

---

# Assessing Population Genetics Evidence for Local Adaptation to Phosphorus Availability in Maize.

## Doctoral Project Proposal

Fausto Villafrade Rodríguez Zapata. M.Sc.

### Thesis directors:

Dr. Andrés Moreno Estrada.  
Dr. Ruben Rellán Álvarez.

### Committee:

Dr. Alexander de Luna  
Dr. Luis Delaeye Arredondo  
Dr. Sean Michael Rovito  
Dr. Tania Hernández Hernández

## Abstract

Plant phosphorus starving response (PSR) is a coordinated set of biochemical, physiological and developmental reactions to low phosphorus supply. I expect divergent selection in the loci involved in PSR if it is the result of local adaptation. Given the absence of opposing evolutionary forces, the main consequence of this selection is an increased frequency of adaptive alleles in populations exposed to low phosphorus availability. Here I use a reverse ecology approach to identify loci that might be involved in PSR and be subject to divergent selection. First I built soilP, an R package for assigning soil phosphorus retention potential to geographic locations. With this retention potential as phenotype I performed environmental GWAS on 3238 georeferenced landraces of *Zea mays* from Latin America and the Caribbean. This resulted in the detection of significant signal from the 13 Mb span of *Inv4m*, a previously reported adaptive retrogression from highland teosinte that includes *ZmPho1;2a* an inorganic phosphate transporter. Finally, I propose analysis of phosphorus response in biparental populations polymorphic for *Inv4m* in order to disentangle population structure from local adaptation or other environmental covariates.

## Introduction

**Phosphorus as limiting nutrient.** Different soil phosphorus compounds can be sorted into separate pools according their availability for direct plant use. The most readily available form for plants is Pi in aqueous solution. Next, is the phosphorus in organic compounds, such as nucleotides, phospholipids, phosphorylated sugars and proteins, from which phosphates can be hydrolyzed by enzymatic activity. After this, the non-occluded P fraction that corresponds to phosphate adsorbed to the surfaces of iron and aluminum oxides and CaCO<sub>3</sub> [41]. One step less available is the occluded P fraction, that is, phosphate physically encapsulated or surrounded by secondary minerals, such as Fe, Al and Mn oxyhydroxides [16,42] by coprecipitation or diffusive penetration [41]. Thus the labile, i.e. more bioavailable [42], phosphorus pool includes

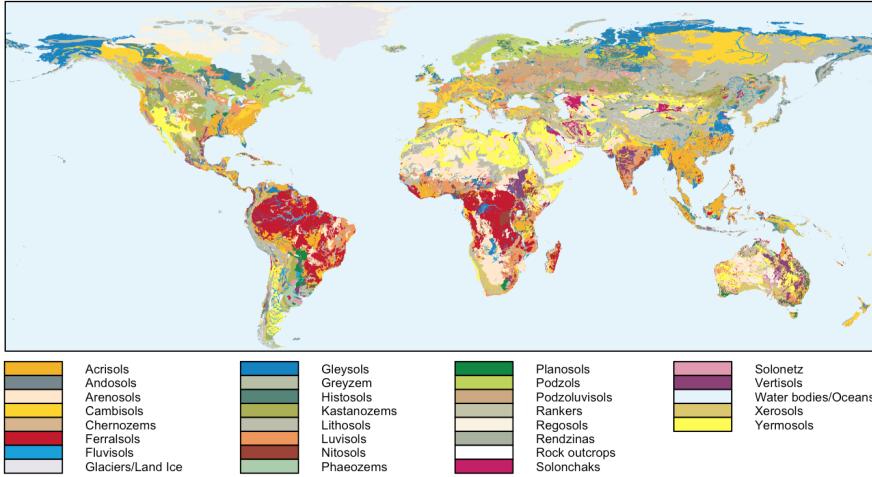
phosphate ions in solution as well as P incorporated in soil organic matter, and to a lesser degree the non-occluded P fraction. The refractory forms, which are not easily bioavailable, include P in non-weathered apatite minerals and the occluded P fraction [16]. In experimental conditions added normal or high Pi concentrations in nutrient solution range from 15  $\mu$ M to 0.1 mM, depending on the subject species [36]. However lower realized concentrations are expected given phosphate precipitation in presence of calcium ions, that are usually part of the solution. Pi concentrations 0.3 to 5  $\mu$ M in soil solution, could be considered low if they do not support optimal plant growth [35, 36]. Unfertilized soils rarely release P fast enough to support the high growth rates of crop plant species [35].

**Phosphate Starving Response (PSR).** When confronted with Pi scarcity, plants present biochemical, physiological and developmental adjustments, collectively known as phosphate starving response , that can improve aquisition, reallocation, mobilization and, in general, phosphorus use efficiency [36]. Pi starvation results in a shift in the developmental program that can increase the fitness of the plant [28]. There are maize landraces in the Mexican highlands that show PSR response in field conditions. In the Purhepecha Plateau, Michoacán [5, 6], local maize landraces showed adaptation to a phosphorus deficient Andisol in traits such as root length, lateral root number, root hairs, and phosphorus use efficiency defined lower inhibition of shoot growth under low P. Cloned genes intervening in PSR and candidates from maize, sorghum and *Arabidopsis* are reviewed in [8] with funtions such as phosphate ion transport, transcription factors, micro-rna regulation, and lipid metabolism.

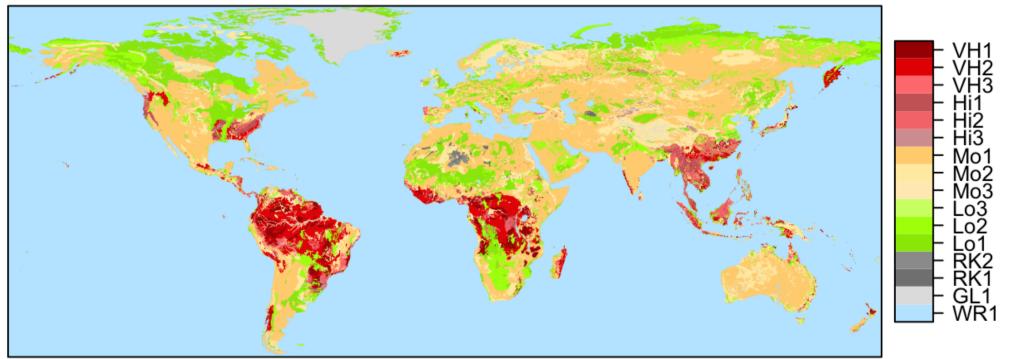
**Local Adaptation.** The classical experimental design to prove local adaptation is the reciprocal transplant. In this design, the strict criterion for local adaptation is that a population must have higher fitness at its native site than any other population introduced to that site [34]. Local adaptation results in higher fitness of a population in a local environment, relative to populations coming from different habitats [23]. It is both the observed pattern and the process dependent on divergent selection, the allele frequency change in the populations in response to selection that varies geographically [39]. Evolutionary forces against local adaptation include gene flow, migration and mutation, because they act contrary to genetic differentiation of the populations. In contrast genetic drift causes differentiation in allele frequencies that is not necessarily adaptive, and is thus a confounding factor when studying the genetic basis of local adaptation [23].

**Phosphorus Retention Potential Map.** Since the time of domestication, around 8700 B.P [18], the adaptation of maize to local soil conditions has mostly taken place before the widespread adoption of industrial fertilizers. In contrast to the current high input crop system, where low phosphorus availability is readily supplemented, preindustrial fertilization practices were far less effective if existent. Therefore, with low or no added phosphates to soil, phosphorus retention potential must have been a an even more limiting parameter for plant growth in preindustrial times. In this work I used retention potential from the Global Phosphorus Retention Potential Map [4] as an indicator of soil phosphorus availability instead of soil phosphorus content because datasets of estimated phosphorus content have serious data gaps in Latin America and the Caribbean and/or are likely to include modern fertilizer input [17, 33, 38, 42]. The base raster for this map is derived from the Digital Soil Map of the World (1:5 million scale) [13, 14], and has 10 km resolution with a total of 4320 x 1686 pixels. The raster value for each pixel encodes a map unit identifier. A raster attribute table links each map unit with probabilities for 8 possible soil units. The resulting map of the dominant,

i.e. most probable, soil units can be seen in Figure 1 (a). An additional table stores the physicochemical properties per soil unit, and a corresponding column for phosphorus retention potential class assigned by Batjes based upon soil unit pH and mineralogy. There are 4 possible values for this phosphorus retention potential class: low (Lo), moderate (Mo), high (Hi), very high(VH). Phosphorus retention potential class probabilities are then defined as the sum of soil unit probabilities per phosphorus retention class in each map unit 1. The resulting probabilities are binned according to its dominant class, i.e. 75-100 %, 50-75 % and 25-50 %, in order to obtain a summary ordinal scale of phosphorus retention score (Figure 1 (b)) that can be coded as a number from 1 to 15.

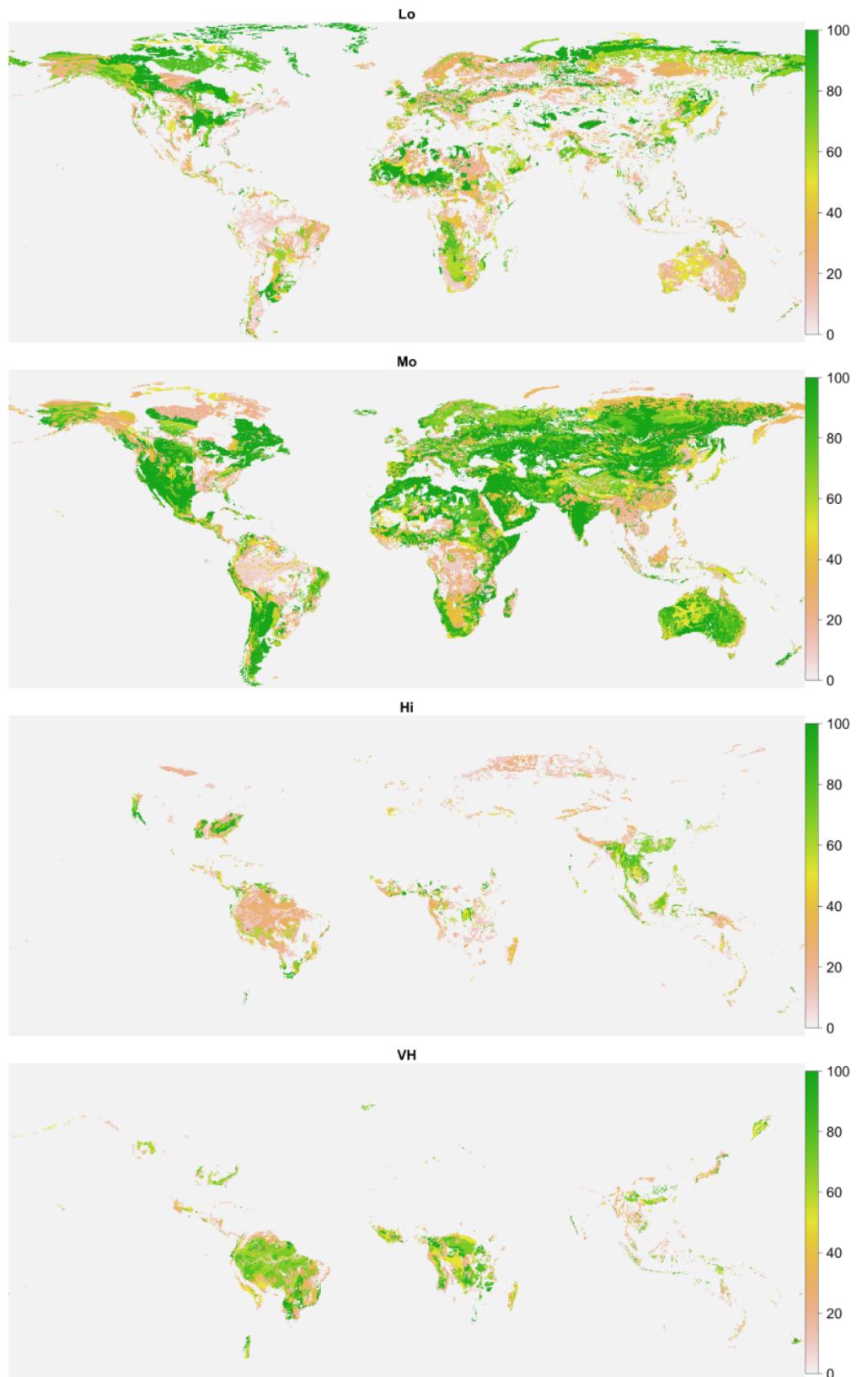


(a) Digital Soil Map of the World. Dominant soil groups [14] legend based upon [13, 15].



(b) Global Phosphorus Retention Potential Map. Retention classes are described by letters, VH: Very High; Hi: High; Mo: Moderate; Lo: Low; RK: Rock Outcrop; GL: Glacier; WR: Water. Class probability bins are represented by numerals, 1: 75-100 %; 2: 50-75 %; 3: 25-50 %. Modified from [4].

Figure 1. Soil class defines phosphorus retention.



**Figure 2. Soil Phosphorus Retention Class Probability Map.** Each channel represents one of the 4 retention classes, VH: Very High; Hi: High; Mo: Moderate; Lo: Low. The same data is summarized in Figure 2b. These probabilities were used as phenotypes for GWAS analysis on the maize landraces of [31]

# Objectives

**Main Goal** To assess the population genetics evidence for local adaptation to soil phosphorus availability in maize, and to dissect the genetic architecture of such adaptation.

## Specific objectives

1. Find candidate loci for local adaptation using environmental GWAS in landrace populations.
2. Find evidence of selection in candidate loci between landrace populations.
3. Disentangle the confounding effects of environmental correlates and population structure through QTL analysis of phosphorus response in a biparental cross.

# Methods

## Environmental GWAS

The GBS genotypes and geographic coordinates of 4711 maize landraces from Latin America and the Caribbean [31] are the base data for this work. I started from the set of biallelic imputed SNPs, B73 RefGen\_V2 as reference genome, having 1584871 markers. Then I selected the 'GWAS dataset' as the 503016 markers with minor allele frequency greater than 0.01 in the 3238 georeferenced unrelated accessions (see below).

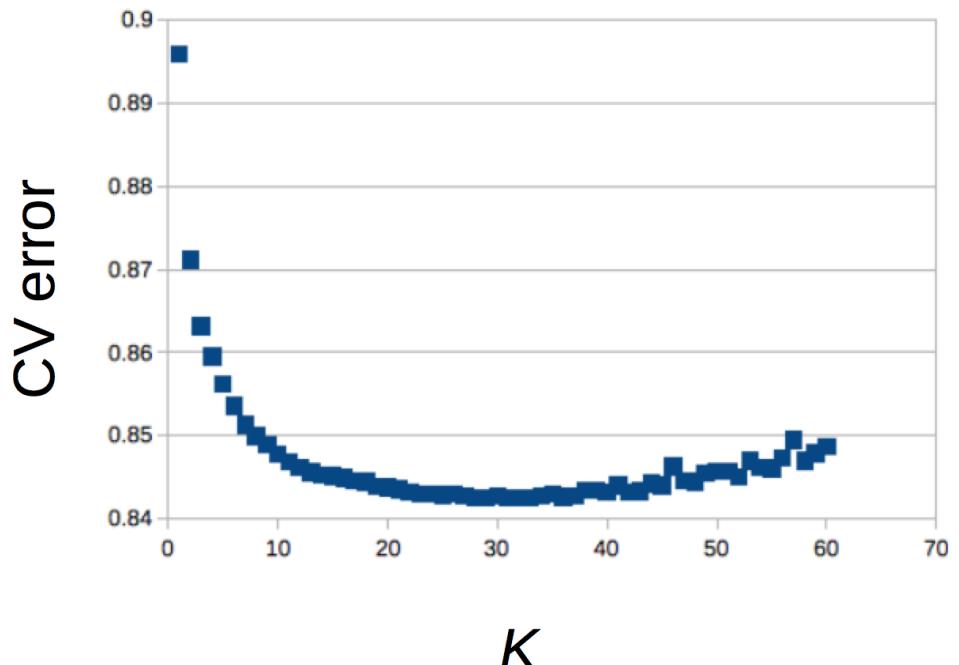
The idea of a genome environment association is using environmental variables instead of plant phenotypes as response variables [24]. Here I use the soil phosphorus retention potential, both, as class probabilities and as summarizing scale from [4]. With this aim I built `soilP` [30], an R package for extracting values from the Global Phosphorus Retention Potential Map and assigning them to arbitrary geographic locations. The `soilP` package and the detailed phosphorus variable extraction procedures are available at <https://github.com/sawers-rellan-labs/soilP>.

In addition to the phosphorus variables, I ran GWAS with selected environmental factors likely correlated to soil phosphorus retention potential. These comprise altitude and latitude [31], topsoil pH, clay and sand content [19], and precipitation of the warmest quarter [22]. All environmental variables were extracted at the landrace coordinates from the corresponding 250 m or 1 km resolution geotiff using the R package `raster` for subsequent analysis [20, 29]

In order to correct for population structure I calculated the ancestry matrix  $\mathbf{Q}$  for different number of postulated sub populations  $K$  using the model based estimation from ADMIXTURE [3]. However ADMIXTURE assumes that genotypes come from unrelated organisms and unlinked markers, hereafter named the 'structure dataset'. I decided to start the search for such a maximal independent set from a high quality matrix consisting of 6456 unimputed markers in 4511 individuals. Using `vcftools` [10] I estimated both the plant relatedness coefficient  $\phi$  according to the Manichaikul [26] method, and marker linkage disequilibrium through  $r^2$ . This allowed me to use `fastindep` [1] to arrive at a 'structure dataset' comprising 4368 unrelated plants at  $\phi < 1 \times 10^{-6}$  and 4908 unlinked SNPs at  $r^2 < 0.1$ . Once this dataset fulfilled ADMIXTURE's independence assumptions I estimated  $\mathbf{Q}$  for  $K = \{1 \dots 50\}$  using its unsupervised method. Instead of using the Evanno  $\Delta K$  method [12] for estimating the optimal number of subpopulations  $K$  I used ADMIXTURE's cross-validation criterion that requires less computational effort as it replaces complete dataset runs with runs of partially masked data (folds) [3, 25]. In this way I found 28 to be the optimal  $K$ ,

minimizing the cross validation error calculated from 5 fold runs 3. Then through  
 ADMIXTURE's supervised method I projected the resulting ancestry on the individuals  
 excluded from the 'structure dataset' for  $K = \{3, 10, 28\}$  spanning a range up to the  
 maximum reasonable  $K$ . After estimating ancestry for all the original 4711 landraces I  
 finnally subsetted the  $\mathbf{Q}$  matrix corresponding to the georeferenced unrelated common  
 variants or 'GWAS dataset'.  
116  
117  
118  
119  
120  
121

I ran the association studies as GLM models for each environmental variable  
 dependent on landrace genotypes with command line TASSEL 5 [7] using custom made  
 shell scripts. As I constructed the 'GWAS dataset' from unrelated individuals I had no  
 need for correcting the GLM model with a kinship matrix  $\mathbf{K}$ . However I did correct for  
 $\{\mathbf{Q}_0, \mathbf{Q}_3, \mathbf{Q}_{10}, \mathbf{Q}_{28}\}$  corresponding to no correction, and correction given the selected  
 subpopulation numbers  $K = \{3, 10, 28\}$  Figure 8 (a-c).  
122  
123  
124  
125  
126  
127



**Figure 3. Cross validation of  $\mathbf{K}$ .** Optimal  $K$  is found to be 28 from 1 run of ADMIXTURE 5 fold Cv on 4368 unrelated plants and 4908 unlinked SNPs.

## Selection Scan

128

I expect divergent selection in the loci involved in PSR if it is the result of local  
 adaptation. Given the absence of opposing evolutionary forces, the main consequence of  
 this selection is an increased frequency of adaptive alleles in populations exposed to low  
 phosphorus availability. Therefore in order to use this pairwise difference of allele  
 frequencies it is necessary to previously define the populations subject to comparison.  
 From here on I will restrict the analysis to Mexican landraces and will be comparing  
 VH(1-3) (1500 accessions) vs Mo(1-3) (507 accessions) because the Lo(1-3) landraces are  
 a smaller set (103 accessions) and are geographically more distant from central México  
 than Mo accessions. However the caveat that highland landraces are more frequently  
 exposed to VH soils than lowland landraces ( $r = 0.39$ ) persists so any difference in allele  
 frequencies can be due to either environmental factor. Fst and its significance would be  
 calculated using `hierfstat` [11]. The XP-CLR statistic will be calculated using both  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140

groups as reference [9], and apropiate parameters for windows size an LD  $r^2$  would be  
141 chosen as compromize between precision and computational time. The AGPv2  
142 consensus genetic map with 7184 markers will be used as the framework for estimating  
143 the SNPs genetic distance as a required input for XP-CLR.  
144  
[https://ftp.maizegdb.org/MaizeGDB/FTP/B73\\_RefGen\\_v2.dumps/](https://ftp.maizegdb.org/MaizeGDB/FTP/B73_RefGen_v2.dumps/).  
145

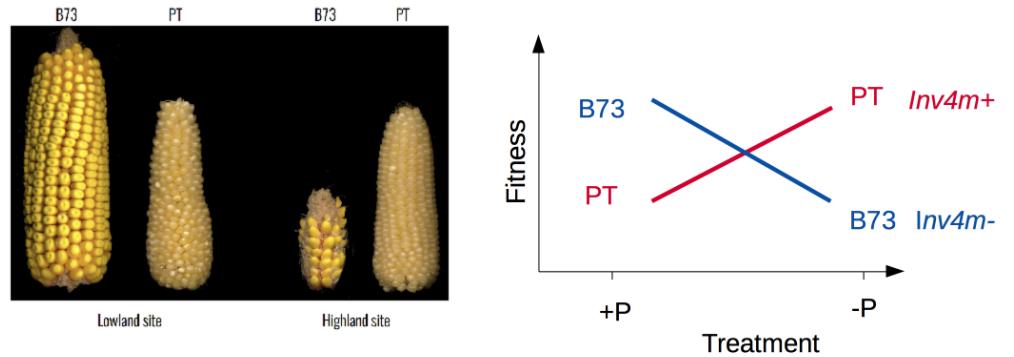
## Biparental Populations Analysis

In order to clarify the role of *Inv4m* in the adaptation to low soil phosphorus availability  
147 I can make use of biparental populations derived from B73 X Palomero Toluqueño  
148 crosses. B73 is a temperate inbred that has been used as the reference genome for  
149 maize [37], while PT is a highland adapted landrace [40]. Having contrasting  
150 phenotypes they can be used for generating segregating populations.  
151

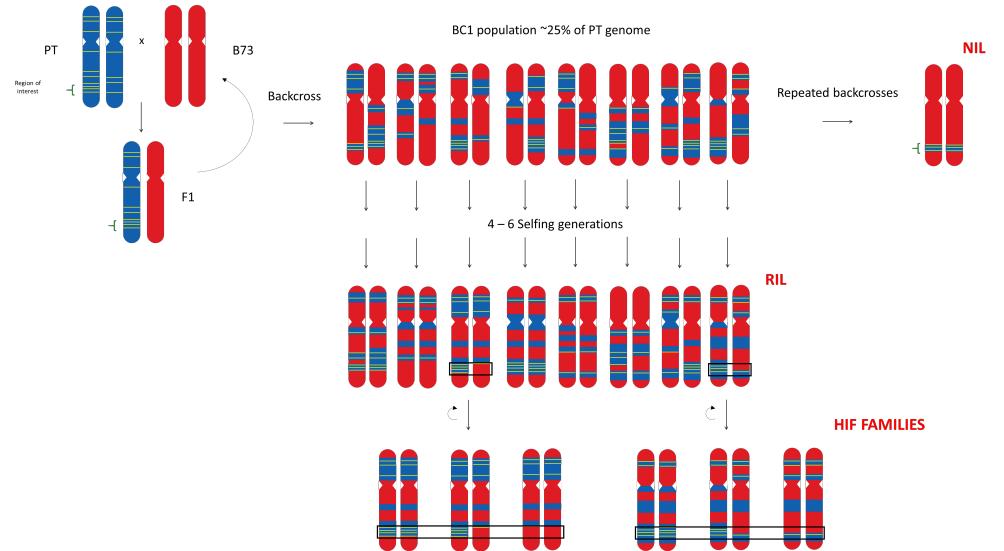
If *Inv4m* has any role in the adaptation to low phosphorus it's introgression into a  
152 B73 genetic background could in principle improve the inbred line phosphorus starving  
153 response. Segregating introgression lines can be derived from a B73 X PT F1 through  
154 different strategies 5. The more intuitive ones though successive back crosses, in each  
155 generation diluting the introgressing genome by half. In another method, after freezing  
156 the BC1 mosaic though several generations of selfing heterozygotes at the locus of  
157 interest are genotyped and kept so that whenever the breeder decides to autopollinate,  
158 it segregates. This method is called Heterologous Inbred Families (HIFs), each HIF  
159 having a distinct mosaic but segregating at the selected locus.  
160

Both NIls and HIFs have segregation either 1:1 or 1:2:1. Therefore the phenotyping  
161 experiment can only be planned over the expected values of segregation in the seed as  
162 is described in 2a. For the HIF strategy I made a power analysis of an unbalanced  
163 ANOVA for differences in phenotypes between the three segregating classes using  
164 custom R scripts. So for a coefficient of variation in each category phenotype of 12.5%  
165 and a significance of 0.05 with a real difference between category means of 10% the  
166 column n number of samples would have 90% of power and is logically feasible. With  
167 a coefficient of variation of 25% the same set samples would have just over 50% of  
168 power. The +P-P column lists the n appropriate for a two way ANOVA, duplicate an  
169 triplicate columns represent the amount of material needed in order to find differences  
170 not, between genotypic classes but between individual lines.  
171

The phenotyping experiment for phosphorus response will be set up at the  
172 LANGEBIO greenhouse at the start of the growing season (April). 3 seeds per line will  
173 be sowed into 3L pots with washed and sterilized sand with a complete randomized  
174 design. After germination only one plant per pot will be allowed to grow. The substrate  
175 will be watered every 3 days with 200 mL of +P (330  $\mu$ M) or -P (0  $\mu$ M) modified  
176 Hoagland solution [21,32]. At 25 days after emergence measurements of shoot length,  
177 number of leaves, root length root/shoot ratio will be taken and tissue for genotyping  
178 and metabolomics will be sampled, flash frozen, and preserved at -80°C. Phosphorus  
179 Ionomics will be carried out with ICP-MS by Ivan Baxter, in the Missouri Danforth  
180 center.  
181



**Figure 4. Contrasting phenotypes of B73 and PT.** Left: Actual phenotype of B73 and PT in high and low altitude (photo Rubén Rellán). Right: Expected reaction curves of B73 and PT if *Inv4m* plays a crucial role in local adaptation to low phosphorus availability through a genetic trade-off.



**Figure 5. Crossing scheme for genetic analysis of B73 and PT.** Reproduced from [2]

**Table 1.** CAPS marker at the *inv4m*, modified from [2]

Properties	
Oligo code	RS 427/428
SNP id	PZE104103103
Position (bp)	179617762
Primer forward	CTGAGCAGGAGATGATGGCCACTC
Primer reverse	GGAAAGGACATAAAAGAAAGGTGCA
5' context (50 bp)	aagaagatgcatgaatggtttgcacgtacacttgcttagggcgtagcga
B73 allele (ref)	A
PT allele (alt)	G
3' context (30 bp)	attcttcttcttggcccttgtgaattccgaattctcgacgattgattctg
Amplicon size (bp)	502
Restriction Enzyme	<i>Hinf</i> I
B73 fragments (bp)	166; 44; 251; 61
PT fragments (bp)	166; 295; 61

**Table 2.** Plant material requirements for HIF design.

genotype	n	+P/-P	Duplicate	Triplet
inv+	25	50	100	150
inv+/inv-	50	100	200	300
inv-	25	50	100	200
B73	20	40	80	160
PT	20	40	80	160
	pots	280	560	650
	seed	840	1680	1950

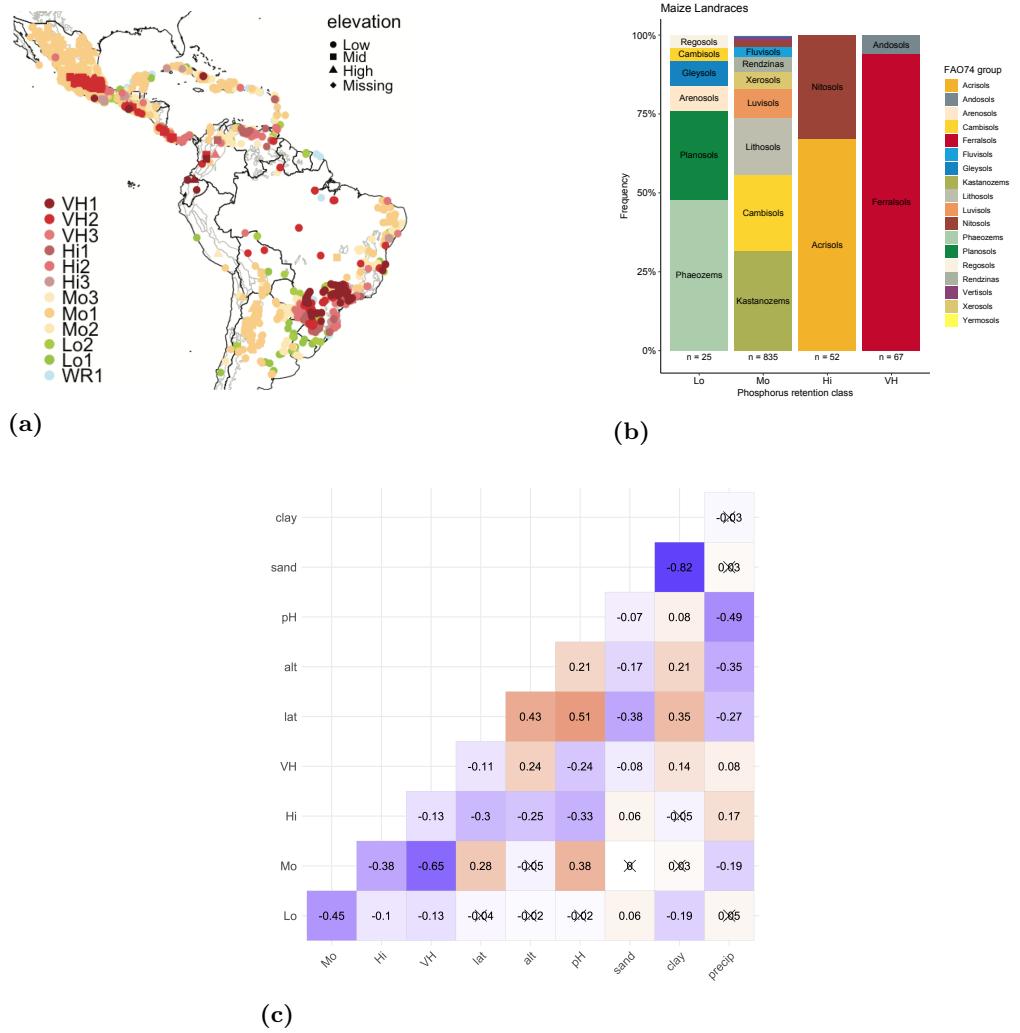
(a) HIF Lines

genotype	n	+P/-P	Duplicate	Triplet
inv+/inv-	25	50	100	150
inv-	20	40	80	120
B73	10	20	40	60
PT	10	20	40	60
	pots	280	260	270
	seed	840	780	810

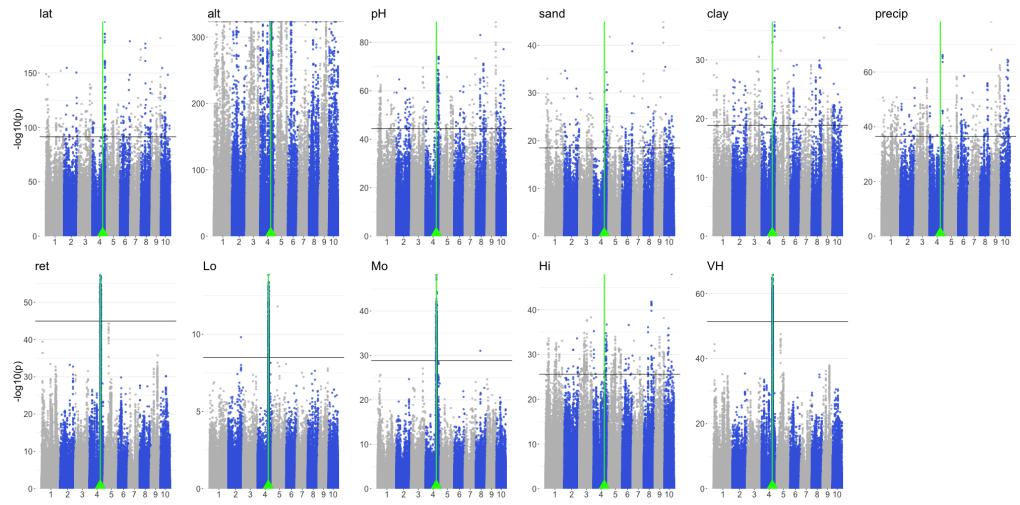
(b) NIL Lines

# Environmental GWAS Results

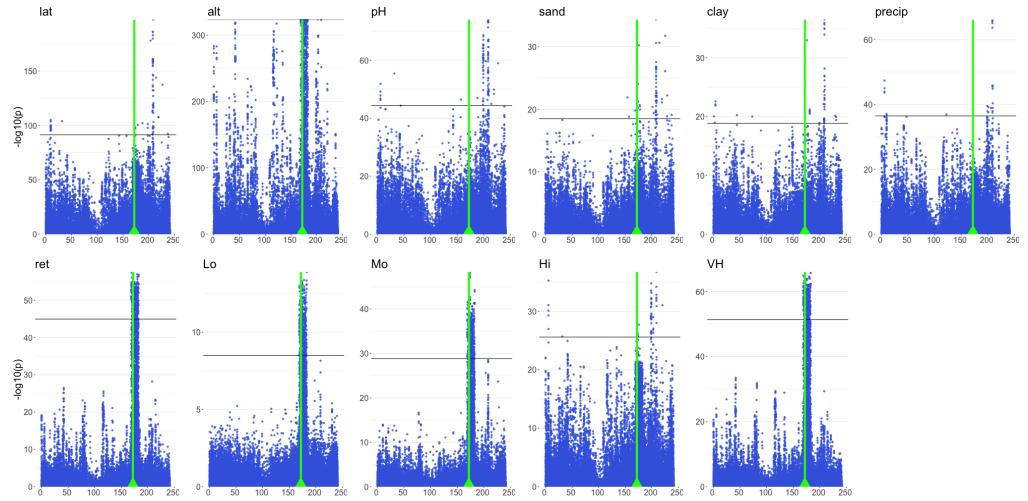
182



**Figure 6. Phosphorus retention potential for LAC landraces, soil distribution and correlation with selected confounding variables.** (a) Distribution map of 3238 geolocated landraces from [31], most of landraces were collected from sites with moderate phosphorus retention potential. There is a lack of sampling in the Northern and Central Andes. (b) Soil distribution among phosphorus retention potential classes. The counted landraces (n) are those that happen to be in locations with 100% probability for their phosphorus retention class. VH soils are restricted to andosols and ferralsols. Hi soils are limited to acrisols and nitosols. The moderate class is the most abundant and most eclectic, but dominated by kastanozems, cambisols and lithosols. The Lo class is dominated by phaeozems and planosols. (c) Altitude correlates moderately with VH and Hi retention potential which can be due to a higher proportion of andosols in the Mexican transvolcanic belt, and a higher proportion of acrisols in southern Brazil and Venezuelan northeast lowlands. Non significant correlations are marked with a cross (t test, Bonferroni corrected  $p < 0.05/36$ ).

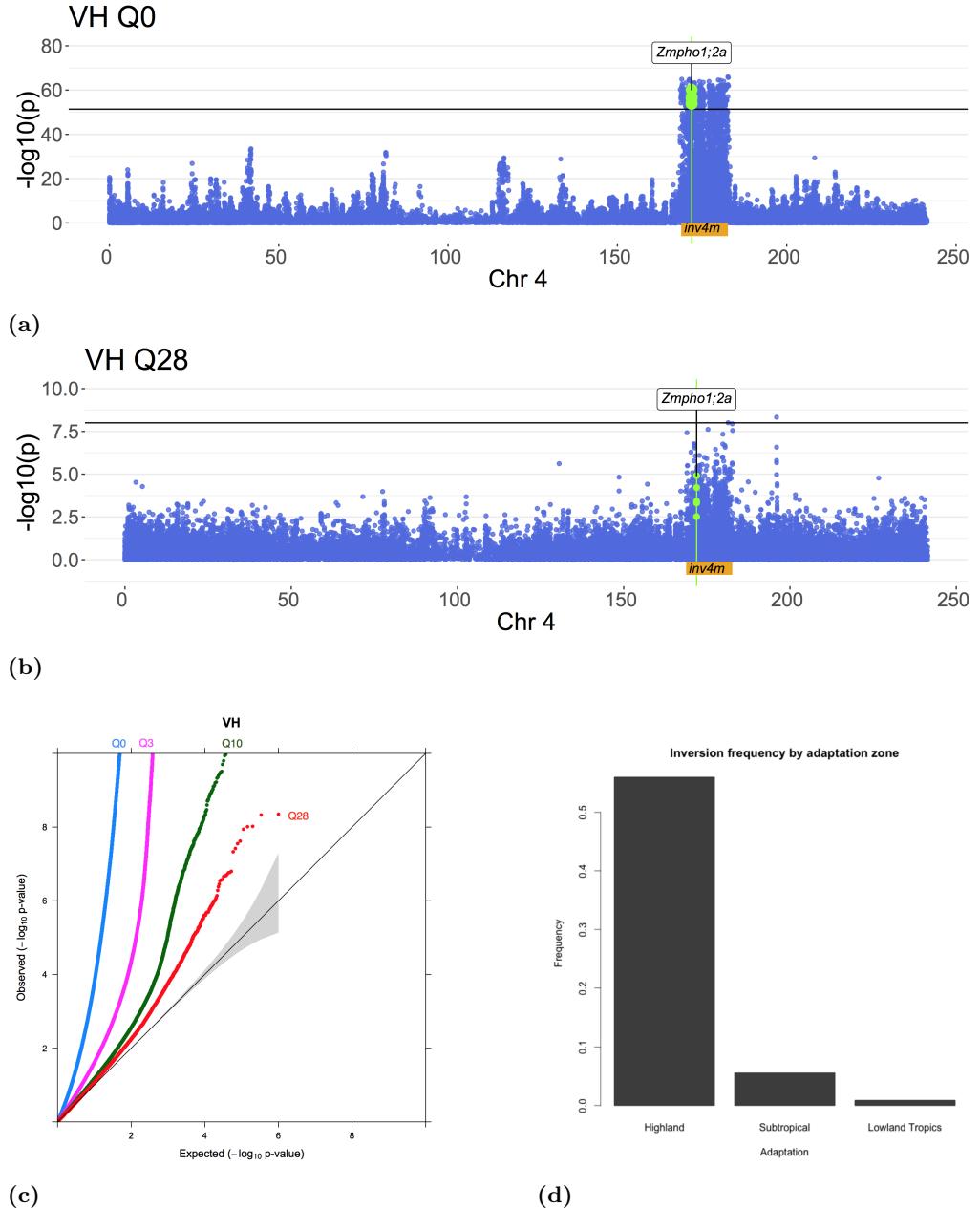


**(a) Genomewide.** For the phosphorus retention variables (bottom panel) the strongest association is observed in Chromosome 4 near 172 Mb (green triangle). A GWAS peak is detected in the same location for altitude (alt, top panel).



**(b) Chromosome 4.** A zoom into this chromosome shows that the signal around 172 Mb is characteristic of altitude and the phosphorus retention variables except Hi. Despite the significant correlations observed in Figure 6, latitude, precipitation, and the other 3 soil variables do not show as strong association in this region.

**Figure 7. GLM GWAS analysis.** Black horizontal bar corresponds to 0.001 quantile of p-values.



**Figure 8. GWAS main signal is lost with population structure correction and corresponds to *Inv4m*, an inversion mostly present in highland landraces.** VH is used as phenotype for illustration. Population structure was accounted for with  $Q$ , the ancestry coefficient matrix of  $K$  subpopulations ( $K = 0, 3, 10, 28$  estimated with [3]), included as fixed effect in the GWAS GLM model. *Inv4m* is a putative adaptive introgression from highland teosinte spanning 13 Mb, (orange bar) [27], (a) At Q0, no ancestry matrix included, there are 4 significant snps (green dots) in the phosphate transporter *ZmPho1;2a* [32] (green line at 172 Mb), black line 0.001 p-value quantile. (b) Most of the significance is lost at Q28, black line p-value = 1e-8. (c) GWAS QQ-plot, p-values are far more significant than expected (over the diagonal) for any  $Q$ . Increasing the number of postulated subpopulations diminishes p-value inflation but even at Q28 they deviate from the 95% CI (grey). (d) *Inv4m* is mostly present in highland landraces (2000 m.a.s.l) [31].

## Schedule

TASK	SEMESTER									
	1	2	3	4	5	6	7	8	9	10
GWAS analysis	X	X								
Selection Scan	X	X	X							
NIL Phenotyping				X	X					
NIL Genotyping				X	X					
Paper writing						X	X			
Final document writing								X	X	

Figure 9. Schedule

## References

1. K. J. Abraham and C. Diaz. Identifying large sets of unrelated individuals and unrelated markers. *Source Code for Biology and Medicine*, 9(1):6, Mar. 2014.
2. M. R. Aguilar Rangel. *The role of introgression from teosinte (*Zea mays ssp. mexicana*) in the adaptation of maize to the highlands of Mexico*. Ph.D., CINVESTAV Unidad Irapuato, Irapuato, 2018.
3. D. H. Alexander, J. Novembre, and K. Lange. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9):1655–1664, Sept. 2009.
4. N. H. Batjes. Global distribution of soil phosphorus retention potential. Technical report, ISRIC-World Soil Information, Wageningen, 2011. Data available at [https://files.isric.org/public/other/Global\\_distribution\\_of\\_soil\\_phosphorus\\_retention\\_potential](https://files.isric.org/public/other/Global_distribution_of_soil_phosphorus_retention_potential)
5. J. S. Bayuelo-Jiménez, M. Gallardo-Valdés, V. A. Pérez-Decelis, L. Magdaleno-Armas, I. Ochoa, and J. P. Lynch. Genotypic variation for root traits of maize (*Zea mays L.*) from the Purhepecha Plateau under contrasting phosphorus availability. *Field Crops Research*, 121(3):350–362, Apr. 2011.
6. J. S. Bayuelo-Jiménez and I. Ochoa-Cadavid. Phosphorus acquisition and internal utilization efficiency among maize landraces from the central Mexican highlands. *Field Crops Research*, 156:123–134, Feb. 2014.

- 
7. P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635, Oct. 2007.
  8. C. Calderon-Vazquez, R. J. H. Sawers, and L. Herrera-Estrella. Phosphate Deprivation in Maize: Genetics and Genomics. *PLANT PHYSIOLOGY*, 156(3):1067–1077, July 2011.
  9. H. Chen, N. Patterson, and D. Reich. Population differentiation as a test for selective sweeps. *Genome Research*, 20(3):393–402, Mar. 2010.
  10. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, Aug. 2011.
  11. T. de Meeùs and J. Goudet. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution*, 7(6):731–735, Dec. 2007.
  12. G. Evanno, S. Regnaut, and J. Goudet. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8):2611–2620, July 2005.
  13. FAO and UNESCO. Soil Map of the World 1: 5000000, Vol. I, Legend. 1974.
  14. FAO and UNESCO. *Digital soil map of the world and derived soil properties*. FAO, Rome, Italy, [version 3.5] edition, 1995.
  15. FAO GIS and WRB. Dominant Soils of the World Map. Aug. 1999. Available at <https://chemistry.beloit.edu/Rain/pages/soil99.html>.
  16. G. M. Filipelli. The global phosphorus cycle. In B. A. S. Rattan Lal, editor, *Soil Phosphorus*. CRC Press, 2016.
  17. HarvestChoice. Updating Soil Functional Capacity Classification System. *International Food Policy Research Institute Washington, DC., and University of Minnesota, St. Paul*, Feb. 2010. Available online at <http://harvestchoice.org/node/1435>.
  18. C. A. Hastorf. Rio Balsas most likely region for maize domestication. *Proceedings of the National Academy of Sciences*, 106(13):4957–4958, Mar. 2009.
  19. T. Hengl, J. M. d. Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2):e0169748, Feb. 2017.
  20. R. J. Hijmans. *raster: Geographic Data Analysis and Modeling*. 2017.
  21. D. R. Hoagland and T. C. Broyer. GENERAL NATURE OF THE PROCESS OF SALT ACCUMULATION BY ROOTS WITH DESCRIPTION OF EXPERIMENTAL METHODS. *PLANT PHYSIOLOGY*, 11(3):471–507, July 1936.

- 
22. D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4:170122, Sept. 2017.
23. T. J. Kawecki and D. Ebert. Conceptual issues in local adaptation. *Ecology Letters*, 7(12):1225–1241, Dec. 2004.
24. J. R. Lasky, H. D. Upadhyaya, P. Ramu, S. Deshpande, C. T. Hash, J. Bonnette, T. E. Juenger, K. Hyma, C. Acharya, S. E. Mitchell, E. S. Buckler, Z. Brenton, S. Kresovich, and G. P. Morris. Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, 1(6):e1400218, July 2015.
25. D. J. Lawson, L. v. Dorp, and D. Falush. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1):3258, Aug. 2018.
26. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867–2873, Nov. 2010.
27. T. Pyhäjärvi, M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. Complex Patterns of Local Adaptation in Teosinte. *Genome Biology and Evolution*, 5(9):1594–1609, Sept. 2013.
28. B. Péret, M. Clément, L. Nussaume, and T. Desnos. Root developmental adaptation to phosphate starvation: better safe than sorry. *Trends in Plant Science*, 16(8):442–450, Aug. 2011.
29. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
30. F. Rodriguez-Zapata. *soilP: Local Adaption to Soil Phosphorus Availability*. 2018.
31. J. A. Romero Navarro, M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, A. Terrón, H. V. Delgado, V. Vidal, A. Ortega, A. E. Banda, N. O. G. Montiel, I. Ortiz-Monasterio, F. S. Vicente, A. G. Espinoza, G. Atlin, P. Wenzl, S. Hearne, and E. S. Buckler. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature Genetics*, 49(3):476–480, Mar. 2017.
32. M. N. Salazar-Vidal, E. Acosta-Segovia, N. Sánchez-León, K. R. Ahern, T. P. Brutnell, and R. J. H. Sawers. Characterization and Transposon Mutagenesis of the Maize (*Zea mays*) *Pho1* Gene Family. *PLOS ONE*, 11(9):e0161882, Sept. 2016.
33. P. A. Sanchez, C. A. Palm, and S. W. Buol. Fertility capability soil classification: a tool to help assess soil quality in the tropics. *Geoderma*, 114(3-4):157–185, June 2003.
34. O. Savolainen, M. Lascoux, and J. Merilä. Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11):807–820, Nov. 2013.
35. D. P. Schachtman, R. J. Reid, and S. Ayling. Phosphorus Uptake by Plants: From Soil to Cell. *Plant Physiology*, 116(2):447–453, Feb. 1998.

- 
36. W.-R. Scheible and M. Rojas-Triana. Sensing, signalling, and CONTROL of phosphate starvation in plants: molecular players and applications. In *Annual Plant Reviews Volume 48*, volume 115, page 23–63. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015.
37. P. S. Schnable, D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T. A. Graves, P. Minx, A. D. Reily, L. Courtney, S. S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S. M. Rock, E. Belter, F. Du, K. Kim, R. M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S. M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M. J. Levy, L. McMahan, P. Van Buren, M. W. Vaughn, K. Ying, C.-T. Yeh, S. J. Emrich, Y. Jia, A. Kalyanaraman, A.-P. Hsia, W. B. Barbazuk, R. S. Baucom, T. P. Brutnell, N. C. Carpita, C. Chaparro, J.-M. Chia, J.-M. Deragon, J. C. Estill, Y. Fu, J. A. Jeddeloh, Y. Han, H. Lee, P. Li, D. R. Lisch, S. Liu, Z. Liu, D. H. Nagel, M. C. McCann, P. SanMiguel, A. M. Myers, D. Nettleton, J. Nguyen, B. W. Penning, L. Ponnala, K. L. Schneider, D. C. Schwartz, A. Sharma, C. Soderlund, N. M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T. K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J. L. Bennetzen, R. K. Dawe, J. Jiang, N. Jiang, G. G. Presting, S. R. Wessler, S. Aluru, R. A. Martienssen, S. W. Clifton, W. R. McCombie, R. A. Wing, and R. K. Wilson. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*, 326(5956):1112–1115, Nov. 2009.
38. W. Shangguan, Y. Dai, Q. Duan, B. Liu, and H. Yuan. A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems*, 6(1):249–263, Mar. 2014.
39. P. Tiffin and J. Ross-Ibarra. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29(12):673–680, Dec. 2014.
40. J.-P. Vielle-Calzada, O. Martinez de la Vega, G. Hernandez-Guzman, E. Ibarra-Laclette, C. Alvarez-Mejia, J. C. Vega-Arreguin, B. Jimenez-Moraila, A. Fernandez-Cortes, G. Corona-Armenta, L. Herrera-Estrella, and A. Herrera-Estrella. The Palomero Genome Suggests Metal Effects on Domestication. *Science*, 326(5956):1078–1078, Nov. 2009.
41. T. W. Walker and J. K. Syers. The fate of phosphorus during pedogenesis. *Geoderma*, 15(1):1–19, Jan. 1976.
42. X. Yang, W. M. Post, P. E. Thornton, and A. Jain. The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences*, 10(4):2525–2537, Apr. 2013.