We then constructed a null distribution by drawing 10,000 samples without replacement of n SNPs from those found within or around 10 kb upstream and downstream of all protein-coding genes and we obtained the mean PBE for this null distribution. With the set of PBE outliers for glycerolipid metabolism in the four populations, we tested for evidence of physiological or pleiotropic constraint using the C^2_χ statistic (47).

QTL analysis of phospholipid levels. We analyzed glycerophospholipid QTLs in a mapping population of 57 RILs (BC1S5) from the cross B73 x PT. These RILs were grown in a highland site in Metepec, Edo de México at 2,600 masl during the summer of 2016 and in Puerto Vallarta, Jalisco at 50 masl during the winter of 2016/17. We analyzed the samples using UHPLC-QTOF, as above, and 67 leaf lipid species were identified. For QTL analysis we calculated the mean across all fields of individual lipid mass signal. We also used as phenotypes the sum total of the following lipid classes: diacylglycerols, triacylglycerols, PCs and LPCs. Furthermore, we also included the log base 10 transformed ratios of LPCs/PCs and the ratios of their individual species. We did a simple single marker analysis with "scanone" using Haley-Knott regression, and assessed the QTL significance with 1,000 permutations.

CRISPR-Cas9 editing of *HPC1* and analysis of the effect of *hpc1*^{CR} mutant on flowering time. CRISPR/Cas9 was used to create a *hpc1* mutant through Agrobacterium *Agrobacterium tumefaciens*-mediated transformation of background line B104 (89; 90). RNA guides (gRNAs) were designed as described in (91) for the B73 reference genome v4. The B104 and B73 sequence for *HPC1* were identical. The gRNA cassette was cloned into pGW-Cas9 using Gateway cloning. Two plants from the T_0 transgenic events were identified through genomic PCR amplification and Sanger sequencing and were self-pollinated. In the next generation, several plants were allowed to self. Plants were genotyped using forward primer CAGTTCTCATCCCATGCACG and reverse primer CCTGATGAGAGCTGAGGTCC. Homozygous *hpc1*^{CR} T2 plants were usedd for lipid analysis in greenhouses at North Carolina State University during Spring 2020 and then self-pollinated. Cas9 positivity was tested for using 0.05% glufosinate ammonium contained in Liberty herbicide. Wild-type plants from the same segregating families were used for analysis.

Expression analysis of *HPC1* in B73, PT and B73xPT F1 hybrids in conditions simulating highland environments. Gene expression data were generated from leaf tissue from B73, PT and the B73xPT F1 hybrids. Plants were grown following the same protocol as in (18). Briefly, kernels were planted in growth chambers set to imitate spring temperature conditions in the Mexican Lowlands (22°C night, 32°C day, 12 hr light) and highlands (11°C night, 22°C day, 12 hr light). Leaf tissue was sampled from the V3 leaf the day after the leaf collar became visible, between 2 and 4 hr after dawn. Tissue was immediately placed in a centrifuge tube, frozen in liquid nitrogen, and stored at –70°C.

RNA-seq libraries were constructed, sequenced, and analyzed following (18). Briefly, randomly primed, strand specific, mRNA-seq libraries were constructed using the BRaD-seq (92) protocol. Multiplexed libraries were sequenced on 1 lane of an Illumina HiSeq X platform. Low quality reads and adapter sequences were removed using Trimmomatic v.0.36 (93), and the remaining paired reads were aligned to the maize reference genome B73 v4 and quantified using `kallisto` v.0.42.3 (94). Gene counts were normalized using the weighted trimmed mean of M-values (TMM) with the *calcNormFactors* function in *edgeR* (95) and converted to log2CPM.

Sanger sequencing of *HPC1* in RILs homozygous at the *qPC/LPC3@8.5* locus. We identified 3 RILs homozygous for the PT allele at the *HPC1* locus and we developed 6 sets of primers to sequence by Sanger across the coding region and the promoter. The location of primers along the locus are shown in Sup. file 8.

Association of *HPC1* with agronomic traits. We re-analyzed phenotypic data from the F_1 Association Mapping (FOAM) panel of Romero-Navarro *et al* (19) and Gates *et al* (20) to more fully characterize association signatures of *HPC1*. Full descriptions of this experiment and data access are described in these references.

We downloaded BLUPs for each trait and line from Germinate 3, and subset the data to only those lines with GBS genotype data from México. We fit a similar model to the GWAS model used by (20) to estimate the effect of the *HPC1-PT* allele on the traits' intercept and slope on trial elevation, accounting for effects of tester ID in each field and genetic background and family effects on the trait intercept and slope using four independent random effects. We implemented this model in the *R* package *GridLMM* (48). We extracted effect sizes and covariances conditional on the REML variance component estimates and used these to calculate standard errors for the total *HPC1-PT* effect as a function of elevation. To test whether the phenotypic effects of *HPC1-PT* on yield components might be explained as indirect effects via flowering time, we additionally re-fit each model using days to anthesis (DTA) as a covariate with an independent effect in each trial.

Identification of features and domains of HPC1. The domains of *HPC1* 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 the S293, D357, and H400 catalytic triad were identified from CDD. H416 was identified as a substitute for H400 in the catalytic triad by protein modeling.

Bacterial optimal growth temperature association with HPC1 flap-lid domain allelic variation. We compared the maize HPC1 protein sequence to prokaryotes with the same sequence to determine whether the identified residue change in maize and accompanying association with low temperature survival was consistent with observations in other organisms. We identified the Pfam domain PF01764 in the B73 protein sequence using the HMMER3 web server, we also identified 982 observations of the PF01764 Pfam domain in 719 prokaryote species using PfamScan (53; 54). We predicted the optimal growth temperature of these species using tRNA sequences as in (55). We aligned the maize and prokaryote PF01764 domain sequences with hmmalign from the hmmer3 package (96), and the aligned Pfam sequences were recoded to reflect nine amino acid physicochemical properties (97). We filtered the sequences to remove gaps in the domain alignment and observations with only partial domain sequences, then we clustered them based on sequence similarity, resulting in two clusters of observations within the domain. For each cluster, positions in the filtered alignment were associated with prokaryote optimal growth temperatures using a linear regression with all nine amino acid physicochemical properties. Seventeen sites in and around the flap-lid region of the protein passed a 10% FDR significance threshold, including the single residue change V211I previously identified in *HPC1-PT*. Welch's two-sided t-test was used to compare the optimal growth temperatures of prokaryote species with the same (V) *HPC1-B73* aminoacid residue to the optimal growth temperatures of prokaryote species with the same (I) *HPC1-B73* allele at this site.

Subcellular localization of *HPC1*. We fused the HPC1-B73 chloroplast transit peptide (CTP, 52 aminocacids) to GFP. Three constructs encoding subcellular localization markers were used as control; cytoplasm (C-GFP), nucleus (N-GFP), and chloroplast (P-GFP). All markers were under the control of the cauliflower Mosaic Virus (CaMV) 35S promoter. These constructs were transiently infiltrated in *Nicotiana benthamiana* leaf cells.

Teosinte *mexicana* introgression in highland maize. To evaluate if the *HPC1-PT* allele is the result of standing variation from teosinte *parviglumis* or an introgression from *mexicana* we used Patterson's D statistic and genome-wide f_d to calculate ABBA-BABA patterns. The data were obtained from (14). The material used in (14) included whole-genome sequence data from three highland outbred individuals: two PT and one Mushito de Michoacán; three lowland landraces: Nal Tel (RIMMA0703) and Zapalote Chico (RIMMA0733) obtained from (75) and BKN022 from (56); two *mexicana* inbred lines: TIL08 and TIL25; three *parviglumis* inbred lines: TIL01, TIL05, TIL10 and *Tripsacum* TDD39103 (56) as an outlier.

Haplotype network analysis of *HPC1* in Mexican maize landraces and teosintes. We extracted SNP genotypes for *HPC1* from the TIL teosinte accessions in the HapMap 3 imputed data (56) and the 3700 Mexican landraces in the SeeD dataset. With the set of 1060 accessions that were homozygous at all sites in this genomic region we calculated a haplotype network depicting the minimal spanning tree for haplotypes covering 90% of the input accessions with the R package `pegas` (98), and haplotype frequencies for three elevation classes in the landraces (0–1,000; 1,000–2,000 and >3,000 masl).

Clustering analysis of *HPC1* in maize inbred lines and teosintes. Using v3 of the B73 genome, *HPC1* SNPs were obtained from PT (this paper), the 282 diversity panel (99), teosinte inbred lines and the German inbred lines from HapMap 3 (56). The selection was made to have a good representation of tropical, temperate and European inbred lines together with teosintes and PT lines. SNPs were aligned using Geneious2020.0.5 and a neighbor-joining cluster analysis was generated. To facilitate visualization and interpretation of the tree we condensed cluster branches from lines with identical haplotypes and from similar geographic locations. The full tree is available as Sup. file 9.

Expression analysis of candidate genes and association with flowering traits in the 282 diversity panel. We used gene expression RNA-seq data obtained from the 282 diversity panel at different developmental stages (57) and BLUP values of several flowering and photoperiod sensitivity traits (27) to study the correlation of *HPC1* expression values with flowering time traits.

Data Availability

Data and code used in the paper can be found in the Github repository of the project.

Acknowledgments

This work was supported by Conacyt Young Investigator (CB-238101), Conacyt National Problems (APN-2983), UC-Mexus and NC State startup funds awarded to Rubén Rellán-Álvarez. Fieldwork and mapping population development was supported by NSF-PGR award 1546719 to Jeffrey Ross-Ibarra, Ruairidh Sawers, Daniel Runcie and Matthew Hufford. Fausto Rodríguez-Zapata (2018-000012-01NACF-12692) and Karla Blöcher-Juárez were supported by CONACYT doctoral fellowships. Allison Barnes was supported by NSF-PGRP PRFB grant 2010703. Andi Kur was supported as an AgBioFEWS fellow by the NSF award 1828820. We thank Josh Strable, Peter Balint-Kurti and Luis Herrera-Estrella for critical review of the manuscript and feedback. We thank the Rellán-Álvarez, Sawers, Hufford, Flint-Garcia lab members and CIMMYT and Puerto Vallarta Winter Nursery crews that have helped generate and evaluate the mapping populations used in this manuscript. We specially want to acknowledge the indigenous people of the Americas and the ingenuity through which they domesticated and made possible the spread and adaptation of maize throughout the continent.

References

[1] C Natarajan, et al., Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **354**, 336–339 (2016).

[2] X Yi, et al., Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).

[3] A Bigham, et al., Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. *PLoS genetics* **6**, e1001116 (2010).

[4] X Liu, et al., EPAS1 gain-of-function mutation contributes to high-altitude adaptation in tibetan horses. *Molecular biology and evolution* (2019).

[5] J Yang, et al., Genetic signatures of high-altitude adaptation in tibetans. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4189–4194 (2017).

[6] JP Velotta, CE Robertson, RM Schweizer, GB McClelland, ZA Cheviron, Adaptive shifts in gene regulation underlie a developmental delay in thermogenesis in High-Altitude deer mice. *Molecular biology and evolution* **37**, 2309–2321 (2020).

[7] F Cicconardi, et al., Genomic signature of shifts in selection in a subalpine ant and its physiological adaptations (2020).

[8] EC Gilmore, Jr., JS Rogers, Heat units as a method of measuring maturity in corn 1. *Agronomy journal* **50**, 611–615 (1958).

[9] JL Hatfield, JH Prueger, Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes* **10**, 4–10 (2015).

[10] Y Matsuoka, et al., A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6080–6084 (2002).

[11] DR Piperno, AJ Ranere, I Holst, J Iriarte, R Dickau, Starch grain and phytolith evidence for early ninth millennium BP maize from the central balsas river valley, mexico. *Proceedings of the National Academy of Sciences* **106**, 5019–5024 (2009).

[12] DR Piperno, KV Flannery, The earliest archaeological maize (zea mays l.) from highland mexico: new accelerator mass spectrometry dates and their implications. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2101–2103 (2001).

[13] MB Hufford, et al., The genomic signature of crop-wild introgression in maize. *PLoS genetics* **9**, e1003477 (2013).

[14] E Gonzalez-Segovia, et al., Characterization of introgression from the teosinte zea mays ssp. mexicana to mexican highland maize. *PeerJ* **7**, e6815 (2019).

[15] JA Aguirre-Liguori, et al., Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (zea mays parviglumis and zea mays mexicana). *Molecular ecology* **24**, 2663 (2019).

[16] N Lauter, C Gustus, A Westerbergh, J Doebley, The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* **167**, 1949–1959 (2004).

[17] AK Hardacre, HA Eagles, Comparisons among populations of maize for growth at 13°c. *Crop science* **20**, 780–784 (1980).

[18] T Crow, et al., Gene regulatory effects of a large chromosomal inversion in highland maize. *PLoS genetics* **16**, e1009213 (2020).

[19] JA Romero Navarro, et al., A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature genetics* **49**, 476 (2017).

[20] DJ Gates, D Runcie, GM Janzen, AR Navarro, others, Single-gene resolution of locally adaptive genetic variation in mexican maize. *bioRxiv* (2019).

[21] J Stephen Athens, et al., Early prehistoric maize in northern highland ecuador. *Latin American Antiquity* **27**, 3–21 (2016).

[22] A Grobman, et al., Preceramic maize from paredones and huaca prieta, peru. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1755–1759 (2012).

[23] S Takuno, et al., Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics*, genetics.115.178327 (2015).

[24] L Wang, et al., Molecular parallelism underlies convergent highland adaptation of maize landraces. *bioRxiv* (2020).

[25] RR da Fonseca, et al., The origin and evolution of maize in the southwestern united states. *Nature plants* **1**, 14003 (2015).

[26] K Swarts, et al., Genomic estimation of complex traits reveals ancient maize adaptation to temperate north america. *Science* **357**, 512–515 (2017).

[27] HY Hung, et al., ZmCCT and the genetic basis of daylength adaptation underlying the postdomestication spread of maize (2012).

[28] Z Liang, et al., Conventional and hyperspectral time-series imaging of maize lines widely used in field trials. *GigaScience* **7**, 1–11 (2018).

[29] L Guo, et al., Stepwise cis-regulatory changes in ZCN8 contribute to maize Flowering-Time adaptation. *Current biology: CB* **28**, 3005–3015.e4 (2018).

[30] ND Coles, MD McMullen, PJ Balint-Kurti, RC Pratt, JB Holland, Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics* **184**, 799–812 (2010).

[31] C Huang, et al., ZmCCT9 enhances maize adaptation to higher latitudes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E334–E341 (2018).

[32] Q Yang, et al., CACTA-like transposable element in 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] JT Brandenburg, et al., Independent introductions and admixtures have contributed to adaptation of european maize and its american counterparts. *PLoS genetics* **13**, e1006666 (2017).

[35] T Degenkolbe, et al., Differential remodeling of the lipidome during cold acclimation in natural accessions of arabidopsis thaliana. *The Plant journal: for cell and molecular biology* **72**, 972–982 (2012).

[36] R Welti, et al., Profiling membrane lipids in plant stress responses: Role of phospholipase Dα in freezing-induced lipid changes in arabidopsis. *The Journal of biological chemistry* **277**, 31994–32002 (2002).

[37] DV Lynch, PL Steponkus, Plasma membrane lipid alterations associated with cold acclimation of winter rye seedlings (secale cereale l. cv puma). *Plant physiology* **83**, 761–767 (1987).

[38] J Jouhet, Importance of the hexagonal lipid phase in biologi-

cal membrane organization. *Front. Plant Sci.* **4**, 494 (2013).
[39] JR Henriksen, TL Andresen, LN Feldborg, L Duelund, JH Ipsen, Understanding detergent effects on lipid membranes: a model study of lysolipids. *Biophys. J.* **98**, 2199–2205 (2010).
[40] 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).
[41] Y Gu, et al., Biochemical and transcriptional regulation of membrane lipid metabolism in maize leaves under low temperature. *Front. Plant Sci.* **8**, 2053 (2017).
[42] Y Nakamura, et al., Arabidopsis florigen FT binds to diurnally oscillating phospholipids that accelerate flowering. *Nature communications* **5**, 3553 (2014).
[43] 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).
[44] 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).
[45] 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).
[46] JE Pool, DT Braun, JB Lack, Parallel evolution of cold tolerance within drosophila melanogaster. *Molecular biology and evolution* **34**, 349–360 (2017).
[47] 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.
[48] DE Runcie, L Crawford, Fast and flexible linear mixed models for genome-wide genetics. *PLOS Genetics* **15**, 1–24 (2019).
[49] KW Broman, H Wu, S Sen, GA Churchill, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
[50] SB Ryu, Phospholipid-derived signaling mediated by phospholipase a in plants. *Trends in plant science* **9**, 229–235 (2004).
[51] 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).
[52] 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).
[53] SC Potter, et al., HMMER web server: 2018 update (2018).
[54] S El-Gebali, et al., The pfam protein families database in 2019. *Nucleic acids research* **47**, D427–D432 (2019).
[55] E Cimen, SE Jensen, ES Buckler, Building a tRNA thermometer to estimate microbial adaptation to temperature. *Nucleic Acids Research* (2020).
[56] R Bukowski, et al., Construction of the third generation zea mays haplotype map. *GigaScience* (2017).
[57] KAG Kremling, et al., Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).
[58] Y Nakamura, Plant phospholipid diversity: Emerging functions in metabolism and Protein–Lipid interactions. *Trends in plant science* **22**, 1027–1040 (2017).
[59] EJ Veneklaas, et al., Opportunities for improving phosphorus-use efficiency in crop plants: Tansley review. *The New phytologist* **195**, 306–320 (2012).
[60] 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).
[61] 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).
[62] SR Marla, et al., Comparative transcriptome and lipidome analyses reveal molecular chilling responses in Chilling-Tolerant sorghums. *Plant Genome* **10** (2017).
[63] KL Mercer, H Perales, Structure of local adaptation across the landscape: flowering time and fitness in mexican maize (zea mays l. subsp. mays) landraces. *Genetic resources and crop evolution* **66**, 27–45 (2019).
[64] C Riedelsheimer, et al., Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics* **44**, 217–220 (2012).
[65] 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).
[66] AJ Waters, et al., Natural variation for gene expression responses to abiotic stress in maize. *The Plant Journal* **89**, 706–717 (2017).
[67] M Kisko, et al., LPCAT1 controls phosphate homeostasis in a zinc-dependent manner. *eLife* **7** (2018).
[68] AK Petersen, et al., On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC bioinformatics* **13**, 120 (2012).
[69] AK Hottes, et al., Bacterial adaptation through loss of function. *PLoS genetics* **9**, e1003617 (2013).
[70] FI Khan, et al., The lid domain in lipases: Structural and functional determinant of enzymatic properties. *Frontiers in bioengineering and biotechnology* **5**, 16 (2017).
[71] S Khan, SC Rowe, FG Harmon, Coordination of the maize transcriptome by a conserved circadian clock. *BMC Plant Biol.* **10**, 126 (2010).
[72] 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).
[73] 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).
[74] T Guo, et al., Optimal designs for genomic selection in hybrid crops. *Molecular plant* **12**, 390–401 (2019).
[75] L Wang, et al., The interplay of demography and selection during maize domestication and expansion. *Genome biology* **18**, 215 (2017).
[76] 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).
[77] H Tsugawa, et al., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* **12**, 523–526 (2015).
[78] S Fan, et al., Systematic error removal using random for-

est for normalizing large-scale untargeted lipidomics data. *Analytical Chemistry* **91**, 3590–3596 (2019) PMID: 30758187.

[79] BC DeFelice, et al., Mass spectral feature list optimizer (msflo): 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.

[80] KJ Adams, et al., Skyline for small molecules: A unifying software package for quantitative metabolomics. *J. Proteome Res.* **19**, 1447–1458 (2020).

[81] MC Whitlock, Evolutionary inference from qst. *Molecular ecology* **17**, 1885–1896 (2008).

[82] D Bates, M Maechler, B Bolker, S Walker, lme4: Linear mixed-effects models using eigen and s4. r package version 1.1-7 (2014).

[83] 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).

[84] 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).

[85] M Kanehisa, Y Sato, M Furumichi, K Morishima, M Tanabe, New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**, D590–D595 (2019) Publisher: Oxford Academic.

[86] JL Portwood, et al., MaizeGDB 2018: the maize multigenome genetics and genomics database. *Nucleic Acids Research* **47**, D1146–D1154 (2019) Publisher: Oxford Academic.

[87] 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).

[88] FR Zapata, sawers-rellan-labs/pglipid: Glycerophospholipid Pathway analysis (2020).

[89] Q Wu, et al., The maize heterotrimeric G protein β subunit controls shoot meristem development and immune responses. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1799–1805 (2020).

[90] SN Char, et al., An agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* **15**, 257–268 (2017).

[91] VA Brazelton, Jr, et al., A quick guide to CRISPR sgRNA design tools. *GM Crops Food* **6**, 266–276 (2015).

[92] 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).

[93] AM Bolger, M Lohse, B Usadel, Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

[94] NL Bray, H Pimentel, P Melsted, L Pachter, Near-optimal probabilistic rna-seq quantification. *Nature biotechnology* **34**, 525 (2016).

[95] MD Robinson, DJ McCarthy, GK Smyth, edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

[96] SR Eddy, Accelerated profile HMM searches. *PLoS computational biology* **7**, e1002195 (2011).

[97] 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).

[98] E Paradis, pegas: an R package for population genetics with an integrated modular approach. *Bioinformatics* **26**, 419–420 (2010).

[99] SA Flint-Garcia, et al., Maize association population: a high-resolution platform for quantitative trait locus dissection: High-resolution maize association population. *The Plant journal: for cell and molecular biology* **44**, 1054–1064 (2005).

Supplementary Tables and Figures

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

Table 1 Pairwise C_{hyper} statistic for population comparisons. "pop1" and "pop2" are the two populations being compared

			Broad Trait Heritability Estimates		
QTL	LOD	% variance	Simple Balanced	Piepho	Cullis
qPCs/LPCs3@8.5	24.5	87	0.751	0.803	0.797
qLPCs3@8.5	9.2	53	0.714	0.759	0.766
qPCs3@8.5	5.6	37	0.629	0.687	0.688

Table 2 Phenotypic variance explained by QTLs at chromosome 3, position 8,542,287 (AGPv3).

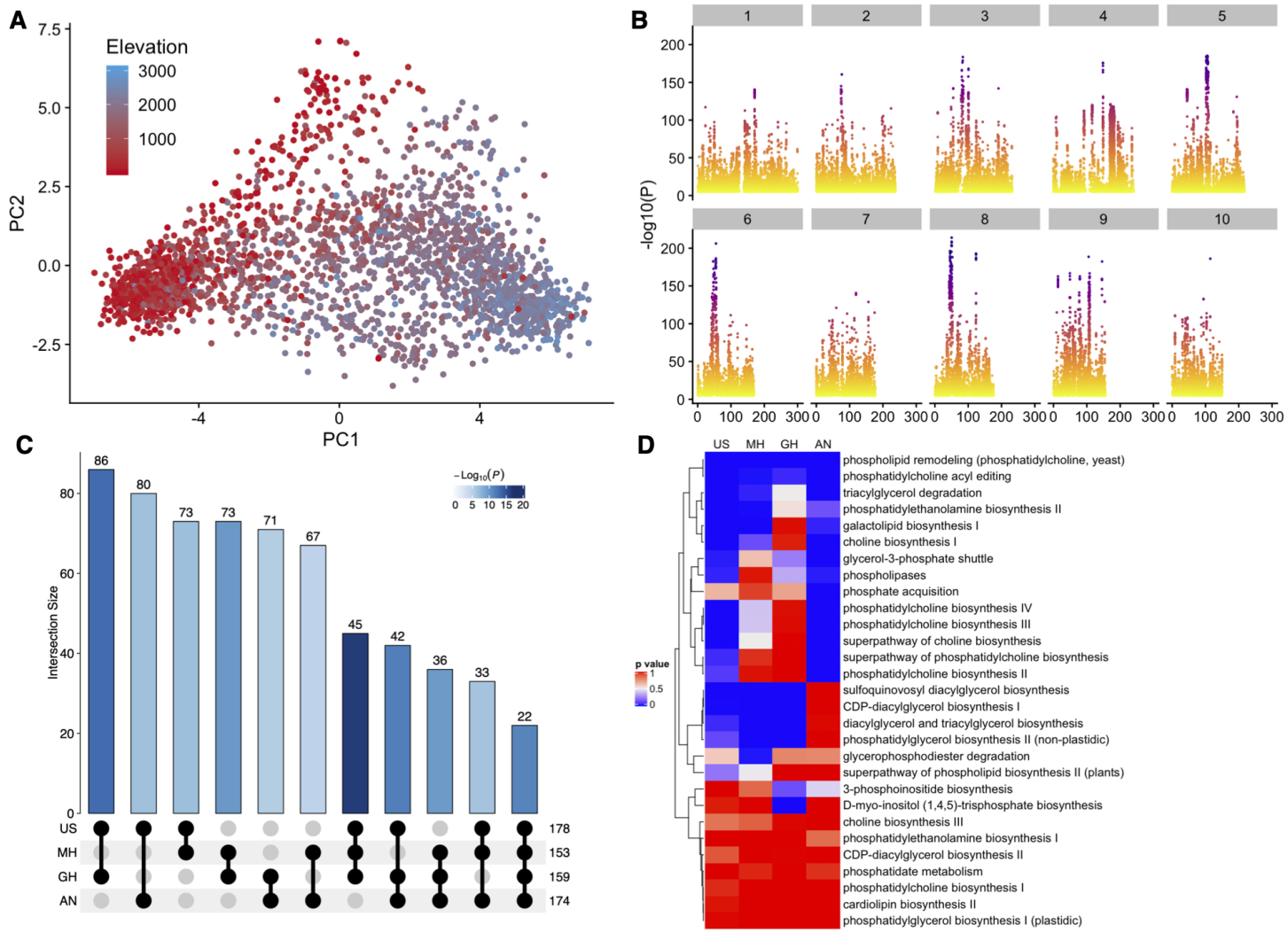

Figure Supplementary Figure 1 *Pcadapt* analysis of Mexican landraces. We used GBS data from the Mexican landraces of the SeeD dataset (19) and run a *pcadat* analysis (45) that identified **A)** elevation as the major driver of population differentiation, separated along Principle Component 1 (PC1). **B)** Genome-wide analysis of *Pcadapt* PC1 outliers. **C)** Population Branch Excess (PBE) Analysis of genes in glycerolipid pathways. Empirical p-value for the observed mean PBE of SNPs in genes annotated with CornCyc glycerolipid pathways. Null distribution calculated from random sampling (without replacement, 10000 replicates) of an equivalent number of SNPs in protein coding genes. **D)** Highland selection of glycerolipid-related pathways using PBE. See methods for details. For each pathway, we first selected all SNPs in the coding regions and 10 kb upstream and downstream of each gene to calculate the mean pathway PBE score. We then constructed a null distribution by drawing 10,000 samples without replacement of *n* SNPs from those found within or around 10 kb upstream and downstream of all protein-coding genes and we obtained the mean PBE for this null distribution. For each pathway, the heat map shows p-values corresponding to the probability of sampling from the null distribution a set of *n* SNPs with the observed glycerolipid-related pathway mean PBE score or higher.

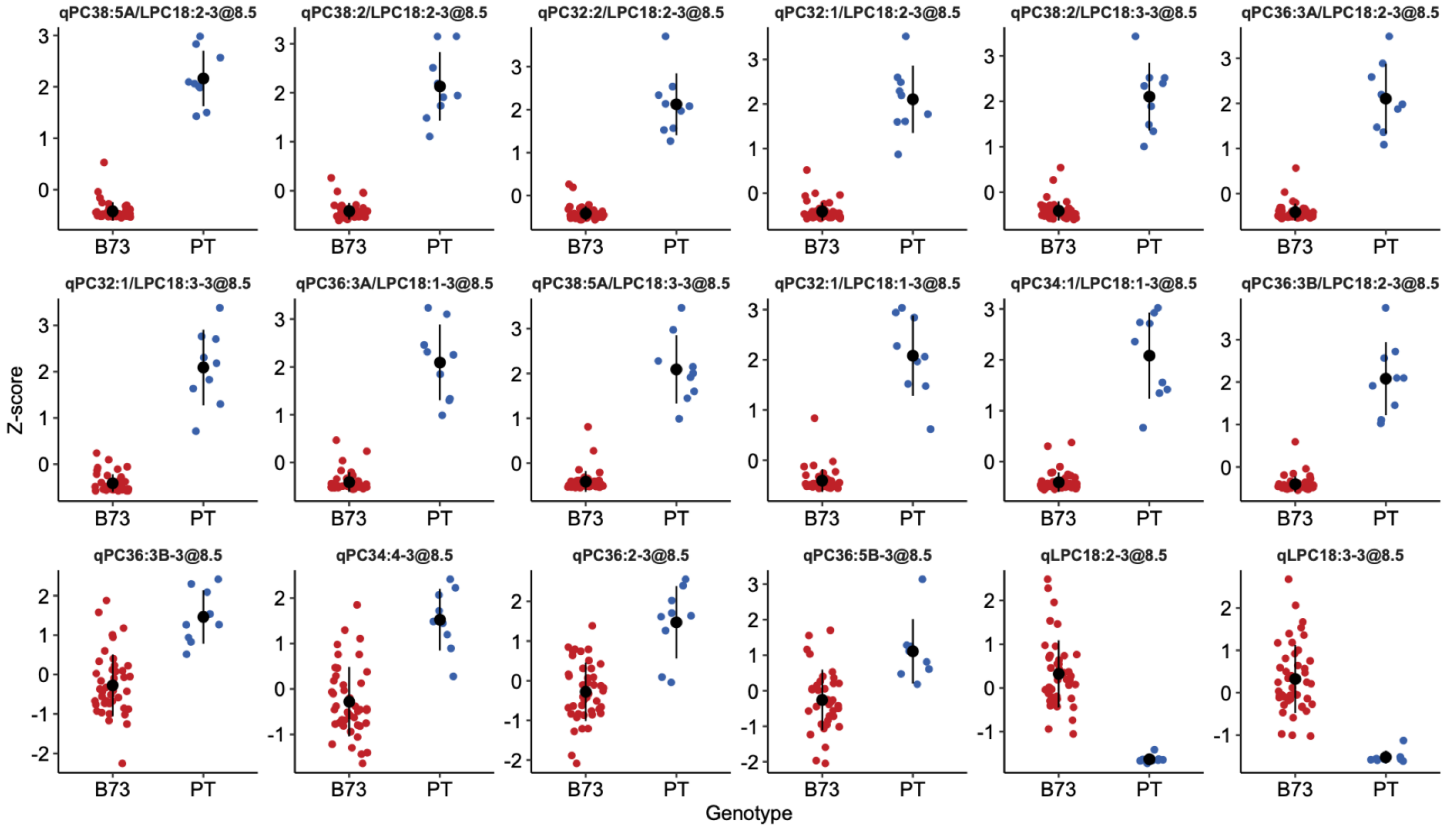

Figure Supplementary Figure 2 Effects of QTLs at chromosome 3 @8.5 Mb for individual PC and LPC species (bottom), and PC/LPC ratios (12 most significant peaks).

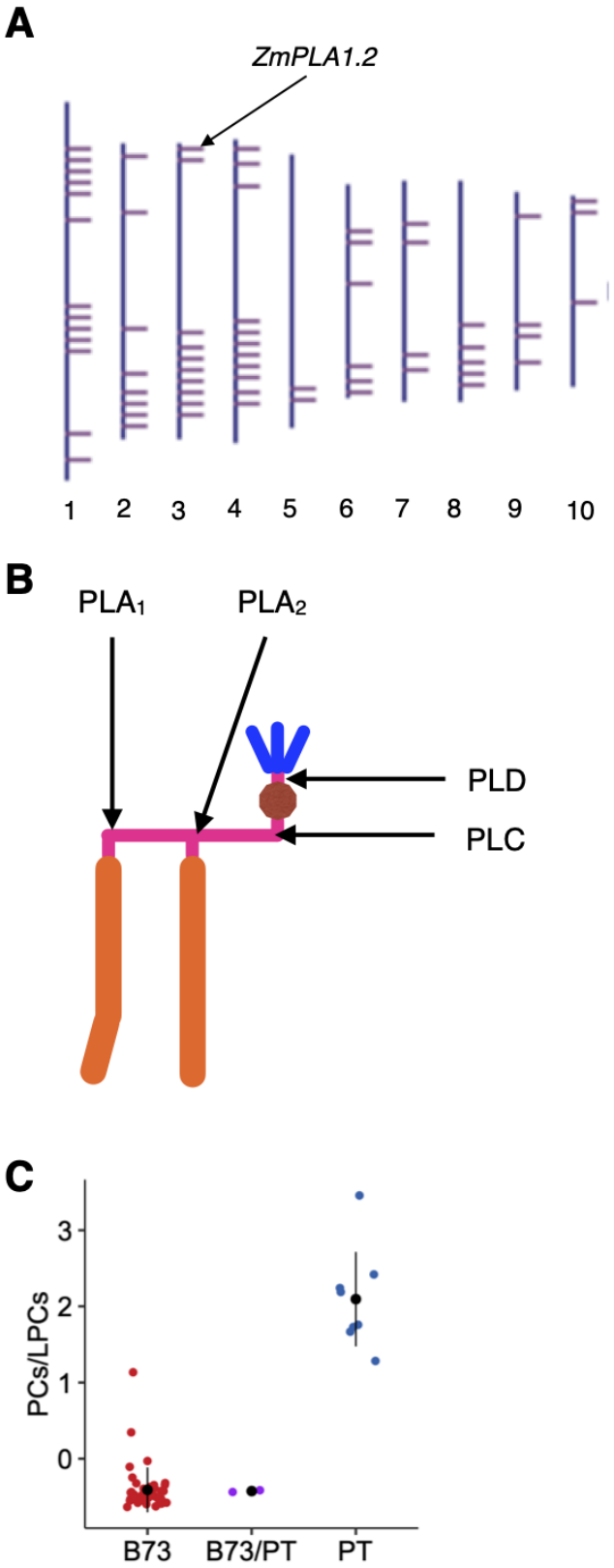

Figure Supplementary Figure 3 A) GenomiclLocation of genes encoding proteins with predicted phospholipase A1 activity. **B)** Site of action of the different types of phospholipases. **C)** Effect sizes of PC/LPC levels at RILs homozyzogous for B73, PT or heterozygous at the QTL qPC/LPC3@8.5.

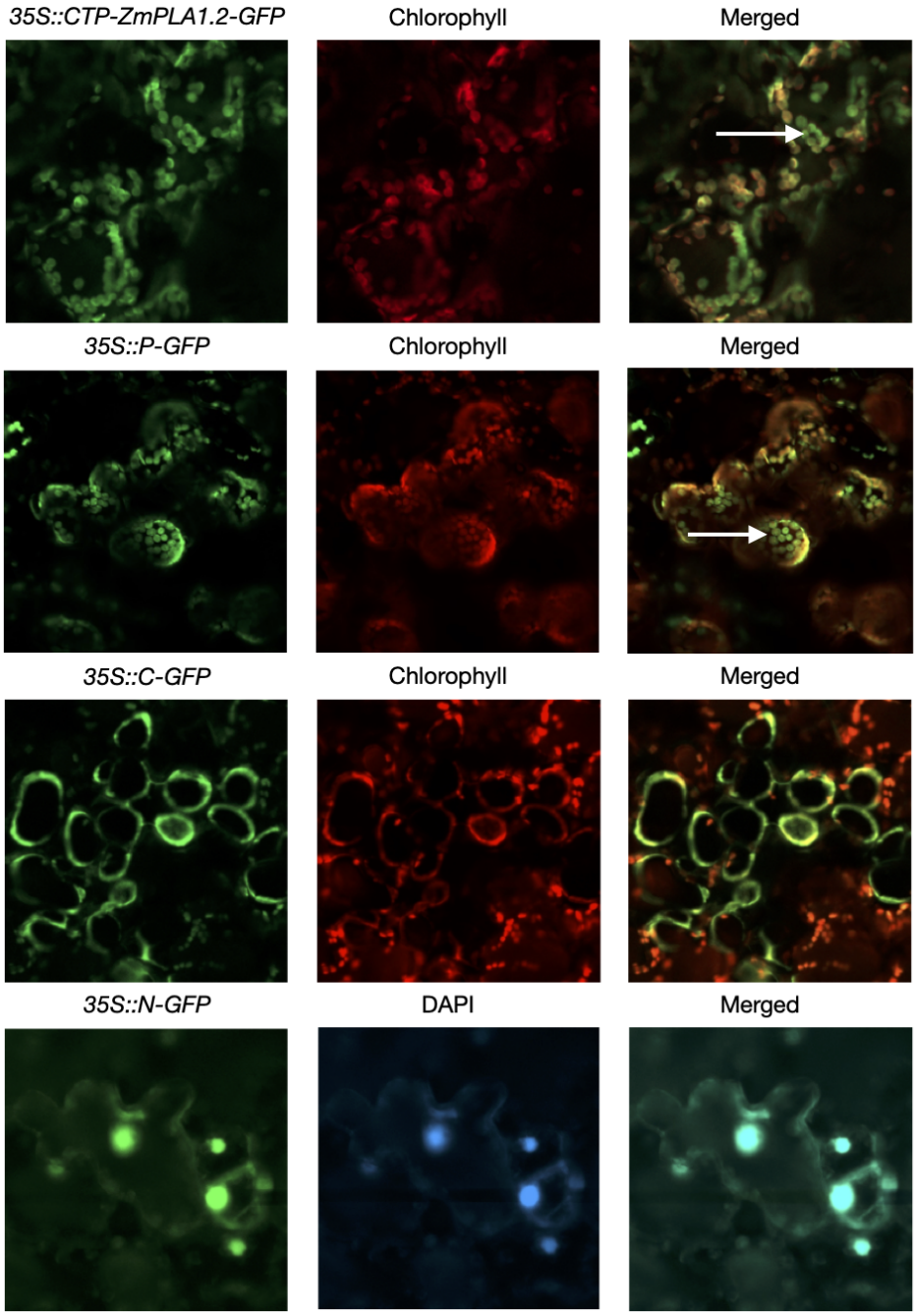

Figure Supplementary Figure 4 Subcellular localization of HPC1 in *Nicotiana benthamiana* epidermal leaf cells. CTP.HPC1-GFP, corresponding to the first 52 amino acids HPC1, were fused to GFP. Three constructs for various subcellular compartments were used as control; cytoplasm (C-GFP), nucleus (N-GFP), and chloroplast (P-GFP). All reporters were under the control of the cauliflower Mosaic Virus (CaMV) 35S promoter. The white arrows show chloroplast and nucleus.

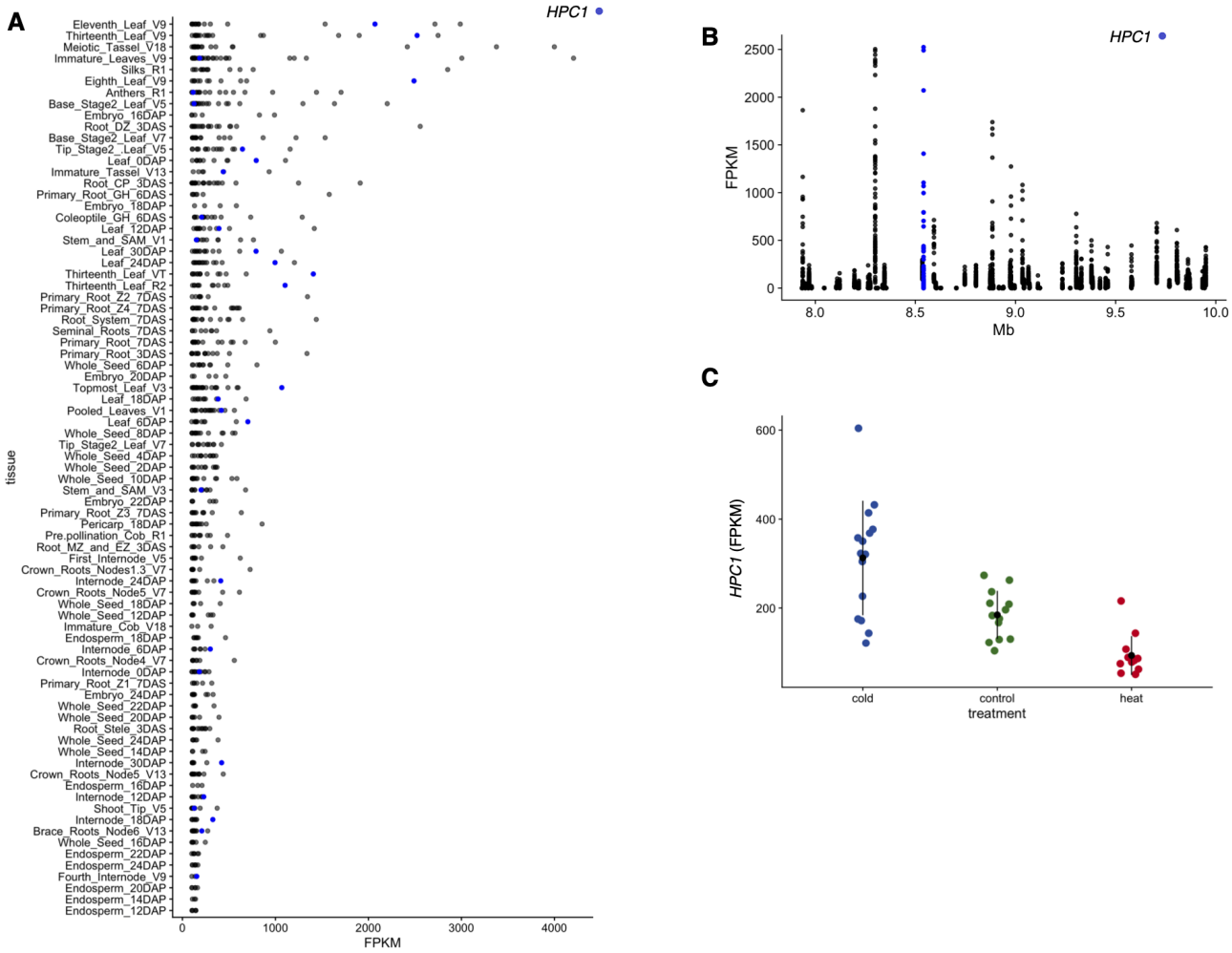

Figure Supplementary Figure 5 A) Expression levels in B73 of genes encoding enzymes with predicted phospholipase A1 activity across different tissues. *HPC1* is indicated in blue. Data from (52). **B)** Expression levels from B73 leaves for genes in the 7.9–10 Mb QTL interval of chromosome 3. Data from (52). **C)** *HPC1* expression levels in the temperate inbred lines B73, Mo17, OH43, and PH207 under cold, control and heat stress conditions. Values taken from (66)

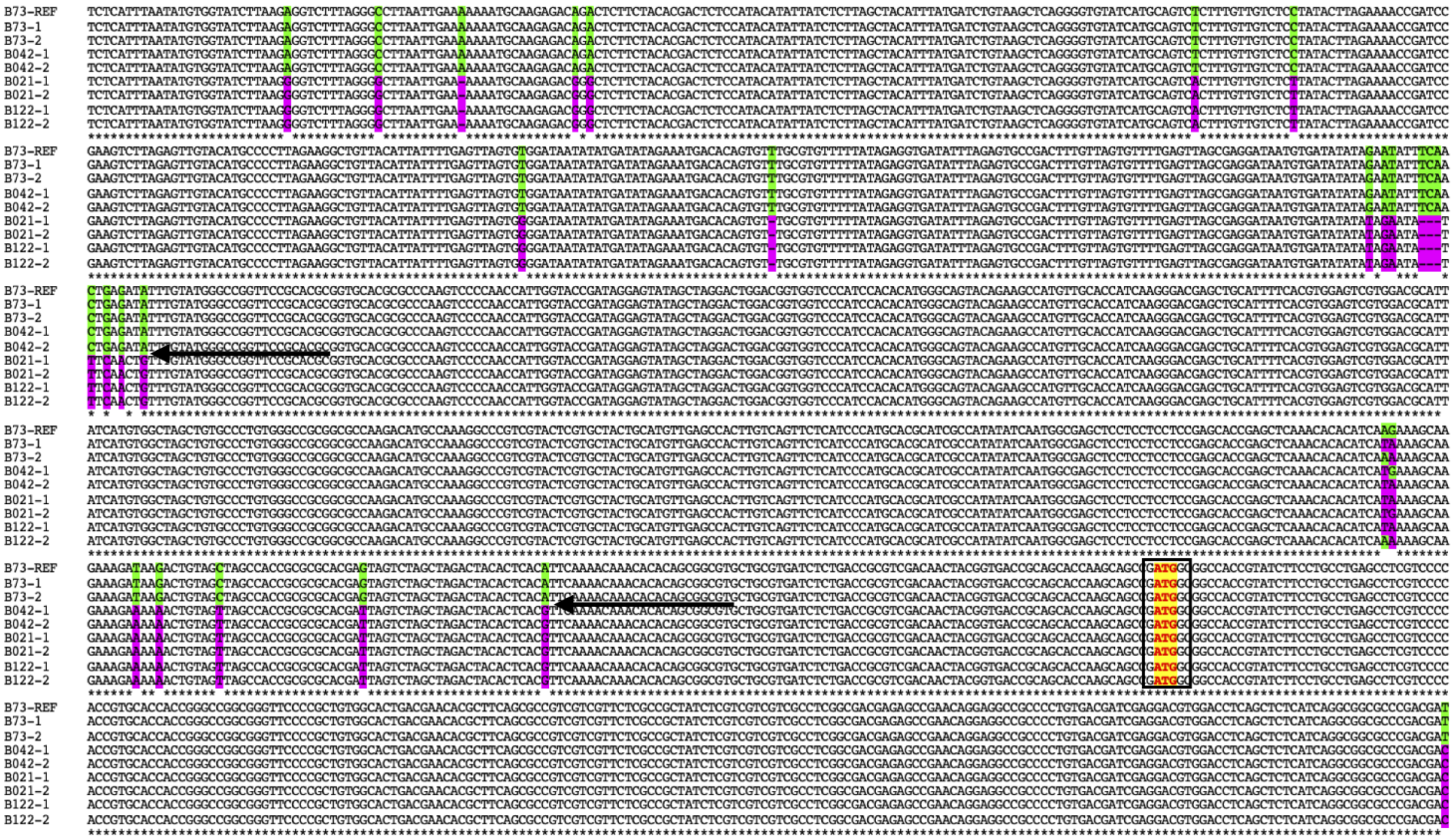

Figure Supplementary Figure 6 Sanger sequencing of the promoter and start of the *HPC1* genomic region obtained from B73 plants and 3 RILs (B042, B021 and B122). A recombination point 500 bp upstream the translation start codon in B1042 is indicated by arrows. B73 alleles are marked in green and PT alleles are marked in pink.

```
B73      MAATVSSCLSLVPTVHHRAGGFPAVALTNTLQRRRRSRRYLVVVASATRAEQEAAPVTIE    60
PT       MAATVSSCLSLVPTVHHRAGGFPAVALTNTLQRRRRSRRYLVVVASATRAEQEAAPVTIE    60
         ************************************************************

B73      DVDLSSHQAAPDDGELAARWPEIHGSNNWEGLLDPIDGVLLQELIRYGEFAQATYDSFDY    120
PT       DVDLSSHQAAPDDGELAARWPEIHGSNNWEGLLDPIDGVLLQELIRYGEFAQATYDSFDY    120
         ************************************************************

B73      DRFSPYCGSCKYPAKTFFHDVGLGGIGYEVTRYLYATCNDLKFPNFGIKTAANAKMWSES    180
PT       DRFSPYCGSCKYPAKNFFHDVGLGGIGYEVTRYLYATCNDLKFPNFGIKTAANTKMWSES    180
         ***************.*************************************:******
                                        ⬇
B73      GTFIGYVAVSTDEETARLGRRDIAVAWRGTITRLEWVADLTANQIPLRETGVPCPDPDVK    240
PT       GTFIRYVAVSTDEETARLGRRDITVAWRGTVTRLEWVADLTANQIPLRETGVPCPDPDVK    240
         **** ******************:******:*****************************

B73      VERGFVALYTDKGTGCRFCRYSAREQVLAEVRKLVDLYHGRGEQVSVTVTGHSLGSALAM    300
PT       VERGFVALYTGKGTGCRFCRYSAREQVLAEVRKLVDLYHERGEQVSVTVTGHSLGSALAM    300
         **********.**************************** *******************

B73      LCAFDIAETRANVSPGDRVAPVCVFSFAGPRVGNVAFRRRFERELGVRALRVVNVHDSVP    360
PT       LCAFDIAETRANVSPGDRVAPVCVFSFAGPRVGNVAFRRRFEQELGVRALRVVNVHDSVP    360
         ******************************************:*****************

B73      KVPGVFFNESAFPELVLRAADRLGLGGVYTHLGVLLQLDHKVSPFLKETLDLSCYHNLEA    420
PT       KVPGVFFNESAFPELVLRAADRLGLGGVYTHLGVLLQLDHKASPFLKEALDLSCYHNLEA    420
         ******************************************.******:***********

B73      HLHLLDGFRGSGAGFEPRGRDPALVNKSTDFLREDHMVPPVWYQAENKGMVRTEDGRWVL    480
PT       HLHLLDGFRGSGAGFELRRRDPALVNKAADFLREDHMVPPAWYQPENKGMVRTEDGRWVL    480
         **************** * ********::***********.*** ***************

B73      PPRQRVLDDHPEDTDHHLQRLGLTA*  505
PT       PPRQRVLDDHPEDTDHHLQRLGLTA*  505
         *************************
```

Figure Supplementary Figure 7 B73 and PT HPC1 proteins aligned using ClustalOmega. Green residues are those from the chloroplast transit peptide. Blue resides are part of the Lipase3 domain. Magenta residues are the flap lid. Arrow denotes I211V mutation. "*" denotes matching sequence, ":" denotes conservation between groups with similar properties, "." denotes conservation of groups with weakly similar properties.

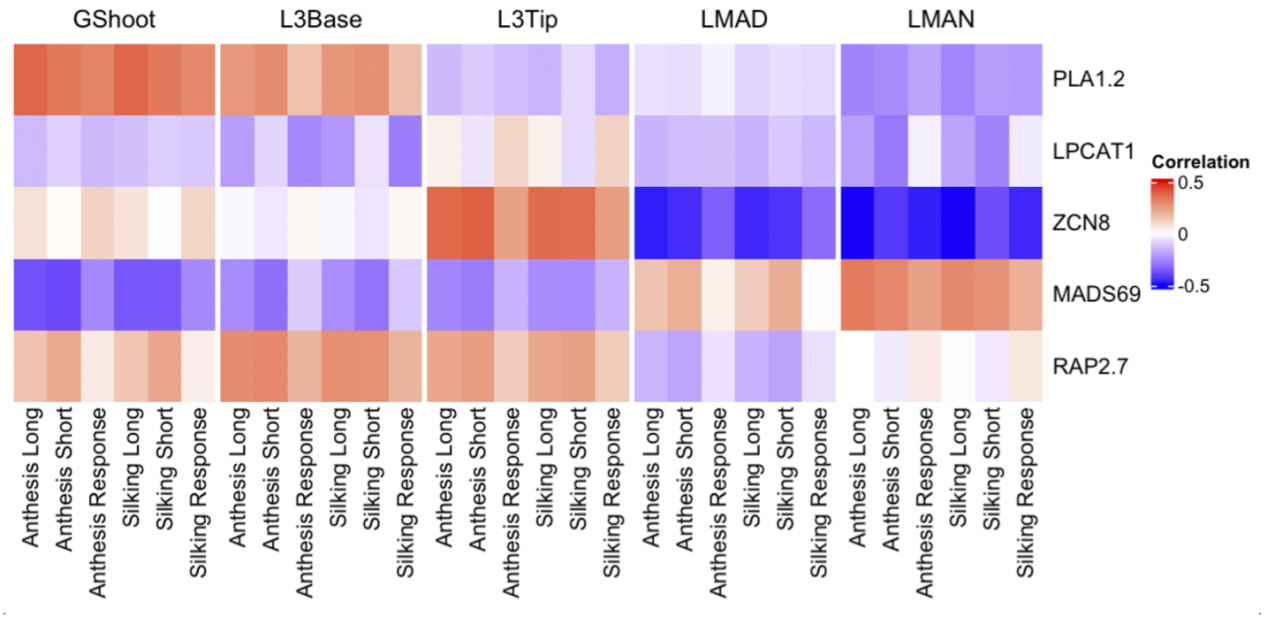

Figure Supplementary Figure 8 Correlation between flowering time and expression of phospholipid-related genes, using flowering time traits from aerial tissues in the 282 diversity panel. Data obtained from (57)