Local adaptation of maize landraces

Garrett Mitchell Janzen

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Genetics Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/17706

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Local adaptation of
maize landraces

by

Garrett Mitchell Janzen

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:
Matthew Hufford, Major Professor
Fredric Janzen
Peng Liu
John Nason
Amy Toth

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2019

Copyright © Garrett Mitchell Janzen, 2019. All rights reserved.
DEDICATION

To my mother and father, Lora and Danny,

who made this possible,

and to my family, Anna, Paul, and Ella,

who make this worth it.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>List</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>xii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER 1. GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Dissertation Organization</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Dissertation Objectives</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER 2. LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Theory</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Model System</td>
<td>5</td>
</tr>
<tr>
<td>2.3 References</td>
<td>10</td>
</tr>
<tr>
<td>CHAPTER 3. DEMONSTRATION OF LOCAL ADAPTATION OF MAIZE LANDRACES BY RECIPROCAL TRANSPLANTATION</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>18</td>
</tr>
<tr>
<td>3.3 Methods</td>
<td>21</td>
</tr>
<tr>
<td>3.3.1 Field Experiment Design</td>
<td>21</td>
</tr>
<tr>
<td>3.3.2 Phenotypic Data Collection</td>
<td>22</td>
</tr>
<tr>
<td>3.3.3 Statistical Analyses</td>
<td>25</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>26</td>
</tr>
<tr>
<td>3.4.1 Adjusted Fitness (FITplant and FITplantveg)</td>
<td>26</td>
</tr>
<tr>
<td>3.4.2 Family Traits (STD, PE, BRN, DTA, DTS, ASI)</td>
<td>27</td>
</tr>
<tr>
<td>3.4.3 Plant Size Traits (PH, EH, TL, TBN, EN)</td>
<td>28</td>
</tr>
<tr>
<td>3.4.4 Yield Traits (PM, LM, GPH, DM)</td>
<td>28</td>
</tr>
<tr>
<td>3.4.5 Water Use Efficiency (d13C)</td>
<td>29</td>
</tr>
<tr>
<td>3.4.6 Anthocyanin Pigmentation and Macrohair Density (P_INTsolid, P_INTspot, P_EXTsolid, P_EXTspot, M_DENsolid, M_DENmarg)</td>
<td>29</td>
</tr>
<tr>
<td>3.4.7 Principal Component Analyses</td>
<td>30</td>
</tr>
<tr>
<td>3.4.8 Pearson Correlation of Highland-Adaptive Traits</td>
<td>30</td>
</tr>
<tr>
<td>3.4.9 Finlay-Wilkinson Regression</td>
<td>31</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>Names and descriptions of all collected phenotypes.</td>
<td>23</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Reaction norm $p$-values</td>
<td>50</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Niche breadth and correlated values</td>
<td>115</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Predictor environmental dataset prediction with categorical fastStructure group membership ($k = 9$, Figure 4.19(a))</td>
<td>120</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Independent environmental dataset prediction with categorical fastStructure group membership ($k = 9$, Figure 4.19(b))</td>
<td>121</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Predictor environmental dataset prediction with Radial Basis function kernel ancestry ($k = 3$, Figure 4.19(c))</td>
<td>122</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Independent environmental dataset prediction with Radial Basis function kernel ancestry ($k = 3$, Figure 4.19(d))</td>
<td>123</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Predictor environmental dataset prediction with kinship matrix (Figure 4.19(e))</td>
<td>124</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Independent environmental dataset prediction with kinship matrix (Figure 4.19(f))</td>
<td>125</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>LFMM results table</td>
<td>126</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Euclidean allele frequency distance between continent/elevation populations</td>
<td>164</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Euclidean allele frequency distance between continent/elevation/latitude sub-populations</td>
<td>164</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Relationships between the first six vectors of the eigen decomposition of the inverted genetic distance matrix and geospatial variables elevation, longitude, annual mean precipitation, and annual mean temperature. Absolute values of slopes explain the strength of the relationship, and $R^2$ indicates how much variance is explained by the geospatial variable. Bold values are the highest in the column.</td>
<td>165</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Geography of 120 landrace accessions and common garden sites. (a) Location of collection sites of accessions and common garden sites. (b) Elevation of collection sites of accessions. Red lines indicate the elevations of the highland and lowland common garden sites. (c) Annual mean precipitation of collection sites of accessions.</td>
</tr>
<tr>
<td>3.2</td>
<td>Reaction norms for all measured phenotypic traits.</td>
</tr>
<tr>
<td>3.3</td>
<td>PCA analysis of family and plant trait values.</td>
</tr>
<tr>
<td>3.4</td>
<td>Pearson correlation between FITplantveg, P_INTsolid, P_EXTsolid, and M_DENsolid in Mexican Highland (a), Mexican Lowland (b), South American Highland (c), and South American Lowland (d) populations. In each subfigure, panels 1 and 2 show correlations within the lowland site (PV) and the highland site (MT), respectively. Blue shapes indicate positive correlation, red shapes indicate negative correlation, color intensity and shape size indicate strength of correlation, and asterisks indicate statistical significance (p-value thresholds = 0.001, 0.01, 0.05). Panel 3 shows the between-garden difference in correlation value for each pairwise correlation (positive/blue values indicate more positive correlations in the highland site than in the lowland site).</td>
</tr>
<tr>
<td>3.5</td>
<td>Pearson correlation between FITplantveg, P_INTsolid, P_EXTsolid, and M_DENsolid in Mexican Highland (a), Mexican Lowland (b), South American Highland (c), and South American Lowland (d) populations. In each subfigure, panels 1 and 2 show correlations within the lowland site (PV) and the highland site (MT), respectively. Blue shapes indicate positive correlation, red shapes indicate negative correlation, color intensity and shape size indicate strength of correlation, and asterisks indicate statistical significance (p-value thresholds = 0.001, 0.01, 0.05). Panel 3 shows the between-garden difference in correlation value for each pairwise correlation (positive/blue values indicate more positive correlations in the highland site than in the lowland site).</td>
</tr>
<tr>
<td>3.6</td>
<td>Finlay-Wilkinson regression of individual landrace FITplantveg fitnesses to mean FITplantveg at each environment (MT, PV2016, PV2017). These lines are plotted all together (a), as well as divided by population ((b), (c), (d), (e)).</td>
</tr>
</tbody>
</table>
Figure 3.7 Relationship between Finlay-Wilkinson regression slope and deviation from mean FITplantveg. (a) Relationship between regression slope and y-intercept for each landrace. Ellipses capture 90% of each population’s lines. Density plots show the distribution of data points along each axes for each population. Dashed lines demark one standard deviation to the left and to the right of the mean of the cumulative distribution, creating ninths in the coordinate space. (b) The number of landrace lines from each population that fall within each ninth from (a).

Figure 4.1 Loadings for the first five environmental PCA components.

Figure 4.2 Pearson Correlation between environmental predictors. Correlations between all predictors (a) resulted in sixteen clusters (less than 30% correlated). We then combined monthly values into seasonal values for solar radiation and water vapor, and combined all wind values into an annual average. In the second correlation (b), eighteen unique environmental layer clusters were found (less than 15% correlated), and within clusters, a single layer was selected.

Figure 4.3 Loadings for the first five Principal Components from the independent environmental dataset.

Figure 4.4 Population genetic structure. (a), (b) Each vertical line represents a genotyped individual. Individuals are grouped by landrace identity, and are colored by membership into genetic groups. Genetic structure complexity was best explained at either $k = 2$ or $k = 3$. (c), (d) Ancestry component of each landrace. (e) Hierarchical clustering dendrogram constructed from mean pairwise $F_{ST}$ values.

Figure 4.5 Maxent model suitability score predictions based on environmental layers.

Figure 4.6 Maxent model suitability score predictions based on PCA-transformed environmental layers.

Figure 4.7 Percent contribution of each predictor to each occurrence group.

Figure 4.8 Percent contribution of each predictor to each occurrence group.

Figure 4.9 Heatmaps and dendrograms of niche overlap (Schoener’s D) between landraces and genetic groups.

Figure 4.10 Heatmaps and dendrograms of niche overlap (Schoener’s D) between landraces and genetic groups, averaged between landrace and genetic group models, and with applied threshold $D = 0.6$.

Figure 4.11 Tangle plot of alignment of dendrograms demonstrating genetic similarity (mean pairwise $F_{ST}$) and niche similarity (Schoener’s D, entanglement coefficient = 0.09).

Figure 4.12 Elevation of sampled landrace accessions and wild relatives.

Figure 4.13 Distribution of AUC values for each occurrence group. Red lines indicate the value of the combined mean of the AUC for all ENM replicates for each category of occurrence groups (landrace, genetic group, wild relative, and random). (a) AUC from models built from environmental data. (b) AUC from models built from PCA-transformed environmental data.
Figure 4.14  Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.  

Figure 4.15  Relative metrics of niche breadth and correlated environmental variables. High values indicate high niche specificity (narrow niche breadth). The pink line is the average of Normalized AUC, Normalized AUC (PCA), 1 - Normalized B1, 1 - Normalized B2, and Normalized SSI. All variables are normalized for comparison.

Figure 4.16  Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.

Figure 4.17  Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.

Figure 4.18  Environmental heterogeneity and structure within landrace groups. Dendrograms of individual similarity based on Euclidean distance of 18 scaled environmental variables at location. TUXPEN branches slightly less at higher height and more at lower height, relative to CONICO and RANDOM.

Figure 4.19  Z-ratios of genotype and landrace name designation at predicting two different environmental datasets. Genotype similarity matrices estimated by categorical fastStructure group membership (k = 9, (a), (b)), Radial Basis function kernel of fastStructure ancestry coefficients (k = 3, (c), (d)), and kinship matrix ((e), (f)).

Figure 5.1  Optimal k values of population structure. (a) sNMF cross-entropy for k = 1 – 15. k = 3 is optimal, k = 2 – 5 are considered in further comparisons. (b) Bayesian Information Criterion (BIC) for k = 1 – 40. k = 3 is optimal.

Figure 5.2  Population genetic structure across North and South America, given k = 2 subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of p-values of population differentiation (F_{ST}) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate = 0.001) are in blue.

Figure 5.3  Population genetic structure across North and South America, given k = 3 subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of p-values of population differentiation (F_{ST}) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate = 0.001) are in blue.
Figure 5.4 Population genetic structure across North and South America, given $k = 4$ subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of $p$-values of population differentiation ($F_{ST}$) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate $= 0.001$) are in blue.

Figure 5.5 Population genetic structure across North and South America, given $k = 5$ subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of $p$-values of population differentiation ($F_{ST}$) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate $= 0.001$) are in blue.

Figure 5.6 Isolation by distance. (a) Distance between all individuals. (b) Distance between all continent/elevation/latitude subpopulations.

Figure 5.7 Relationship between eight continent/elevation/latitude subpopulations. (a) Neighbor-joining tree. (b) Popgraph. (c) Heatmap of isolation by graph distance.

Figure 5.8 Principal Components Analysis demonstrating genetic population structure within Mexican and South American maize landrace accessions. (a) 4 populations. (b) 8 subpopulations. (c) sNMF $k = 2$. (d) sNMF $k = 3$. (e) sNMF $k = 4$. (f) sNMF $k = 5$.

Figure 5.9 DAPC analysis and group membership prediction. Scatter plots (left) represent relationships between clusters across discriminant functions. Circles represent training data, squares represent testing data. Contingency tables (right) show assignment of individuals into groups. Columns represent the true group of the individual, and rows represent group predicted by the discriminant functions. (a), (b) Population clusters determined by DAPC at optimal $k = 3$. (c), (d) 4 continent/elevation subpopulations. (e), (f) 8 continent/elevation/latitude subpopulations.

Figure 5.10 $P_{ST}$-$F_{ST}$ comparison between highland and lowland Mexican populations (blue), between highland and lowland South American populations (green), and between all four populations (black).

Figure 5.11 Ear height divergence across PCs of relatedness. Each point represents the trait value of a genotyped individual, and color represents its membership in a population. Solid purple lines show the linear regression of trait values across PCs. Dashed blue lines represent 95% confidence intervals (used only for visualization purposes). Significant phenotypic divergence is found when the linear regression is greater or lower than the confidence interval. (a), (b), (c) Ear height of plants from both common gardens plotted against PC2, PC3, and PC6. (d), (e), (f) Ear height of plants the lowland garden (PV) plotted against PC2, PC3, and PC6. (g), (h), (i) Ear height of plants the highland garden (MT) plotted against PC2, PC3, and PC6.
Figure 5.12  Q-values of significantly higher ((a), (b), (c)) or lower ((d), (e), (f)) phenotypic divergence than expected given axes of population structure captured in the top 11 PCs. Comparisons are taken with the lowland garden (PV, (a), (d)), the highland garden (MT (b), (e)), or from both gardens ((c), (f)).

Black = 0-0.001, Red = 0.001-0.01, Orange = 0.01-0.05, Yellow = 0.05-0.1, White = 0.1-1.

Figure 5.13  Population structure across PC1 (a) and PC2 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density (k groups labeled MexHigh, Low, and SAIHigh) and population groups (MexHigh, MexLow, SAIHigh, and SAILow).

Figure 5.14  Population structure across PC3 (a) and PC4 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density (k groups labeled MexHigh, Low, and SAIHigh) and population groups (MexHigh, MexLow, SAIHigh, and SAILow).

Figure 5.15  Population structure across PC5 (a) and PC6 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density (k groups labeled MexHigh, Low, and SAIHigh) and population groups (MexHigh, MexLow, SAIHigh, and SAILow).
ACKNOWLEDGMENTS

I cannot begin to express my thanks for Matthew Hufford for welcoming me into his new lab at Iowa State University and for the years of instruction, support, and encouragement. I would also like to thank every postdoctoral associate, graduate student, undergraduate, and lab manager that has left a mark on the community of this lab, and on me as well. Li Wang, David Hufnagel, and Nancy Manchanda instructed and assisted me in the use of computational tools, Sarah Pedersen worked with me in the field collecting data, Samantha Snodgrass worked with me on multiple projects, and Kelsey Bryant, Kathryn Kananen, Lucas Roberts, Aimee Schulz, and the rest of the lab provided hours of collaborative discussion and debate. Their skill, knowledge, and work ethic are matched only by the warmth of their friendship, and I will miss them as I move on the next step of my career.

I would like to thank my dissertation committee members, Fred Janzen, Peng Liu, John Nason, and Amy Toth, for their support over years as my ideas developed into a completed study. I also wish to thank John Nason, Chanda Skelton, and Nicole Valenzuela for providing mentorship and support as I instructed labs and courses.

I am indebted to Jeffrey Ross-Ibarra, Daniel Runcie, and Sherry Flint-Garcia for their substantial input and guidance throughout my research. I am deeply thankful for the members of the labs of Ruairidh Sawers and Rubén Rellán-Álvarez at CINVESTAV, in particular Eric González-Segovia, Karla Juarez, and Carolina Cíntora-Martínez, who assisted in a great deal of field and lab work for my research and made me feel welcome in their country and in their lives. I would like to extend a special word of gratitude to María Rocío Aguilar-Rangel, for assisting in my data collection, for hosting me at her home, and for driving me to the doctor when I almost died from eating a bad papaya.
Thanks as well to collaborators Anthony Studer and Daniel Gates for their analyses and insights. I’d also like to recognize the assistance of Diego Jarquin and Lauren Mellenthin, whose insights helped guide my analyses.
ABSTRACT

Local adaptation is the process by which local populations evolve traits that increase their fitness relative to individuals of foreign populations. Traditional maize landraces across the Americas have long been observed to exhibit local adaptations to abiotic environmental variables, chiefly across elevational gradients. However, aside from field observations and studies of limited geographic and/or phenotypic scope, the strength and breadth of these local adaptations have remained unexplored.

To investigate local adaptations in maize landraces broadly, I conducted a broad-scale reciprocal transplant experiment with four populations of maize landraces (Mexican highland, Mexican lowland, South American highland, and South American lowland). I found that landraces grown near their native elevation have higher fitness than landraces from distant elevations. Several traits show higher between-population variance than expected given neutral genetic variance, suggesting that those traits are under divergent selective pressures. Furthermore, putatively highland-adaptive traits have strong, elevation-dependent correlations with fitness, further supporting their role in highland adaptation.

To investigate patterns of local adaptation across Mexico with greater resolution, I conducted Ecological Niche Modeling (ENM), Environmental Association Analyses (EAA), and environmental prediction using mixed effect models. I characterized landrace niches by associations with numerous environmental variables, and found that not all landraces are equally locally adapted; highland landraces have narrower niches and higher niche overlap than lowland landraces. Also, niche models of genetic groups on average had wider niches, suggesting that landrace groups aggregate adaptively similar accessions into categorical groups at least as well as Bayesian clustering algorithms. However, additive genetic variance estimated with a kinship matrix was a better predictor of environment than was landrace group, demonstrating the genetic foundation of local adaptation.
Finally, I identify SNPs strongly associated with environmental variables that constrain landrace distributions, and assert that these SNPs are good local adaptation candidates.

In my third chapter, I use genotypic data collected in the reciprocal transplant experiment and further investigate the partitioning of adaptive genetic variation into groups across geographic space. I find again that landrace genetic population structure is largely shaped by geographic distance, latitude, and elevation, and is consistent with demographic historical migration and bottleneck events. To incorporate these complex patterns of relatedness across the Americas, I looked for phenotypic trends across the axes of variation of an eigen-decomposition of a genetic distance matrix which capture elements of genetic structure between populations that correlate to latitude and elevation. Traits that varied more or less than expected given neutral genetic variance were identified as under divergent or stabilizing selection, receptively. I also identified genic SNPs with high $F_{ST}$ between highland and lowland populations in either Mexico or South America, or both, to find genes involved in highland adaptation unique to one continent or the other or parallel adaptation.

In summary, my research identifies strong phenotypic and genetic evidence of local adaptation in maize landraces across the Americas, and characterizes the local niches that have evolved.
CHAPTER 1. GENERAL INTRODUCTION

1.1 Dissertation Organization

This dissertation is organized into six chapters.

In Chapter 1, I introduce key theoretical concepts and the maize landrace model system at the center of this dissertation. I also describe the organization of the dissertation and the objectives of the research detailed in this dissertation.

Chapter 2 is a literature review of local adaptation and New World maize landraces. I detail the demographic history of maize domestication and expansion, the adaptations found among various maize landrace populations, and the scientific, economic, and cultural value of these populations.

Chapter 3, the first of my research chapters, details a large-scale reciprocal transplant experiment involving 120 landrace accessions from four populations (Mexican Highland, Mexican Lowland, South American Highland, South American Lowland) grown in lowland and highland common garden sites in Mexico. Landraces exhibited plasticity in fitness-relevant phenotypes between the two sites, wherein fitness trait values were higher at the accession’s native elevation, evincing local adaptation. Patterns differed between Mexican and South American accessions, suggesting that Mexican and South American landraces have adapted to highland conditions differently. Finally, putatively highland-adaptive traits anthocyanin pigmentation and leaf sheath macrohair density show high plasticity between garden sites. Both of these traits show pattern-specific correlations with fitness which differ between Mexican and South American landraces, and between highland and lowland landraces, indicating that these traits are part of lineage-specific highland adaptation evolutionary strategies. Genetic sequence data collected from plants grown in this reciprocal transplant experiment are analyzed in Chapter 5.

Chapter 4 employs Ecological Niche Modeling (ENM) and Environmental Association Analyses (EAA) to compare the ecological niches of 989 Mexican maize landrace accessions from 10 landraces.
ENMs were constructed for groups of accessions binned according to landrace identity or genetic population structure, and were compared with reference to environmental variable importance, niche breadth, and niche similarity/overlap. Niche similarity is highly positively correlated with genetic similarity, supporting genetic bases for modeled niches. We find that highland groups (landraces and genetic groups) have narrower niches than lowland groups, and that this effect is stronger than expected given neutral expectations. Comparative niche analyses indicate that ENMs built from landrace identity or from genetic structure perform roughly equitably at predicting occurrence points. However, mixed model equations find that a genetic kinship matrix explains more environmental variance than does a landrace membership matrix, suggesting that population genetic structure below the level of landrace exists and is important for local adaptation, and that adaptation may be due to fewer large-impact genetic loci. Finally, we identify candidate SNPs strongly associated with environmental variables. Many of these SNPs are uncharacterized and warrant further investigation.

Chapter 5 further investigates local adaptation between populations with high gene flow and complex demographic histories. Phenotype data from the reciprocal transplant experiment discussed in Chapter 3 are re-analyzed in light of genetic structure. $Q_{ST} - F_{ST}$ analyses reveal that many traits are significantly diverged between populations. However, $Q_{ST} - F_{ST}$ comparisons necessitate defining distinct and equally related populations (even when relatedness patterns are gradual and complex, as with maize landraces) and good estimates of additive genetic variance (which typically require multi-generational controlled breeding experiments), and for these reasons, $Q_{ST} - F_{ST}$ can be prone to high rates of false positives. The method $Q_{PC}$ utilizes the eigen-decomposition of a kinship matrix to separate between-population (high-order eigenvectors) from within-population (low-order eigenvectors) genetic variance, assisting in the estimation of both significant axes of population structure and additive genetic variance. Plant height, ear height, tassel branch number, macrohair density, and flowering time show signatures of divergent selection across population structure axes (particularly those that capture structure across an elevational gradient), and traits related directly to fitness (stand count, ear-producing stand count, and anthesis-silking interval)
show signatures of stabilizing selection. As in Chapter 4, we identified candidate SNPs strongly associated with altitude, mean annual temperature, and mean annual precipitation.

I conclude with general remarks in Chapter 6.

1.2 Dissertation Objectives

My research objectives are:

1. Characterize the strength and breadth of New World maize landrace adaptations to local abiotic environments.

2. Demonstrate the adaptive nature of the putatively highland-adaptive traits anthocyanin pigmentation and leaf sheath pilosity.

3. Map the relationships between environmental variable, adaptive phenotype, and genetic architecture.

4. Determine the degree to which landrace names capture adaptive genetic variation.
CHAPTER 2. LITERATURE REVIEW

2.1 Theory

Species exist across ranges of geography which vary for any number of environmental components. Though some spaces are more environmentally homogeneous than others, no space is devoid of environmental variability at some spatial or temporal scale. Likewise, though some species are tolerant of a broader range of environmental conditions than others, all species have some limits to the kinds of environments in which they can survive. Ideally, any organism could make itself well-suited (either morphologically, physiologically, or behaviorally) for any environment, but such perfect phenotypic plasticity is not possible (Bradshaw, 1965). To the degree that complete phenotypic plasticity is impossible, species evolve to succeed in some subset of habitats while foregoing others, resulting in fitness trade-offs. At the level of intra-specific populations (or demes), the process of adapting to local environmental conditions is called local adaptation. More formally, demes are said to be locally adapted when they meet the “local vs. foreign” criterion of local adaptation, in which a local population must exhibit higher fitness than foreign populations (Kawecki and Ebert, 2004).

Local adaptation is the result of natural selection, a fundamental force behind the evolution of biodiversity. Though local adaptation is common (over 70% of plant studies find local adaptation (Leimu and Fischer, 2008; Hereford, 2009)), it is not universal. Local adaptation can be slowed or prevented outright by gene flow, genetic drift, temporal environmental stochasticity, antagonistic pleiotrophy, and local extinction (Reznick and Ghalambor, 2001; Kawecki and Ebert, 2004). Because homogenizing factors like drift and gene flow act to erode the phenotypic divergence between locally adapting populations, populations that persistently display divergent adaptive phenotypes must be experiencing recent or ongoing divergent selective pressure. Therefore, because metapopulations consist of locally adapted demes which maintain divergent adaptive phenotypes, and because
those phenotypes are associated with (and putatively maintained by) divergent present-day en-
vironmental selective pressures, biological systems exhibiting local adaptation are well-suited for the
investigation of causal relationships between adaptation and environment (Reznick and Ghalam-
bor, 2001). Conversely, phenotypic and genetic population divergence due to local adaptation can
result in reproductive isolation and facilitate further divergence and possibly speciation (Schluter,
2001; Sobel et al., 2010).

Numerous methods have been developed for the detection and characterization of local adap-
tation. The most diagnostic test of local adaptation is reciprocal transplantation, which directly
tests whether local demes have higher fitness than members of foreign demes in their native habitat
(Savolainen et al., 2013). $F_{ST}$ methods identify SNPs that vary in frequency between pairs of
demes, but such methods are prone to high false positive rates (Fourcade et al., 2013). Methods
that correlate allele frequencies with environmental variables (such as Bayenv (Coop et al., 2010),
SAMβADA (Stucki et al., 2017), and LFMM (Frichot et al., 2013)) are less prone to Type 1 errors.
Methods employing $Q_{ST} - F_{ST}$ tests compare the proportion of quantitative trait variance that is
due to within-population additive genetic variance ($Q_{ST}$) to the proportion of total genetic variance
that is due to within-population variance ($F_{ST}$). If $Q_{ST} > F_{ST}$, trait divergence between popula-
tions ($Q_{ST}$) exceeds neutral expectations ($F_{ST}$), likely due to diversifying selection. Stabilizing or
uniform selection can result in $Q_{ST} < F_{ST}$, and trait divergence can be explained by drift if $Q_{ST}$
roughly equals $F_{ST}$ (Leinonen et al., 2013).

2.2 Model System

Men and women have long been aware of the relationship between living things and their local
environments. Geographic patterns of phenotypic variation were noted long before we uncovered
the mechanisms of their origination and transmission. Early agriculturalists experimented (with
unknown levels of foreknowledge and intent) with transporting domesticated seeds and plants from
environment to environment, and with imparting traits from one population to another via hy-
bridization and introgression. Thousands of years of such experiments resulted in a rich diversity of domesticated plants, and more specifically, many varieties with unique and beneficial phenotypes.

Maize (Zea mays subsp. mays) is a model system of high agronomic (Shiferaw et al., 2011; Hake and Ross-Ibarra, 2015), economic (Shiferaw et al., 2011; Ranum et al., 2014), cultural (Fernandez Suarez et al., 2013; Perales, 2016), and scientific (Dumas and Mogensen, 1993; Fedoroff, 2001; Stern et al., 2004; Hake and Ross-Ibarra, 2015) value, and as such, a wealth of resources have been developed for the study of maize. Maize was domesticated in the lowlands of the Balsas River Valley in Mexico from the teosinte taxon Zea mays subsp. parviglumis roughly 9000 years BP (Matsuoka et al., 2002). From there, maize was carried across across North America into South America as early as 6000 years BP (Bush et al., 1989; Grobman et al., 2012), north into present-day United States by about 4500 years BP (Merrill et al., 2009), and around the world (Tenaillon and Charcosset, 2011; Van Heerwaarden et al., 2011; Hake and Ross-Ibarra, 2015). Presently, maize is grown under a broad range of temperatures, precipitations, and soil types, and a wider range of elevations and latitudes than any other crop (Ruiz Corral et al., 2008; Shiferaw et al., 2011).

Adaptation to divergent selective ecological pressures is one of the major modes of speciation (Schluter, 2001; Sobel et al., 2010; Nakazato et al., 2010), and the same is true of maize landrace formation (Eagles and Lothrop, 1994; McCouch, 2004). At points along the range expansion of maize, farmers selected plants and seed desirable for human applications and consumption and suitable for growth in their local environment. Between 2 and 4% of maize genes (about 1200 genes in total) show signs of selection during maize domestication and improvement (Wright et al., 2005). Over generations of propagation and selection, this process formed varietal populations called landraces. In Mexico alone, 59 landraces have been defined (Vielle-Calzada and Padilla, 2009). These landraces are grown and maintained to the present by smallholder farmers day as dynamic, evolving populations (Mijangos-Cortes et al., 2007; Dyer and López-Feldman, 2013) that diverged and remain distinct despite considerable gene flow (Ortega, 1995; Pressoir and Berthaud, 2004). Gene flow is typically a constraint to local adaptation, but populations with unstable structure, large-scale geographic range shifts, and/or frequent extinction and recolonization events
(as may be the case when seed types are abandoned or swapped between farmers) may benefit from gene flow (Slatkin, 1987).

Maize landraces exhibit many unique morphological, physiological, and phenological characteristics, many of which covary with climate, soil type and quality, and geography. Farmers consciously select primarily for ear characteristics (kernel filling, large ears, varietal consistency, (Louette and Smale, 2000)), the environment selects for plant survival and reproduction (Cleveland and Soleri, 2007), and the combination of these selective factors comprise the agroecosystem to which landraces evolve (Villa et al., 2005). (See Villa et al. (2005) for a review of the defining characteristics of landraces.)

Of particular interest are adaptations to highland and lowland conditions (Eagles and Lothrop, 1994). Phenotypic and genetic patterns covary strongly with environmental variables along elevational gradients (Dyer and López-Feldman, 2013). Highland adaptations are hypothesized to be imparted via introgression from Zea mays subsp. mexicana which is adapted to cool, dry highland conditions (Hufford et al., 2012) with features discussed by Lauter et al. (2004), notable among these anthocyanin pigmentation and increased macrohair pilosity on leaf sheaths. Lauter et al. (2004) review high-elevation, cold temperature adaptations which differentiate highland mexicana from mid- and low-elevation parviglumis, and these same highland characteristics have long been reported to differentiate highland and lowland maize landraces as well, in some cases due to adaptive introgression of these traits from mexicana into highland Mexican landraces (Hufford et al., 2013). Hufford et al. (2013) found that mexicana introgression into sympatric maize overlapped regions containing QTL identified by Lauter et al. (2004) that control pilosity and pigmentation, giving a clear picture of the adaptive importance of mexicana introgression. Macrohairs, depending on structure, density, and distribution on the plant, can increase surface friction, thereby decreasing wind speed across the surface and increasing the thickness of the air boundary layer around the plant, reducing heat loss and transpiration (Schuepp, 1993). Haplotypes introgressed from mexicana which overlap QTL controlling pilosity and pigmentation are found in Guatemalan and southwestern USA landraces (Da Fonseca et al., 2015), but not in highland landraces in the South
American Andes (Wang et al., 2017). Takuno et al. (2015) also found little evidence of convergent evolution between highland Mexican and Andean maize landraces. Highland-adaptive loci from Mexico were unlikely to be retained during the slow migration through lowland Central and South America (environments in which such loci would presumably result in lower fitness) to the Andes. These conditions and observations support the premise that smallholder-managed landrace populations evolve adaptations to local environment (Bellon et al., 2018).

Landraces are of critical importance for both Mexican small-scale farms and modern large-scale farming operations. More than 75% of the maize produced in Mexico comes from small (3 or fewer tons per hectare) farms (Bellon et al., 2018), and up to 80% of the arable land in Mexico is devoted to growing landrace maize (Louette et al., 1997). Landraces are frequently out-yielded by modern hybrids, but in their own home environments, landraces can and often do out-perform hybrids (Bellon et al., 2018; Perales, 2016; Bellon et al., 2003; Mercer and Perales, 2018). Due to their distinctive ear morphologies and unique culinary qualities, some landraces fill niche market demands, safeguarding them from abandonment and replacement by agronomically superior modern hybrids (McLean-Rodríguez et al., 2019). About 31.6% of the land area of Mexico is “Marginally Suitable” (compared to 6.4% “Suitable”, 25.1% “Moderately Suitable”, and 36.9% “Not Suitable”, (Monterroso Rivas et al., 2011)), and it is in marginally suitable areas that local landraces have an adaptive edge over modern improved lines (Ceccarelli, 1994). (Here, “Marginally Suitable” means mainly that precipitation is below 600mm annually, and such lands in Mexico are more frequently found at higher elevation.) Conversely, the areas of highest landrace diversity are the areas projected to suffer the least under climate change (Monterroso Rivas et al., 2011), though almost every landrace is expected to incur some level of range contraction in the next 10 to 30 years (Ureta et al., 2012). As the “Suitable” land of Mexico is projected to be most adversely affected and reduced by future climate change, and because areas of high landrace richness are projected to be largely spared by climate change, landraces are positioned to become even more important in the future. Landraces that harbor advantageous alleles are desirable for modern maize breeders. The intense breeding programs that have developed modern agronomic inbred
lines have reduced genetic diversity and capacity for adaptive plasticity (McCouch, 2004; Gage et al., 2017). Reincorporation of landrace germplasm can restore key genetic variants that impart various adaptations, including those to environmental conditions. Efforts by breeders to incorporate highland Mexican landrace germplasm into improved hybrids began in the 1940s and continue to the present day, and have produced lines adapted to temperate regions around the world (Eagles and Lothrop, 1994; Varshney et al., 2018). Highland maize landraces have traits like pilose, pigmented stems and droopy, leathery ears, which are adaptive in cold and low precipitation environments, and resistance to insect and fungal field pests (Eagles and Lothrop, 1994). Some highland lines express the latente ("latency") trait, in which the plant avoids drought stress damage by halting growth during drought conditions and resuming growth quickly upon recovery irrigation (Eagles and Lothrop, 1994). Better adapted crops are one of the least environmentally harmful and most secure strategies for increasing productivity and security of worldwide food supply (Byerlee, 1996).

Though the value of maize landraces is increasingly recognized and is projected to increase over time, many landraces are threatened with replacement and subsequent loss. Within a metapopulation framework (Levins, 1969), replacement of local varieties with foreign varieties is equivalent to extinction due to specialist competitors from another deme, and the anthropogenic forces that lead to that replacement are equivalent to habitat loss or "persecution" by humans (Hanski, 1998). Though landraces are increasingly replaced in by modern hybrids, landraces persist in areas of marginal environments (Ceccarelli, 1994) and marginal social/economic/technological infrastructure (Byerlee, 1996; de la Barrera and Martínez, 2018; McLean-Rodríguez et al., 2019). Numerous national and international agencies (reviewed in part by Cohen et al. (1991)) engage in in situ and ex situ conservation of maize landraces and wild relatives. When prioritizing populations for conservation efforts, a host of criteria are considered, including potential economic value, importance to research, importance to culture, importance to ecosystem functions, current geographic distribution, current abundance within that distribution, risk of distribution and/or abundance decline due to environment change caused by land use or climate change, endemism, and genetic/geographic/environmental relatedness to other accessions (Bellon et al., 2003; Perales et al.,
Though *ex situ* conservation is certainly an indispensable tool in most conservation programs, *in situ* conservation is usually regarded as superior when feasible. Open-pollinated landrace populations maintained by traditional farming methods by smallholder farmers promote adaptive evolution by maintaining large effective populations, preserving standing genetic diversity, and permitting maize populations to adapt to diverse and changing environments (Bellon et al., 2018).

### 2.3 References


CHAPTER 3. DEMONSTRATION OF LOCAL ADAPTATION OF MAIZE
LANDRACES BY RECIPROCAL TRANSPLANTATION

Adapted from a paper to be submitted to Evolutionary Applications.

Authors: Garrett M Janzen\textsuperscript{1}, María Rocío Aguilar-Rangel\textsuperscript{2}, Carolina Cíntora-Martínez\textsuperscript{2}, Karla Azucena Blöcher-Juárez\textsuperscript{2}, Eric González-Segovia\textsuperscript{2}, Anthony Studer\textsuperscript{3}, Daniel E Runcie\textsuperscript{4}, Sherry A Flint-García\textsuperscript{5,6}, Rubén Rellán-Álvarez\textsuperscript{2,7}, Ruairidh J H Sawers\textsuperscript{2,8}, Matthew B Hufford\textsuperscript{1}

\textsuperscript{1}Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA
\textsuperscript{2}Langebio, Cinvestav; Km 9.6 Libramiento Norte Carretera León; Irapuato, Guanajuato, Mexico 36821
\textsuperscript{3}College of Agricultural, Consumer, & Environmental Sciences, University of Illinois Urbana-Champaign; 1102 South Goodwin Avenue; Urbana, Illinois, USA 61801
\textsuperscript{4}Department of Plant Sciences, University of California-Davis; 278 Robbins; Berkeley, California, USA 95616
\textsuperscript{5}Agricultural Research Service, United States Department of Agriculture, Columbia, Missouri, 65211
\textsuperscript{6}University of Missouri; 301 Curtis Hall; Columbia, Missouri, USA 65211
\textsuperscript{7}Molecular and Structural Biochemistry, North Carolina State University; 128 Polk Hall; Raleigh, North Carolina, USA 27695-7622
3.1 Abstract

Populations are said to be locally adapted when they exhibit higher fitness than foreign populations in their native habitat. Examples of maize landrace adaptations to highland and lowland conditions are widely reported, and are of interest to researchers and breeders. Although many landraces are commonly believed to be locally adapted, only a few studies of limited scope have addressed the question. To determine the prevalence and strength of local adaptation, we have performed a reciprocal transplant experiment across an elevational gradient in Mexico. We grew 120 landrace accessions from North and South America in highland and lowland common gardens and collected phenotypes relevant to fitness, including reported highland-adaptive traits such as anthocyanin pigmentation and macrohair density.

We find phenotypic patterns consistent with local adaptation, though these patterns differ between Mexican and South American populations. Generally, Mexican landraces had higher fitness than South American landraces when grown in our Mexican sites. However, populations exhibit higher values of fitness including yield metrics when grown at their native elevation. Highland populations expressed generally higher anthocyanin pigmentation than lowland populations, and moreso in the highland site than in the lowland site. Macrohair density was largely non-plastic, but Mexican landraces and highland landraces were generally more pilose. Analysis of $\delta^{13}$C indicated that lowland populations have lower water use efficiency. Each population demonstrated garden-specific correlations between highland trait expression and fitness, with stronger positive correlations in the highland site.

These results give substance to the long-held presumption of local adaptation in New World maize landraces, to both elevation and other environmental variables across North and South America.
### 3.2 Introduction

Populations evolve adaptations to selective pressures imparted by their biotic and abiotic environment. Over time, given sufficiently low genetic drift and gene flow, theory predicts that a population will adapt to the particular selective pressures of its local environment (Leimu and Fischer, 2008). In particular, populations are said to be locally adapted when they meet the “Foreign vs. Local” criterion of local adaptation, in which a local population must exhibit higher fitness than foreign populations when grown in the same location (Kawecki and Ebert, 2004).

Traditionally, attempts to validate and quantify local adaptation in natural populations have relied on common garden experiments (Savolainen et al., 2013; Turesson, 1922; Clausen et al., 1940; Fraser et al., 2011). Exposing individuals from populations spanning environmental gradients to a common environment permits assessment of genotype-by-environment ($G \times E$) interactions (Savolainen et al., 2013). Reciprocal transplant experiments are preferable to common garden experiments in many cases, as the scale and complexity of the environments of the included populations can be modeled more holistically, rather than being reduced to single or few environmental variables in controlled laboratory settings (Kawecki and Ebert, 2004; Savolainen et al., 2013; Limpens et al., 2012). For many species, common garden and reciprocal transplant experiments are logistically impractical and potentially unethical, which is why they are more commonly implemented with plants (Clausen et al., 1948).

Maize (*Zea mays* subsp. *mays*) is an extensively studied model system of high agronomic (Shiferaw et al., 2011), economic (Shiferaw et al., 2011; Ranum et al., 2014), cultural (Fernandez Suarez et al., 2013; Perales, 2016), and scientific (Dumas and Mogensen, 1993; Fedoroff, 2001; Stern et al., 2004) value. Maize was domesticated in the lowlands of the Balsas River Valley in Mexico from the teosinte taxon *Zea mays* subsp. *parviglumis* roughly 9000 years BP (Matsuoka et al., 2002). From there, maize was carried across across North America into South America as early as 6000 years BP (Grobman et al., 2012; Bush et al., 1989), north into the present-day United States by about 4500 years BP (Merrill et al., 2009), and around the world as part of the Columbian exchange (Tenaillon and Charcosset, 2011; Van Heerwaarden et al., 2011). Presently,
maize is grown under a broad range of temperature, precipitation, and soil types, being cultivated across more elevations and latitudes than any other crop (Ruiz Corral et al., 2008; Shiferaw et al., 2011).

At points along the range expansion of maize, farmers selected plants and seed that were both desirable for human applications and consumption and suitable for growth in their local environment. Over generations of propagation and selection, this process formed varietal populations called landraces. These landraces are grown and maintained by smallholder farmers as dynamic, evolving populations (Mijangos-Cortes et al., 2007; Dyer and López-Feldman, 2013) with low but significant gene flow between them (Ortega, 1995). Most of the maize production of Mexico comes from small-scale subsistence farms that extensively cultivate landraces (Bellon et al., 2018). Landraces are typically out-yielded by modern hybrids, but in their own home environments, landraces can and often do out-perform hybrids (Bellon et al., 2018; Perales, 2016; Bellon et al., 2003; Mercer and Perales, 2018).

Maize landraces exhibit many unique morphological, physiological, and phenological characteristics, many of which covary with climate, soil type and quality, and geography (Wellhausen et al., 1952). (See Villa et al. (2005) for a review of the defining characteristics of landraces.) While farmers consciously select primarily for ear characteristics that are indirectly related to fitness outside of an agronomic context (kernel filling, ear size, varietal consistency (Louette and Smale, 2000)), the environment selects for plant survival and reproduction (Cleveland and Soleri, 2007). The combination of these selective factors comprise the agroecosystem to which landraces evolve adaptations (Villa et al., 2005).

Some of the most striking adaptations in maize landraces are to high- and low-elevation conditions (Eagles and Lothrop, 1994). Marked phenotypic variation and genetic structure are linked to elevation (though elevation itself may not be the causal agent (Dyer and López-Feldman, 2013)). In at least some high-elevation regions, adaptations are hypothesized to be imparted via introgression from the maize wild relative *Zea mays* subsp. *mexicana*, which is adapted to cool, dry highland conditions (Hufford et al., 2012; Lauter et al., 2004). Notable similarities between highland maize
and *mexicana* include highly pigmented and hairy (pilose) leaf sheaths. Hufford et al. (2013) found that *mexicana* introgression into sympatric maize in Mexico overlapped chromosomal regions identified as QTL by Lauter et al. (2004) for pilosity and pigmentation. These features reduce both heat loss and transpiration and can be advantageous in cool, dry regions (Schuepp, 1993; Chalker-Scott, 1999). Interestingly, haplotypes from *mexicana* are not found in highland landraces in the South American Andes (Wang et al., 2017). Takuno et al. (2015) also found little evidence of convergent evolution between highland Mexican and Andean maize landraces. Highland-adaptive loci from Mexico were unlikely to be retained during the slow migration to the Andes through lowland Central and South America, environments in which such loci would presumably be deleterious. No studies have evaluated and compared the extent of high elevation adaptation in Mexican and South American maize despite many reporting the prevalence of local adaptation among landraces as fact (Harlan, 1975; Villa et al., 2005; Navarro et al., 2017).

A better understanding of local adaptation may prove critical for modern maize breeders. The intense breeding programs that have developed modern inbred lines have drawn from limited germplasm and, through selection, have further reduced genetic diversity and capacity for adaptive plasticity (Gage et al., 2017). Reincorporation of landrace germplasm can restore key genetic variants that impart adaptations to challenging environments. Despite this potential, research has not fully addressed whether maize landraces do, in fact, exhibit reciprocal home-site advantage, the definition of local adaptation. Landrace geographical extents have been shown to correspond to elevational and climatic factors (Ruiz Corral et al., 2008; Arteaga et al., 2016), supporting (but not demonstrating) local adaptation. Mercer et al. (2008) conducted a reciprocal transplant experiment across three common garden elevations (lowland, midland, and highland) using landraces found in Chiapas, Mexico, and found that landraces from this area exhibit differential reaction norms in response to elevation. However, this study was limited by the small number and geographical range of maize landraces included, as well as logistical issues preventing analysis of their lowland site. Their follow-up experiment (Mercer and Perales, 2018), though successfully collecting data from the lowland site, was similarly narrow in geographical range and landrace diversity.
To investigate the extent of local adaptation between highland and lowland maize landraces, we conducted an elevational reciprocal transplant experiment. We compare highland-adaptive traits, yield traits, and isotopic data from maize landrace accessions from highland and lowland populations from Mexico and South America (four populations, 30 landraces per population) grown in highland and lowland Mexican field sites. Finally, we correlate values of highland-adaptive traits with fitness traits to investigate their elevation-specific impact on fitness.

3.3 Methods

3.3.1 Field Experiment Design

Landrace accessions from CIMMYT that met the following criteria were considered for inclusion:

1) Accessions are present in the Seeds of Discovery (SeeDs) dataset (Pixley et al., 2017).
2) Accessions had latitude and longitude data from North or South America.
3) The elevation of the accession was below 1000 m or above 2000 m.

From eligible accessions, 30 pairs of highland and lowland accessions were chosen from both Mesoamerica and South America with no more than a single pair in each one-degree bin of latitude. These samples are described throughout the manuscript as four discrete populations (Mexican Highland, Mexican Lowland, South American Highland, South American Lowland, abbreviated as Mex High, Mex Low, SA High, and SA Low, respectively) with 30 accessions per population. Note, however, that two accessions included in the “Mexican” populations are in fact from Guatemala.

The two common garden sites that comprise this reciprocal transplant are the Winter Services nursery site near Puerto Vallarta in the Pacific coastal lowlands (elevation 54 m) of Mexico, labeled PV, and a CIMMYT field site near the town of Metepec in the highlands (elevation 2852 m) of the Mexican Central Plateau, labeled MT. Seed lines were regenerated at the field site for one generation prior to the experiment to reduce seed storage and maternal effects. Best local practices for irrigation, fertilizer, and pest/weed control were used across sites. The Metepec field experiment was conducted in the summer of 2016. The Puerto Vallarta field experiment was conducted in the winter of 2016, but virus damage led us to repeat the field experiment at the same site in the winter
of 2017. Certain traits were collected from both years of the Puerto Vallarta field site. A map of the field sites and geographical origin of each accession and boxplots summarizing the elevational and annual precipitation compositions of these four populations can be found in Figure 3.1.

Each field was arranged in a complete block design with two blocks of 120 rows of 15 seeds of a landrace accession. Landraces from latitudinal pairs were planted in adjacent rows.

### 3.3.2 Phenotypic Data Collection

Phenotypes (Table 3.1) were collected from Metepec and both years of the Puerto Vallarta common garden experiment. Two healthy, representative plants from the interior of each row were selected and tagged. Individual plant phenotype data (plant height, ear height, ear number, tassel length, and tassel branch number) were collected from tagged plants.

Other traits were collected at the row level, such as stand count, ear-producing stand count, and barrenness. Days to anthesis and days to silking were recorded as the number of days until 50% of the row exhibited silk emergence or anther exertion on more than half of the main tassel spike, respectively. Anthesis-silking interval was calculated as the difference in these two values.

Primary ears from tagged plants from the Metepec and Puerto Vallarta 2016 field sites were returned to the lab to be photographed and processed for analysis. Total ear weight, ear length, ear diameter, and number of kernels per ear row were measured.

Methods for field visual assessment of anthocyanin pigmentation and macrohair are derived with modification from Lauter et al. (2004). Pigment was scored for pattern, intensity, and extent. The extent of leaf sheath anthocyanin pigmentation was visually scored on a scale of 0-4 (at 25% intervals), from ground level up. The intensity of leaf sheath pigmentation across the plant was visually scored on a scale of 0-4. Though all pigmentation patterns share some degree of genetic and environmental control, spots and banded patterns frequently co-occur as an induced response to pathogenic stress (Selinger and Chandler, 1999), whereas uniform pigmentation (and leaf sheath macrohair expression) is shown to be inducible by highland conditions in some landraces (particularly those harboring introgressed QTL from *mexicana*, (Hufford et al., 2013)). For this reason, the
Table 3.1  Names and descriptions of all collected phenotypes.

<table>
<thead>
<tr>
<th>Code</th>
<th>Trait Name</th>
<th>Unit of Measurement</th>
<th>Trait Description</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>STD</td>
<td>Stand Count</td>
<td>count</td>
<td>Number of plants surviving to sexual maturity</td>
<td>Row</td>
</tr>
<tr>
<td>PE</td>
<td>Ear-Producing Stand Count</td>
<td>count</td>
<td>Number of plants surviving to produce ears</td>
<td>Row</td>
</tr>
<tr>
<td>BRN</td>
<td>Barrenness</td>
<td>1-(PE/STD)</td>
<td>Percent of plants in row that produce no ears</td>
<td>Row</td>
</tr>
<tr>
<td>DTA</td>
<td>Days to Anthesis</td>
<td>count</td>
<td>Number of days between planting and 50% of plants in anthesis</td>
<td>Row</td>
</tr>
<tr>
<td>DTS</td>
<td>Days to Silking</td>
<td>count</td>
<td>Number of days between planting and 50% of plants silking</td>
<td>Row</td>
</tr>
<tr>
<td>ASI</td>
<td>Anthesis/Silking Interval</td>
<td>DTS-DTA</td>
<td>Number of days between 50% silking and 50% anthesis</td>
<td>Row</td>
</tr>
<tr>
<td>PH</td>
<td>Plant Height</td>
<td>cm</td>
<td>Distance between the ground and the ligule of the flag leaf</td>
<td>Plant</td>
</tr>
<tr>
<td>EH</td>
<td>Ear Height</td>
<td>cm</td>
<td>Distance between the ground and the primary ear-bearing node</td>
<td>Plant</td>
</tr>
<tr>
<td>TL</td>
<td>Tassel Length</td>
<td>cm</td>
<td>Distance from tip of the main spike to the bottom branch</td>
<td>Plant</td>
</tr>
<tr>
<td>TBN</td>
<td>Tassel Branch Number</td>
<td>count</td>
<td>Number of tassel branches that attach to the main spike</td>
<td>Plant</td>
</tr>
<tr>
<td>EN</td>
<td>Ear Number</td>
<td>count</td>
<td>Number of seed-producing ears produced</td>
<td>Plant</td>
</tr>
<tr>
<td>PM</td>
<td>Ear Mass</td>
<td>g</td>
<td>Mass of the primary ear</td>
<td>Plant</td>
</tr>
<tr>
<td>LM</td>
<td>Ear Length</td>
<td>cm</td>
<td>Length of the primary ear</td>
<td>Plant</td>
</tr>
<tr>
<td>GPH</td>
<td>Kernels per Row</td>
<td>count</td>
<td>Number of kernels in a row on the primary ear</td>
<td>Plant</td>
</tr>
<tr>
<td>DM</td>
<td>Ear Diameter</td>
<td>cm</td>
<td>Diameter of the primary ear</td>
<td>Plant</td>
</tr>
<tr>
<td>d13C</td>
<td>$\delta^{13}$C</td>
<td>$[R_{\text{Sample}}/R_{\text{Standard}} - 1] \times 1000$</td>
<td>Degree of inclusion of $^{13}$C in flag leaf tissue</td>
<td>Plant</td>
</tr>
<tr>
<td>FITplant</td>
<td>Agronomic Plant Fitness</td>
<td>$PE/15 \times \sqrt{EN} \times PM$</td>
<td>Adjusted plant fitness</td>
<td>Plant</td>
</tr>
<tr>
<td>FITplantveg</td>
<td>Vegetative Plant Fitness</td>
<td>$PE/15 \times \sqrt{EN}$</td>
<td>Adjusted plant fitness</td>
<td>Plant</td>
</tr>
<tr>
<td>P_INTsolid</td>
<td>Pigment Intensity (Solid Pattern)</td>
<td>visual 0-4 code scale</td>
<td>Intensity of anthocyanin pigmentation</td>
<td>Plant</td>
</tr>
<tr>
<td>P_INTspot</td>
<td>Pigment Intensity (Spot Pattern)</td>
<td>visual 0-4 code scale</td>
<td>Intensity of anthocyanin pigmentation</td>
<td>Plant</td>
</tr>
<tr>
<td>P_EXTsolid</td>
<td>Pigment Extent (Solid Pattern)</td>
<td>visual % code scale</td>
<td>Extent of anthocyanin pigmentation from the ground up</td>
<td>Plant</td>
</tr>
<tr>
<td>P_EXTspot</td>
<td>Pigment Extent (Spot Pattern)</td>
<td>visual % code scale</td>
<td>Extent of anthocyanin pigmentation from the ground up</td>
<td>Plant</td>
</tr>
<tr>
<td>M_DENsolid</td>
<td>Macrohairs Density (Sheath)</td>
<td>visual 0-4 code scale</td>
<td>Density of macrohairs of the second leaf sheath from top</td>
<td>Plant</td>
</tr>
<tr>
<td>M_DENmemarg</td>
<td>Macrohairs Density (Sheath Margin)</td>
<td>visual 0-4 code scale</td>
<td>Density of macrohairs of the second leaf sheath margin</td>
<td>Plant</td>
</tr>
</tbody>
</table>
“solid” pattern may better represent pigmentation that lends towards highland adaptation, and the “spot” pattern may represent a stress response to other biotic and abiotic factors. Plants were given the categorical qualitative score of either “banded” (or “streaked” (Selinger and Chandler, 1999)), “spotted”, or “uniform”. When a plant exhibited multiple patterns, the highest category took precedence (uniform, then banded, then spotted). Plants with patterns of “banded” or “spotted” were binned into a “spot” group. Macrohair density on the second leaf sheath from the top of the plant was visually scored on a scale of 0-4. The “uneven/patchy” macrohair pattern likely differs from the “uniform” pattern only by degree, but pubescence restricted to the sheath margin may indeed be under different genetic control and play a different or null role in adaptation. Therefore, plants were grouped by macrohair trait pattern (“solid” or “margin”).

Two adjusted fitness metrics were computed from the combination of several fitness traits (adapted from (Mercer et al., 2008)). Agronomic plant fitness (FITplant) incorporates percent survival to produce ears, the number of ears produced, and primary ear weight. To calculate adjusted fitness for plants that either did not produce ears by the time of harvest or were not harvested for collection of ear traits, a second plant fitness trait (vegetative plant fitness, FITplantveg) excludes ear weight. Both formulations of adjusted fitness square-root transform ear number to account for diminishing yield returns of secondary, tertiary, and subsequent ears. We calculate adjusted fitness thusly:

\[
FIT\text{plant} = \frac{PE}{15} \times \sqrt{EN} \times PM
\]

\[
FIT\text{plantveg} = \frac{PE}{15} \times \sqrt{EN}
\]

Flag leaves from tagged plants from Metepec and Puerto Vallarta 2016 were collected for Carbon isotope discrimination analysis. Carbon isotopic composition δ\text{13C} was calculated in reference to a standard (an in-house B73 population). Percent Carbon and percent Nitrogen were also calculated as a control measure (data not shown). The equation for δ\text{13C} is as follows:

\[
\delta^{13}\text{C} = \left\{ \left[ \frac{^{13}\text{C}_{\text{sample}}}{^{12}\text{C}_{\text{sample}}} \right] / \left[ \frac{^{13}\text{C}_{\text{standard}}}{^{12}\text{C}_{\text{standard}}} \right] - 1 \right\} \times 1000
\]
Leaf tissue samples were collected from a subset of 92 landraces in both Puerto Vallarta and Metepec garden sites.

### 3.3.3 Statistical Analyses

#### 3.3.3.1 $G \times E$ Interactions

We used a linear mixed-effects model (R package lme4, Bates et al. (2014)) to test for significant effects (Lenth, 2012). The formula calls as fixed effects garden, continent, elevation, all interaction combinations therein, and block/garden interaction, and calls as random effects with random intercept accession line, latitudinal pair/garden interaction, and accession line/garden interaction. Traits are modeled as below:

\[
\text{TRAIT} \sim \text{GARDEN} \times \text{CONTINENT} \times \text{ELEVATION} + \text{BLOCK} : \text{GARDEN} + (1|\text{PAIR}) + (1|\text{LINE}) + (1|\text{PAIR} : \text{GARDEN}) + (1|\text{LINE} : \text{GARDEN})
\]

We compared each population’s phenotypes between field sites ($G \times E$ interactions), between highland and lowland populations from the same continent within each field site (highland-lowland local adaptation), and between Mexican and South American populations from the same elevation (local adaptation to continent-specific factors).

#### 3.3.3.2 Phenotypic Correlations

Phenotypes were transformed via Principal Component Analysis (R function prcomp) to determine relatedness between phenotypic patterns. Data were normalized via centering and scaling. Yield traits and $\delta^{13}$C were not taken from both years of the Puerto Vallarta garden site, and so were excluded from PCA.

To determine the degree to which putatively highland-adaptive traits correlate with fitness, and with each other, Pearson correlations between highland traits and fitness (FITplantveg) were calculated for each population in each common garden site. FITplantveg was used rather than FITplant because FITplantveg has less missing data. Additionally, the magnitude and direction of differences
in fitness/highland-adaptive trait correlation coefficients between sites are taken as evidence of the trait’s adaptive role between sites. Each set of correlations is specific to a population, a common garden site, and a class of pigment and macrohair patterns (either “solid” or “spot”/“margin”). The difference in trait correlation between garden sites was also calculated, giving insight into how these correlations and Genotypic vary across elevation.

Thirdly, we conducted a Finlay-Wilkinson regression (Finlay and Wilkinson, 1963), R package FW (Lian and de los Campos, 2016) to assess the stability and “adaptability” (adaptive plasticity) of landrace accessions. In a Finlay-Wilkinson regression, the mean yield of each individual landrace accession is linearly regressed to the mean yield of all landraces at each garden site. Mean yield is used as a metric of site quality or suitability (rather than yield, we used FITplantveg.) The slope of the accession’s regression is therefore equivalent to the accession’s plasticity between sites (an indicator of adaptation), and the y-intercept is the general fitness of the accession across tested environments. For each accession, the location in x/y coordinate space of slope value plotted against y-intercept value gives insight into the accession’s plasticity and fitness (Finlay and Wilkinson, 1963).

### 3.4 Results

A full accounting of the statistical significance of each contrast is provided in Table 3.2. Compensation for multiple comparisons within each trait is achieved with Bonferroni correction (alpha value is $0.05/3 = 0.0167$).

#### 3.4.1 Adjusted Fitness (FITplant and FITplantveg)

Both agronomic fitness (FITplant, Figure 3.2(a)) and vegetative fitness (FITplantveg, Figure 3.2(b)) show strong patterns of home-site advantage. Though FITplant is highest in the highlands for all populations, both traits show crossing reaction norms indicative of local adaptation.

FITplant values in PV are not significantly different from one another, though lowland populations show a modest advantage over highland populations and Mexican populations have an
advantage over South American populations. In the highland MT site, these patterns cross the significance threshold. In general, we would expect to see greater yield (and therefore greater FIT-plant) in PV due to more tropical growth conditions, but these data show the opposite trend. This is likely due to generally poorer field conditions in PV during the particular year that yield data were collected. For comparison, see Ear Weight (PM, Figure 3.2(n)).

FITplantveg are largely reflective of PE (Figure 3.2(d)). In PV, Lowland populations have higher FITplantveg than highland populations (though this difference is statistically insignificant between Mex Low and Mex High), and Mexican populations have higher fitness than SA populations. In MT, highland populations have higher FITplantveg than lowland populations, though Mex/SA differences are not significant. The only population that does not vary significantly between garden sites is Mex Low. Surprisingly, although the Mex Low reaction norm has a greater regression coefficient than Mex High, the Mex High slope is significant and the Mex Low slope is not. This is due to lower standard error within Mex High, particularly in MT (data not shown).

3.4.2 Family Traits (STD, PE, BRN, DTA, DTS, ASI)

All four populations have significantly lower Stand Count (STD, Figure 3.2(c)) in MT than in PV, though a weak pattern of home-site advantage emerges. Ear-Producing Stand Count (PE, Figure 3.2(d)) shows stronger home-site advantage. South American populations cross reaction norms, and all four populations have higher PE at their native elevation (though this trend is insignificant for Mex High). Notably, at the PV site, PE values from Mex High and Mex Low nearly converge, whereas SA High and SA Low values diverge widely. Barrenness (BRN, Figure 3.2(e)) is essentially the inverse of PE, and therefore shows patterns similar to PE (lower BRN at native elevation).

Flowering time traits Days to Anthesis (DTA, Figure 3.2(f)) and Days to Silking (DTS, Figure 3.2(g)) show that flowering takes much longer in MT. Though all populations show similar patterns, South American populations take longer to flower than Mexican populations, and lowland
populations take longer than highland populations. Anthesis/Silking Interval (ASI, Figure 3.2(h)) is generally lower in MT. The only significant ASI contrast is SA High between PV and MT.

### 3.4.3 Plant Size Traits (PH, EH, TL, TBN, EN)

Plant Height (PH, Figure 3.2(i)) and Ear Height (EH, Figure 3.2(j)) are lower in MT than in PV. Mex High is the only population that does not significantly vary between sites. Mex Low has higher PH and EH than Mex High in both sites.

Lowland populations have much higher Tassel Length (TL, Figure 3.2(k)) in PV than MT, but neither highland population varies substantially. Only SA Low varies between sites for Tassel Branch Number (TBN, Figure 3.2(l)), but there is a strong genetic effect between populations in both sites. South American populations and lowland populations have greater TBN than Mexican and highland populations.

Ear Number (EN, Figure 3.2(m)) is largely static between sites and between populations, except for SA High, which has a lower value in PV and a higher value in MT.

### 3.4.4 Yield Traits (PM, LM, GPH, DM)

Ear Weight (PM, Figure 3.2(n)), Ear Length (LM, Figure 3.2(o)), and Ear Diameter (DM, Figure 3.2(q)) are all greater in MT than in PV. These depressed PV data trends may be due in part to virus damage in the 2016 PV field.

PM and DM show crossing reaction norms indicative of home-site advantage, and Mexican populations have greater PM than South American populations from the same elevation. Lowland populations have greater LM than highland populations in PV, but values nearly converge in MT.

Kernels Per Row (GPH, Figure 3.2(p)) does not exhibit the same depression at the PV site as PM, LM, and DM. South American populations and Mex Low have lower GPH in MT, though Mex High has a strong opposite reaction norm.
3.4.5 Water Use Efficiency (d13C)

In C4 plants like maize, there is a negative correlation between water use efficiency (WUE) and $\delta^{13}$C (Ellsworth and Cousins, 2016). Lowland populations and Mex High did not vary greatly for $\delta^{13}$C (d13C, Figure 3.2(r)). Mex High has lower $\delta^{13}$C than both lowland populations in both sites. SA High shows a peculiar pattern of high $\delta^{13}$C in PV, similar to both lowland populations, and low $\delta^{13}$C in MT, similar to Mex High.

3.4.6 Anthocyanin Pigmentation and Macrohair Density (P_INTsolid, P_INTspot, P_EXTsolid, P_EXTspot, M_DENsolid, M_DENmarg)

Solid-Pattern Anthocyanin Intensity (P_INTsolid, Figure 3.2(s)) and Spot-Pattern Anthocyanin Intensity (P_INTspot, Figure 3.2(t)) increase in MT relative to PV. All four populations show similar rates of increase between sites, but the increase in P_INTspot is statistically significant for all populations, and the increase in P_INTsolid is significant only for Mex High. In all cases, SA High had the highest intensity of both patterns of pigmentation, and in all cases, highland populations had higher intensity than lowland populations.

Solid-Pattern Anthocyanin Extent (P_EXTsolid, Figure 3.2(u)) and Spot-Pattern Anthocyanin Extent (P_EXTspot, Figure 3.2(v)) increase in MT relative to PV. This increase of P_EXTspot was significant for all populations, and the increase in P_EXTsolid was significant for the highland populations. In all cases, SA High had the highest pigmentation extent of both patterns of pigmentation, and in all cases, highland populations had higher extents than lowland populations.

Leaf Sheath Macrohair Density (M_DENsolid, Figure 3.2(w)) and Leaf Sheath Margin Macrohair Density (M_DENmarg, Figure 3.2(x)) demonstrated distinct patterns. None of the populations varied significantly in M_DENsolid between garden sites. Mex High had greater M_DENsolid than Mex Low and SA High in both gardens, but otherwise, no significant differences were found.

The only population that varied significantly in M_DENmarg between garden sites was SA High. No other significant differences were found for this trait.
3.4.7 Principal Component Analyses

Traits with low missing data between the three gardens (MT, PV16, and PV17) were used to perform Principal Components Analysis (Figure 3.3). The first two components distinguish individuals from MT, PV16, and PV17. PV16 and PV17 share a higher degree of PC space overlap than either shares with MT.

High values of P_INTsolid, P_EXTsolid, DTA, and DTS characterize plants from MT. High values of several fitness-related traits and low values of M_DENsolid distinguish PV17 from PV16 and MT.

3.4.8 Pearson Correlation of Highland-Adaptive Traits

Pearson correlation values between adjusted fitness, pigment traits, and macrohair traits vary between all four populations and between both gardens.

In all cases, P_INTsolid is positively correlated with P_EXTsolid (Figure 3.4), and P_INTspot is positively correlated with P_EXTspot (Figure 3.5). Likewise, in all cases, the strength of the correlation between P_INTsolid and P_EXTsolid is the same or greater in MT than in PV (Figure 3.4). Conversely, the correlation between P_INTspot and P_EXTspot is weaker in MT than in PV (Figure 3.5).

3.4.8.1 P_INTsolid, P_EXTsolid, and M_DENsolid

The correlation between M_DENsolid and either P_INTsolid or P_EXTsolid varies between populations and gardens, but is generally either weak (positive or negative) or strongly positive. In Mex High, these correlations are strongly positive in both gardens, though stronger in PV. In Mex Low, these correlations are only strongly positive in MT. In SA Low, these correlations are only strong in PV, and in SA High, these correlations are weak in both gardens, but may be marginally positive in PV and negative in MT.

In PV, for all four populations, FITplantveg was either uncorrelated or negatively correlated with P_INTsolid, P_EXTsolid, and M_DENsolid, with the exception of a very weak positive cor-
relation between FITplantveg and P_INTsolid within Mex High (Figure 3.4(a), panel 1). Most of these negative correlations are weak (the exception being a strong negative correlation between FITplantveg and M_DENsolid within Mex Low, Figure 3.4(b), panel 1).

In MT, for all four populations, FITplantveg was positively correlated with P_INTsolid, positively correlated with P_EXTsolid (except for Mex Low), and either uncorrelated or negatively correlated with M_DENsolid. The only significant correlations listed above are found in the SA High population.

Panel 3 in each subfigure of Figure 3.4 communicates difference in correlation between PV and MT. For all populations, the correlation between FITplantveg and P_INTsolid is stronger in MT than in PV. The same is true of P_EXTsolid, except for Mex High, in which there is no change in correlation between garden sites. On the other hand, Mexican populations show an increase in correlation between FITplantveg and M_DENsolid in MT relative to PV, and South American populations show the opposite trend. Also, all four populations except for Mex Low show weaker correlation between M_DENsolid and either P_INTsolid or P_EXTsolid in MT relative to PV (though the Mex High correlation coefficients are still both significantly positive in both PV and MT).

### 3.4.8.2 P_INTspot, P_EXTspot, and M_DENmarg

In both gardens, M_DENmarg is negatively correlated with P_INTspot and P_EXTspot in South American populations, and weakly or positively correlated in Mexican populations. In PV, P_INTspot, P_EXTspot, and M_DENmarg are weakly or negatively correlated with FITplantveg in all populations. This correlation becomes positive for the lowland populations in MT, but patterns in the highland populations are mixed.

### 3.4.9 Finlay-Wilkinson Regression

The 2017 planting at PV had the highest mean vegetative plant fitness, and the 2016 PV planting and MT were nearly equal in mean fitness (Figure 3.6(a)). The landrace accessions in
each population demonstrate a range of regression slopes above and below the mean (Figure 3.6(b)-3.6(e)). When slope and deviation from mean fitness were plotted together, the four populations overlapped significantly, though the lowland populations’ distribution included more of the upper-right regions of coordinate space (Figure 3.7(a)).

3.5 Discussion

3.5.1 Local Adaptation and Plasticity

Landraces may respond to environmental changes in up to four ways: plasticity, evolution, gene flow, or extinction (Mercer and Perales, 2010). The failure of an organism to plastically adapt to all available environments promotes the evolution of adaptations to particular environments, and with those adaptations, adaptive trade-offs. When a population evolves traits that give it a home-site advantage over non-native populations, this fits the “Local vs. Foreign” model of local adaptation (Kawecki and Ebert, 2004).

All four of our elevational/continental landrace populations differ in fitness component values between our highland and lowland Mexican field sites. We observe that populations exhibit reciprocal home-site advantage in several ways. Populations grown at sites near their native elevation have higher agronomic (Figure 3.2(a)) and vegetative (Figure 3.2(b)) fitness, stand count (Figure 3.2(c)), ear-producing stand count (Figure 3.2(d)), ear weight (Figure 3.2(n)), ear diameter (Figure 3.2(q)), and lower barrenness (Figure 3.2(e)) than populations foreign to that site’s elevation, as indicated by crossing reaction norms between populations from the same continent. Other traits show evidence of home-site advantage for populations from one continent, but not the other, indicating that highland and lowland populations from different continents do not have the same adaptations, and/or that highland conditions in the Central Mexican Plateau impart different selective forces than those in South American Andes.

In several cases, populations also fit the “Home vs. Away” model of local adaptation (Kawecki and Ebert, 2004), in which a population has greater fitness in the site corresponding to their native “home” elevation than in the “away” site, regardless of the fitness of the foreign population. The
Mexican Highland and Lowland populations demonstrate this pattern most clearly when looking at ear-producing stand count (Figure 3.2(d)) and δ¹³C (Figure 3.2(r)). Though their reaction norms do not cross, both populations have higher fitness in their home sites. We might consider that, when populations meet the requirements for both models of local adaptation, there is a particularly strong case for local adaptation.

Several traits showed strong environmental effects ($E$) but minimal genotype-by-environment interaction ($G \times E$). All populations respond similarly to site effects for several traits, including days to anthesis (Figure 3.2(f)), days to silking (Figure 3.2(g)), plant height (Figure 3.2(i)), and to lesser extents, P_INTspot (Figure 3.2(t)) and P_EXTspot (Figure 3.2(v)). Others have reported that highland conditions plastically depress plant height and extend maturation (Mercer and Perales, 2018; Hufford et al., 2013), and these reports are corroborated here.

We note that our reciprocal transplant design is not fully reciprocal in that common garden sites in South American locales were unavailable. Though we may expect South American populations to exhibit higher fitness than Mexican populations in such locales, this is currently speculative.

### 3.5.2 Highland Adaptation Traits

Highland conditions present challenges for maize survival. At higher elevation, the atmosphere is thinner, leading to colder temperatures and less filtering of solar radiation. Adaptations to low temperature and high solar radiation are therefore important at high elevation.

Though many environmental conditions co-vary with altitude, the strength and direction of these correlations vary across geographic space. Precipitation and temperature correlate with altitude differently between Mexico and South America, and between lowland habitats west and east of highland ranges. In general, across Mexico, lowland conditions range from tropical to temperate, whereas highland conditions are cooler and drier (Medina et al., 1998). In South America, eastern lowlands neighbor the Amazon Basin, western coastal regions are arid, southern highlands and lowlands become drier with increasing distance from the equatorial tropics (Sarmiento, 1975). The Andean rain shadow produces geographic regions with elevational gradients of cooler, moister
highlands and hotter, dryer lowlands, across which indigenous farmers continue to cultivate maize and other crops (Brush, 1976). Because precipitation and temperature do not uniformly correlate with elevation, landraces that have evolved adaptations to high-elevation bioclimatic conditions in South America may be ill-suited for conditions found at the same elevation in Mexico.

### 3.5.2.1 Anthocyanin Pigmentation and Macrohair Density

Leaf sheath anthocyanin pigmentation and pilosity have long been reported to help plants acquire and retain heat in cold environments (Doebley, 1984; Schuepp, 1993; Lauter et al., 2004). Anthocyanin pigmentation is plastically up-regulated in response to increased light exposure (Vanderauwera et al., 2005) and cold temperatures (Christie et al., 1994; Hufford et al., 2013). We find that the intensity and extent of anthocyanin pigmentation on leaf sheaths is elevated in the highland garden site. In general, highland populations have greater overall pigmentation intensity and extent, though all populations demonstrate similar environmental effects. Also, a difference in predominance of pattern emerges between Lowland populations, wherein Mexican Lowland has greater intensity and extent of solid anthocyanin, and South American Lowland has greater intensity and extent of anthocyanin spots. The correlations between both patterns of anthocyanin and fitness appear to become more positive with increasing elevation, though solid anthocyanin pigmentation has a somewhat more positive correlation with fitness than does anthocyanin spots.

Macrohair density also exhibits plasticity, and is positively correlated with cold temperatures (Hufford et al., 2013) and with maize grain yield in cold environments (Kaur et al., 1985). Unexpectedly, leaf sheath macrohair density was largely non-plastic to the environmental variation present in this study. Leaf sheath macrohair density is greater in Mexican Highland maize than in the other populations, likely due in part to the introgression of alleles that increase macrohairs from *mexicana* into Highland Mexican maize, but no population responds plastically to garden site. The reduction in sheath margin macrohair density (particularly evident in the highland populations) is not statistically significant, except for that of South American Highland. The fitness consequences of macrohair density are different in Mexican and South American maize landraces (Figure 3.4).
The correlation between plant fitness and leaf sheath macrohair density becomes stronger at the highland site for Mexican landrace populations, but becomes more negative for South American landrace populations, indicating that leaf sheath macrohairs are adaptive for Mexican landraces in the Mexican Central Plateau highlands, but are not adaptive for South American landraces. The marginally higher density of sheath margin macrohairs among Mexican populations relative to South American populations (Figure 3.2(x)) offer some support to this conclusion. Because leaf sheath macrohair density is (weakly) negatively correlated with anthocyanin pigmentation in the highland site for South American populations, and positively correlated with anthocyanin pigmentation intensity for Mexican populations in the highland site, this macrohair/fitness correlation may simply be a reflection of the fitness consequences of anthocyanin pigmentation. On the other hand, macrohair’s negative correlation with fitness is stronger than its negative correlation with anthocyanin intensity, suggesting some relationship between macrohair density and fitness beyond conflation with anthocyanin pigment.

The only population in which (solid) anthocyanin pigmentation and leaf sheath macrohair density are strongly correlated regardless of environment is Mexican Highland. Whereas the other three populations show weaker or environmentally conditional correlation between anthocyanin and leaf sheath macrohair density, Mexican Highland maize shows these traits to be linked. This pattern could be explained by linked QTL for these traits within the well-known inversion polymorphism Inv4m (Lauter et al., 2004), introgressed from mexicana into Mexican Highland maize landraces.

### 3.5.2.2 Flowering Time/Plant Maturation

Flowering time is a complex, multigenic trait that plays a crucial role in elevation adaptation (Buckler et al., 2009; Navarro et al., 2017). Due to lower temperatures, highland regions have shorter growing seasons, so fast flowering time is a critical component of highland adaptation. Maize plants from all four populations had longer flowering time in the colder highland site, but highland populations matured more quickly than lowland populations, and this difference was pronounced in the highland site.
Positive values of ASI indicate that a plant is releasing pollen before silks are developed and receptive which can lead to incomplete pollinations and reduced yield. Positive values of ASI negatively correlates with yield (Mercer and Perales, 2018), but low and slightly negative values of ASI are likely less detrimental, as silks can remain receptive for several days, and a single plant that is shedding pollen early can pollinate many plants. For this reason, high values of ASI are typically regarded as an indicator of stress. All four populations experienced slightly higher ASI in PV (though only South American Highland varied significantly), perhaps due to virus damage in the PV site. ASI reaction norms for Mexican populations are roughly parallel, while the South American reaction norms cross. This may be because ASI is capturing the additional stress of being transplanted not just across elevations, but from one continent to another. Both South American populations have ASI values resembling Mexican populations of the same elevation when grown at their native elevation, and then deviate more strongly when grown at the alternative elevation.

3.5.2.3 Water Use Efficiency and $\delta^{13}$C

In C$_4$ plants like maize, there is a negative correlation between water use efficiency (WUE) and $\delta^{13}$C (Ellsworth and Cousins, 2016). Individuals with higher/less negative $\delta^{13}$C scores have higher ratios of $^{13}$C:$^{13}$C, meaning that they discriminate less effectively against $^{13}$C. Though the precise mechanism underlying this relationship is unclear, Avramova et al. (2019) found a region on Chromosome 7 which influences $\delta^{13}$C, WUE, and sensitivity to drought through a mechanism involving reduced abscisic acid and modified stomatal behavior. Because precipitation decreases with increasing elevation in Mexico, higher WUE may play a role in highland adaptation.

Both lowland populations show consistently high $\delta^{13}$C scores (Figure 3.2(r)), indicating low WUE. The Mexican highland population had a consistently lower $\delta^{13}$C at both sites, indicating higher water use efficiency. This finding is in accord with other published studies that detail the various drought-adapted landraces of the Mexican highlands (Eagles and Lothrop, 1994; Hayano-Kanashiro et al., 2009). Only the South American Highland population differed for $\delta^{13}$C significantly between sites. South American Highland maize, like Mexican Highland maize, had high
WUE in MT, but WUE dropped significantly in PV. This unexpected drop in WUE seen in South American Highland maize may be the result of accumulated stress from being outside its native elevation and continent, though similar extreme drops in values of other fitness-relevant traits in South American Highland maize are not observed.

### 3.5.3 Plasticity and Stability

Landrace accessions in the bottom three ninths of Figure 3.7(b) are accessions with slopes one standard deviation below the mean slope. These ninths are populated more by highland landraces than by lowland. Likewise, the top three ninths are populated mostly by lowland landraces. These results indicate that highland landraces are adapted to “unfavourable” highland conditions, and lowland landraces are adapted to “favourable” lowland conditions.

The x-axis density plot in Figure 3.7(a) also shows that more lowland landraces have yields above the mean, though this difference is less than that seen in the x-axis density plot. This suggests that lowland landraces may have on average slightly higher fitness than highland landraces.

### 3.5.4 Asymmetrical Patterns of Local Adaptation

Mercer et al. (2008) found that highland populations suffer a greater reduction in fitness in lowland conditions than lowland populations do in highland conditions. They describe this pattern as asymmetrical local adaptation. Our data do not fully replicate this finding. Our agronomic fitness data approach this pattern, with relatively stable lowland fitness and more variable highland fitness, but vegetative fitness shows an opposite asymmetry with more variable lowland populations and more stable highland populations. As Mercer and colleagues focused on agronomic fitness, these results are in alignment.

If any population is less variable between garden sites, it is Mexican Highland. Mexican Highland varies significantly across garden sites for eight traits, and all other populations vary significantly for fifteen (Table 3.2), though it is highly questionable whether such a comparison is meaningful, considering trait non-independence. Furthermore, any asymmetry of local adaptation
found here may rather be due to yearly fluctuations in $G \times E$ interactions at a site (Mercer and Perales, 2018). Further studies would be required (and are recommended) to see whether patterns of asymmetry break down or are retained over time.

### 3.5.5 Selective Forces in Maize Adaptation

Agroecosystems exert various and at times conflicting selective pressures on maize populations. Fitness is defined as an organism’s ability to survive and reproduce successfully in a particular environment. Fit maize plants must survive the myriad forces at work in the field (due to climate, elevation, soil type and quality, pest and weed pressure, as well as farmer-mediated modifications to the land, such as tilling, irrigation, fertilizer, and crop rotation) to germinate, mature, develop numerous healthy seeds, and resist post-harvest spoilage and loss. Furthermore, fit maize plants must also satisfy the desires of farmers to such a degree that the farmers will be convinced to replant the seed line in subsequent seasons. In fact, farmers more commonly report consciously selecting for culinary traits than selecting for adaptations or yield (Bellon et al., 2003). While maize plants continually evolve to balance between competing selective pressures, farmer practices and consumption patterns also evolve to maximize yield, minimize required inputs, and produce seed with desired grain type.

This complexity is liable to blur the revelation of predictable patterns of adaptation to abiotic clines like elevation. The clear patterns of adaptation to elevation found in this reciprocal transplant experiment are more impressive when considering the complicating and significant force of anthropogenic (or “artificial”) selection.

### 3.6 Conclusions

These results demonstrate local adaptation to elevation among maize landraces from Mexico and South America. Landraces adapted to diverse environmental conditions are an invaluable resource for breeding efforts that rely on fewer costly and ecologically harmful inputs (Dwivedi et al., 2016). The myriad forces that influence the *in situ* conservation status of landraces are
complex and dynamic, though locally adapted and evolving populations are more resilient and less likely to be supplanted by modern varieties (Perales et al., 2003). The importance of landraces as an agronomic resource is likely in increase due to growing global food demands as well as the effects of global climate change, which will likely substantially alter the conditions of many corn-producing regions (Bassu et al., 2014; Xu et al., 2016). In particular, drought-adapted landraces may provide adaptations suited for environments with increasingly infrequent or undependable precipitation. As climate change and the proliferation of modern inbred lines threaten the diversity and prevalence of maize landraces, the identification and preservation of landraces with particularly promising adaptations should continue to be a concern.

3.7 Acknowledgements

This research is made possible by a grant from the National Science Foundation (The Genetics of Highland Adaptation in Maize, Award Number 1546719).

3.8 References


Figure 3.1 Geology of 120 landrace accessions and common garden sites. (a) Location of collection sites of accessions and common garden sites. (b) Elevation of collection sites of accessions. Red lines indicate the elevations of the highland and lowland common garden sites. (c) Annual mean precipitation of collection sites of accessions.
Figure 3.2 Reaction norms for all measured phenotypic traits.
Figure 3.3  PCA analysis of family and plant trait values.
Figure 3.4 Pearson correlation between FITplantveg, P_INTsolid, P_EXTsolid, and M_DENsolid in Mexican Highland (a), Mexican Lowland (b), South American Highland (c), and South American Lowland (d) populations. In each subfigure, panels 1 and 2 show correlations within the lowland site (PV) and the highland site (MT), respectively. Blue shapes indicate positive correlation, red shapes indicate negative correlation, color intensity and shape size indicate strength of correlation, and asterisks indicate statistical significance (p-value thresholds = 0.001, 0.01, 0.05). Panel 3 shows the between-garden difference in correlation value for each pairwise correlation (positive/blue values indicate more positive correlations in the highland site than in the lowland site).
Figure 3.5 Pearson correlation between FITplantveg, P_INTsolid, P_EXTsolid, and M_DENsolid in Mexican Highland (a), Mexican Lowland (b), South American Highland (c), and South American Lowland (d) populations. In each subfigure, panels 1 and 2 show correlations within the lowland site (PV) and the highland site (MT), respectively. Blue shapes indicate positive correlation, red shapes indicate negative correlation, color intensity and shape size indicate strength of correlation, and asterisks indicate statistical significance ($p$-value thresholds = 0.001, 0.01, 0.05). Panel 3 shows the between-garden difference in correlation value for each pairwise correlation (positive/blue values indicate more positive correlations in the highland site than in the lowland site).
<table>
<thead>
<tr>
<th>Garden</th>
<th>PV</th>
<th>PV</th>
<th>PV</th>
<th>PV</th>
<th>MT</th>
<th>MT</th>
<th>MT</th>
<th>MT</th>
<th>PV/MT</th>
<th>PV/MT</th>
<th>PV/MT</th>
<th>PV/MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continent</td>
<td>Mex</td>
<td>SA</td>
<td>Mex/SA</td>
<td>Mex/SA</td>
<td>Mex</td>
<td>SA</td>
<td>Mex/SA</td>
<td>Mex/SA</td>
<td>Mex</td>
<td>SA</td>
<td>SA</td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>High/Low</td>
<td>High/Low</td>
<td>High</td>
<td>Low</td>
<td>High/Low</td>
<td>High/Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>FITplant</td>
<td>0.423</td>
<td>0.304</td>
<td>0.762</td>
<td>0.662</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.688</td>
<td>0.000</td>
<td>0.020</td>
<td>0.016</td>
<td>0.337</td>
</tr>
<tr>
<td>FITplantveg</td>
<td>0.628</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td>0.330</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>STD</td>
<td>0.764</td>
<td>0.000</td>
<td>0.920</td>
<td>0.001</td>
<td>0.021</td>
<td>0.157</td>
<td>0.116</td>
<td>0.543</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PE</td>
<td>0.373</td>
<td>0.001</td>
<td>0.006</td>
<td>0.151</td>
<td>0.005</td>
<td>0.070</td>
<td>0.188</td>
<td>0.820</td>
<td>0.442</td>
<td>0.067</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td>BRN</td>
<td>0.197</td>
<td>0.000</td>
<td>0.000</td>
<td>0.399</td>
<td>0.000</td>
<td>0.008</td>
<td>0.600</td>
<td>0.410</td>
<td>0.247</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
</tr>
<tr>
<td>DTA</td>
<td>0.003</td>
<td>0.172</td>
<td>0.017</td>
<td>0.410</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.126</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>DTS</td>
<td>0.008</td>
<td>0.458</td>
<td>0.003</td>
<td>0.320</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.034</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>ASI</td>
<td>0.705</td>
<td>0.278</td>
<td>0.318</td>
<td>0.730</td>
<td>0.307</td>
<td>0.198</td>
<td>0.950</td>
<td>0.027</td>
<td>0.122</td>
<td>0.030</td>
<td>0.006</td>
<td>0.740</td>
</tr>
<tr>
<td>PH</td>
<td>0.001</td>
<td>0.126</td>
<td>0.336</td>
<td>0.335</td>
<td>0.167</td>
<td>0.845</td>
<td>0.275</td>
<td>0.613</td>
<td>0.017</td>
<td>0.000</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>EH</td>
<td>0.001</td>
<td>0.290</td>
<td>0.010</td>
<td>0.807</td>
<td>0.030</td>
<td>0.185</td>
<td>0.259</td>
<td>0.838</td>
<td>0.738</td>
<td>0.009</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>TL</td>
<td>0.002</td>
<td>0.102</td>
<td>0.203</td>
<td>0.757</td>
<td>0.853</td>
<td>0.748</td>
<td>0.342</td>
<td>0.654</td>
<td>0.709</td>
<td>0.000</td>
<td>0.313</td>
<td>0.000</td>
</tr>
<tr>
<td>TBN</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.267</td>
<td>0.423</td>
<td>0.097</td>
<td>0.000</td>
</tr>
<tr>
<td>EN</td>
<td>0.384</td>
<td>0.147</td>
<td>0.026</td>
<td>0.957</td>
<td>0.760</td>
<td>0.044</td>
<td>0.089</td>
<td>0.950</td>
<td>0.335</td>
<td>0.100</td>
<td>0.000</td>
<td>0.091</td>
</tr>
<tr>
<td>PM</td>
<td>0.090</td>
<td>0.252</td>
<td>0.806</td>
<td>0.554</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.262</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>LM</td>
<td>0.000</td>
<td>0.003</td>
<td>0.177</td>
<td>0.787</td>
<td>0.160</td>
<td>0.297</td>
<td>0.052</td>
<td>0.607</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>GPH</td>
<td>0.000</td>
<td>0.047</td>
<td>0.232</td>
<td>0.000</td>
<td>0.047</td>
<td>0.001</td>
<td>0.000</td>
<td>0.348</td>
<td>0.000</td>
<td>0.000</td>
<td>0.607</td>
<td>0.008</td>
</tr>
<tr>
<td>DM</td>
<td>0.000</td>
<td>0.000</td>
<td>0.013</td>
<td>0.686</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.022</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>δ13C</td>
<td>0.007</td>
<td>0.958</td>
<td>0.004</td>
<td>0.920</td>
<td>0.000</td>
<td>0.000</td>
<td>0.589</td>
<td>0.926</td>
<td>0.206</td>
<td>0.804</td>
<td>0.000</td>
<td>0.956</td>
</tr>
<tr>
<td>P,INTsolid</td>
<td>0.550</td>
<td>0.000</td>
<td>0.044</td>
<td>0.370</td>
<td>0.172</td>
<td>0.000</td>
<td>0.129</td>
<td>0.323</td>
<td>0.004</td>
<td>0.028</td>
<td>0.026</td>
<td>0.024</td>
</tr>
<tr>
<td>P,INTspot</td>
<td>0.865</td>
<td>0.015</td>
<td>0.000</td>
<td>0.178</td>
<td>0.479</td>
<td>0.007</td>
<td>0.015</td>
<td>0.535</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>P,EXTsolid</td>
<td>0.246</td>
<td>0.002</td>
<td>0.341</td>
<td>0.398</td>
<td>0.058</td>
<td>0.000</td>
<td>0.455</td>
<td>0.384</td>
<td>0.006</td>
<td>0.032</td>
<td>0.010</td>
<td>0.023</td>
</tr>
<tr>
<td>P,EXTspot</td>
<td>0.331</td>
<td>0.004</td>
<td>0.076</td>
<td>0.990</td>
<td>0.024</td>
<td>0.036</td>
<td>0.256</td>
<td>0.133</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>M,DENsolid</td>
<td>0.000</td>
<td>0.485</td>
<td>0.000</td>
<td>0.340</td>
<td>0.000</td>
<td>0.262</td>
<td>0.000</td>
<td>0.242</td>
<td>0.408</td>
<td>0.231</td>
<td>0.102</td>
<td>0.411</td>
</tr>
<tr>
<td>M,DENmarg</td>
<td>0.311</td>
<td>0.433</td>
<td>0.189</td>
<td>0.231</td>
<td>0.758</td>
<td>0.646</td>
<td>0.170</td>
<td>0.080</td>
<td>0.063</td>
<td>0.256</td>
<td>0.005</td>
<td>0.107</td>
</tr>
</tbody>
</table>
Figure 3.6 Finlay-Wilkinson regression of individual landrace FITplantweg fitnesses to mean FITplantweg at each environment (MT, PV2016, PV2017). These lines are plotted all together (a), as well as divided by population ((b), (c), (d), (e)).
Figure 3.7 Relationship between Finlay-Wilkinson regression slope and deviation from mean FITplantveg. (a) Relationship between regression slope and y-intercept for each landrace. Ellipses capture 90% of each population’s lines. Density plots show the distribution of data points along each axes for each population. Dashed lines demark one standard deviation to the left and to the right of the mean of the cumulative distribution, creating ninths in the coordinate space. (b) The number of landrace lines from each population that fall within each ninth from (a).
CHAPTER 4. ECOLOGICAL NICHE ANALYSES OF MAIZE LANDRACES

Authors: Garrett M Janzen\textsuperscript{1}, Daniel Gates\textsuperscript{2}, Daniel E Runcie\textsuperscript{2}, Jeffrey Ross-Ibarra\textsuperscript{2}, Matthew B. Hufford\textsuperscript{1}

\textsuperscript{1}Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA
\textsuperscript{2}Department of Plant Sciences, University of California-Davis; 278 Robbins; Berkeley, California, USA 95616

4.1 Abstract

Maize landrace designations reflect both biological and cultural delineations. To determine the degree to which landrace names represent true units of biological diversity, we performed comparative studies on maize landrace accessions partitioned both by landrace name and genetic similarity. Ecological Niche Modeling (ENM) was used to determine environmental variable importance, niche breadth, and niche overlap between landraces and between genetic groups. Niche similarity is highly positively correlated with genetic similarity, supporting genetic bases for modeled niches. We find that highland groups (landraces and genetic groups) have narrower niches than lowland groups, and that this effect is stronger than expected given neutral expectations. Though landrace models and genetic group models predict occurrence points with roughly equitable accuracy, mixed model equations find that additive genetic variance explains more environmental variance than does the landrace identity matrix, suggesting that population genetic structure below the level of
landrace is important for local adaptation, and that environmental associations may be due to fewer large-impact genetic loci which are not reflected in genetic clustering methods. Environmental Association Analyses (EAA) identify candidate SNPs (many of which are uncharacterized) strongly associated with environmental variables that may play important roles in abiotic adaptation.

4.2 Introduction

Maize (\textit{Zea mays} subsp. \textit{mays}) is a model system of high agronomic (Shiferaw et al., 2011; Hake and Ross-Ibarra, 2015), economic (Shiferaw et al., 2011; Ranum et al., 2014), cultural (Fernandez Suarez et al., 2013; Perales, 2016), and scientific (Dumas and Mogensen, 1993; Fedoroff, 2001; Stern et al., 2004; Hake and Ross-Ibarra, 2015) value, and as such, a wealth of resources have been developed for the study of maize. Maize was domesticated in the lowlands of the Balsas River Valley in Mexico from the teosinte taxon \textit{Zea mays} subsp. \textit{parviglumis} roughly 9000 years BP (Matsuoka et al., 2002). From there, maize was carried across across North America into South America as early as 6000 years BP (Bush et al., 1989; Grobman et al., 2012), north into present-day United States by about 4500 years BP (Merrill et al., 2009), and around the world (Tenaillon and Charcosset, 2011; Van Heerwaarden et al., 2011; Hake and Ross-Ibarra, 2015). Presently, maize is grown under a broad range of temperatures, precipitations, and soil types, and a wider range of elevations and latitudes than any other crop (Ruiz Corral et al., 2008; Shiferaw et al., 2011).

Adaptation to divergent selective ecological pressures is one of the major modes of speciation (Schluter, 2001; Sobel et al., 2010; Nakazato et al., 2010), and the same is true of maize landrace formation (Eagles and Lothrop, 1994; McCouch, 2004). At points along the range expansion of maize, farmers selected plants and seed desirable for human applications and consumption and suitable for growth in their local environment. Over generations of propagation and selection, this process formed varietal populations called landraces. These landraces are grown and maintained by smallholder farmers to the present day as dynamic, evolving populations (Mijangos-Cortes et al., 2007; Dyer and López-Feldman, 2013) that diverge despite considerable gene flow (Ortega, 1995; Pressoir and Berthaud, 2004). Gene flow is typically a constraint to local adaptation, but
populations with unstable structure, large-scale geographic range shifts, and/or frequent extinction and recolonization events (as may be the case when seed types are abandoned or swapped between farmers) may benefit from gene flow (Slatkin, 1987; Dyer and Taylor, 2008).

Landraces are a critically important source of corn product across Mexico. More than 75% of the maize produced in Mexico comes from small (3 or fewer tons per hectare) farms (Bellon et al., 2018), and most of the arable land in Mexico (up to 80% (Louette et al., 1997)) is devoted to growing landrace maize. Landraces are frequently out-yielded by modern hybrids, but in their own home environments, landraces can and often do out-perform modern lines (Bellon et al., 2018; Perales, 2016; Bellon et al., 2003; Dyer and Taylor, 2008; Mercer and Perales, 2018)). Due to their distinctive ear morphologies and unique culinary qualities, some landraces fill niche market demands, safeguarding them from abandonment and replacement by agronomically-superior modern hybrids (McLean-Rodríguez et al., 2019). About 31.6% of the land area of Mexico is “Marginally Suitable” (compared to 6.4% “Suitable”, 25.1% “Moderately Suitable”, and 36.9% “Not Suitable”, (Monterroso Rivas et al., 2011)), and it is in marginally-suitable areas that local landraces have an adaptive edge over modern improved lines (Ceccarelli, 1994). Conversely, the areas of highest landrace diversity are those projected to suffer the least under climate change (Monterroso Rivas et al., 2011), though almost every landrace is expected to incur some level of range contraction in the next 10 to 30 years (Ureta et al., 2012). As the “Suitable” land of Mexico is projected to be most adversely affected and reduced by future climate change, and because areas of high landrace richness are projected to be largely spared by climate change, landraces are positioned to become even more important in the future.

Furthermore, landrace populations are an important reservoir of genetic diversity that has been largely purged from inbred breeder lines (McCouch, 2004). Efforts by breeders to incorporate landrace highland Mexican germplasm into improved hybrids began in the 1940s and continue to the present day, and have produced lines adapted to temperate regions around the world (Eagles and Lothrop, 1994; Varshney et al., 2018). Highland maize landraces have traits like pilose, pigmented stems and droopy, leathery ears, which are adaptive to cold and low precipitation environments,
and resistance to field pests (Eagles and Lothrop, 1994). Some highland lines express the latente ("latency") trait, in which the plant avoids drought stress damage by halting growth during drought conditions and resuming growth quickly upon rehydration (Eagles and Lothrop, 1994; Hayano-Kanashiro et al., 2009). Better adapted crops are one of the least environmentally harmful and most secure strategies for increasing productivity and security of worldwide food supply (Byerlee, 1996).

Though the value of maize landraces is increasingly recognized and is projected to increase over time, many landraces are threatened with replacement and subsequent loss. Landraces are increasingly replaced in by modern hybrids, though landraces persist in areas of marginal environments (Ceccarelli, 1994) and marginal social/economic/technological infrastructure (Byerlee, 1996; de la Barrera and Martínez, 2018; McLean-Rodríguez et al., 2019; Dyer and Taylor, 2008). Numerous national and international agencies engage in in situ and ex situ conservation of maize landraces and wild relatives (Cohen et al., 1991). When prioritizing populations for conservation efforts, a host of criteria are considered, including potential economic value, importance to research, importance to culture, importance to ecosystem functions, current geographic distribution, current abundance within that distribution, risk of distribution and/or abundance decline due to environment change caused by land use or climate change, endemism, and genetic/geographic/environmental relatedness to other accessions (Bellon et al., 2003; Perales et al., 2003; Barazani et al., 2008; Burke et al., 2009; Ramírez-Villegas et al., 2010; Phillips et al., 2016). Though ex situ conservation is an indispensable tool in most conservation programs, in situ conservation is usually regarded as superior when feasible, as open-pollinated landrace populations maintained by traditional farming methods by smallholder farmers promote adaptive evolution by maintaining large effective populations, preserving standing genetic diversity, and permitting maize populations to adapt to diverse and changing environments (Bellon et al., 2018).

Agrodiversity is frequently analyzed and estimated at the level of named landraces (Ruiz Corral et al., 2008; Burke et al., 2009; Monterroso Rivas et al., 2011; Ureta et al., 2012; de la Barrera and Martínez, 2018). Landrace diversity is also espoused when discussions turn to the threat of climate
change (Burke et al., 2009; Ureta et al., 2012), which is projected to result in major yield losses (Challinor et al., 2014). However, as stated by Arteaga et al. (2016), a maize landrace is not a static group, but rather an “open and evolving genetic system.” While some landraces are narrowly distributed and share a narrow genetic base, others are more cosmopolitan and have high intra-race genetic and/or morphological variability (Arteaga et al., 2016; Van Heerwaarden et al., 2011; Pressoir and Berthaud, 2004). Some landraces are highly adapted to a single or a few habitats, while others are generalists and are tolerant to a wide breadth of environmental conditions. Seed trade between regions is common, and landrace names are flexibly applied to varieties that possess particular traits, sometimes regardless of origin or phylogeny (Dyer and Taylor, 2008). Therefore, landrace names may not be reliable indicators of biologically meaningful delineations (Goodman and Bird, 1977). In some cases, landrace names are more granular than the maize lines they are applied to, effectively separating a population into several small populations distinguished by delineations in a single trait of interest to farmers. In other cases, a landrace name can be broadly applied to a geographically-diverse group of related yet uniquely locally-adapted lines, conflating them into one entity. For this reason, if capturing and preserving desired genetic diversity is the end goal of ex situ conservation, using landrace name as the guiding criterion in conservation prioritization is of questionable utility. Other criteria focus on density of accessions collected in geographic space or diverse environmental conditions (in addition to taxonomic representation (Ramírez-Villegas et al., 2010; Burke et al., 2009)). Barazani et al. (2008) include geographic distribution, abundance, endemism, and rarity of suitable habitat among their list of seven criteria in selecting crop biodiversity for conservation. These criteria may be better heuristics of adaptive genetic diversity than landrace name when prioritizing populations for conservation and sampling.

In this study, we investigate the degree to which maize landrace designations accurately reflect patterns of local adaptation to abiotic environment. We constructed and compared the ecological niche models (ENMs) of Mexican maize landrace accessions grouped by landrace name or by genetic similarity. Metrics of niche similarity were compared to both published phylogenetic relationships and genetic clustering results (fastStructure, $F_{ST}$) to determine the degree to which genetic relat-
edness and/or similarity result in ENM similarity. From the predictive performance of each ENM, we find the degree to which each group’s distribution is constrained by particular environmental variables (i.e., the degree of adaptation to abiotic environment). From the breadth of each ENM, we determine where individual landraces fall along the gradient between generalists and specialists. We also used mixed effect models to determine whether landrace name or genetic similarity better predict the environmental conditions at which landrace individuals grow. To identify genetic variants involved in adaptation, we conducted Environmental Association Analyses (EAA) to find SNPs that associate most strongly with environmental variables that define ecological niches or constrain geographic distribution. We find strong but variable signals of local adaptation among landraces and comprised genetic groups, and that genotype in general is a better predictor of association with environment than landrace name.

4.3 Methods

4.3.1 Accession Data

Our accessions’ genotype (DArTSeq SNPs) and passport data were taken from the International Maize and Wheat Improvement Center (CIMMYT) Seeds of Discovery (SeeDs) dataset (Pixley et al., 2017). Accessions were selected from the dataset if they came from landraces with 30 or more represented individuals from Mexico that passed genotype quality filters (less than 20% missing data) and included latitude and longitude passport data. In total, 989 accessions were retained (40 Bolita (BOLITA), 100 Celaya (CELAYA), 96 Chalqueño (CHALQU), 229 Cónico (CONICO), 91 Cónico Norteño (CONNOR), 38 Olotillo (OLOTIL), 41 Pepitilla (PEPITI), 49 Tabloncillo (TABLON), 274 Tuxpeño (TUXPEN), and 31 Vandeño (VANDEN)). Additionally, longitude and latitude coordinates for maize wild relatives *Zea mays* subsp. *parviglumis* and *Zea mays* subsp. *mexicana* (*parviglumis* (PARVIGLUMIS) and *mexicana* (MEXICANA) hereafter) were obtained from Hufford et al. (2012).

We used fastStructure (Raj et al., 2014) to sort the 989 maize landrace individuals into genetic groups. Optimal $k$ values were returned by the fastStructure utility tool *chooseK*. Population
structure is explained at $k = 3$ groups, but marginal likelihood is maximized at $k = 2$ groups. Therefore, analyses proceeded with both $k$ values, and a total of five non-independent genetic groups were defined for analyses. Genetic groups were populated with individuals with 75% or greater attribution to the $k$ group. Group names here are identified by $k$ value (2 or 3) followed by group within the $k$ value (at $k = 2$, Group2_1 and Group2_2, and at $k = 3$, Group3_1, Group3_2, and Group3_3). We use the term “occurrence group” to refer to landraces, genetic groups, and teosinte subspecies.

Neutral genetic divergence (mean $F_{ST}$ (Weir and Cockerham, 1984)) between all landraces was calculated with vcftools (Danecek et al., 2011). Population structure within the two most populous landraces, Cónico and Tuxpeño, was estimated with sparse nonnegative matrix factorization (R function snmf (Frichot and François, 2015)). Clusters of environmental Euclidean distance within each landrace were estimated with a consensus of 30 clustering indices (R function NbClust (Charrad et al., 2014)).

### 4.3.2 Environmental Data

Climatic data (bioclimatic variables, solar radiation, wind speed, and water vapor pressure) and Shuttle Radar Topography Mission elevation data were downloaded from WorldClim (Hijmans et al., 2005) at 30 seconds (approximately 1 km$^2$) resolution. Solar radiation, wind speed, and water vapor pressure values were each downloaded as twelve monthly average rasters. Soil data (bulk density, cation exchange capacity of soil, weight percentage of clay particles, soil organic carbon content, pH index measured in water solution, weight percentage of silt particles, and weight percentage of sand particles) were downloaded from ISRIC (World Soil Information database, http://www.isric.org/; (Hengl et al., 2014)) at 1 km resolution and at various soil depths (0 m, 0.3 m, 1 m, and 2 m), as maize roots can grow to depths as low as 2 meters (Dardanelli et al., 1997). In keeping with Bandillo et al. (2017), for each soil variable, we averaged values across depths to return a single value. Each of these variables (including the mean soil variables) were
then standardized to a mean of 0 and standard deviation of 1. This returned 63 environmental variables for consideration.

Non-independence of predictors in ENM will bias results and produce misleading results (See Discussion). Two methods were employed to reduce non-independence of environmental predictors.

First, all 63 environmental layers were PCA transformed (function rasterCPA, r package RStoolbox). The first five principal components (EnvPC1, 2, 3, 4, and 5) were used as environmental variables in the first iterations of ENM construction (Figure 4.1). EnvPC1 is most strongly defined by positive correlation with (high values of) minimum temperature of the coldest month, mean temperature of the coldest quarter, water vapor pressure, annual precipitation, and negative correlation with (high values of) annual temperature range and seasonality. EnvPC2 is most strongly defined by positive correlation with high values of altitude, spring and summer solar radiation, water vapor pressure, and isothermality, and negative correlation with high summer temperatures and water vapor pressure. EnvPC3 is most strongly defined by positive correlations with (high values of) wind speed, and water vapor pressure, and silty soil, and negative correlations with solar radiation from spring to fall. EnvPC4 is positively correlated with solar radiation, wind speed, and weight percentage of sand particles, and negatively correlated with weight percentage of silt particles. EnvPC5 is negatively correlated with precipitation of the coldest, warmest, and driest quarters, precipitation of the driest month, and spring/summer solar radiation.

Second, we dropped colinear (non-informative) environmental rasters from inclusion. To reduce the number of environmental variables considered in ENM construction, we calculated Pearson correlation between all scaled variables, and clustered them hierarchically (Figure 4.2(a)). Because monthly rasters for solar radiation and water vapor pressure varied significantly between months and clustered into different groups, select consecutive months were averaged into rasters of seasonal values. With these new environmental rasters, collinearity was calculated again, this time at a higher cutoff value of 0.85, returning eighteen clusters (Figure 4.2(b)). One variable from each cluster was then selected for ENM construction.
Solar radiation values were split between several correlation groups (March, Spring, June, Summer, October, and Winter, Figure 4.2(a)), but values for March and Summer (July through September) were selected for inclusion (Figure 4.2(b)). Water vapor pressure clustered into two groups (July through August and September through June, Figure 4.2(a)), which were dropped in favor of correlated traits Altitude and Mean Temperature of the Coldest Quarter, respectively (Figure 4.2(b)). (We note that altitude/elevation is a geographic rather than environmental variable.) Wind speed clustered entirely into one group (Figure 4.2(a)), so an annual mean wind speed raster was created.

4.3.3 Ecological Niche Modeling

For each occurrence group, suitability scores were predicted across Mexico, based on ENMs built either from the sufficiently uncorrelated standardized environmental predictors or the PCA-transformed environmental predictors. Hereafter, we refer to the ENMs built from PCA-transformed environmental variables as the PCA ENMs, and the ENMs built from the environmental variables as the Environmental ENMs. The percent contribution of each predictor to the model was calculated. Heatmaps and dendrograms illustrate niche overlap (Schoener’s $D$ (Schoener, 1968)) between landraces and between genetic groups. Environmental ENM and PCA ENM niche overlap values were averaged, and groups with mean niche overlap greater than or equal to 0.6 were considered similar enough to be deemed identical, in keeping with published literature (Wallace Jr, 1981; Zaret and Rand, 1971; Mathur, 1977).

ENMs were constructed using the Maximum Entropy (Maxent) algorithm. Maxent operates on the principle of favoring maximal entropy, estimating the most uniform distribution possible that conforms to provided constraints (Phillips et al., 2004, 2006; Phillips and Dudík, 2008), and has a very good track record in ENM applications (Baldwin, 2009; Costa et al., 2010).

We subsampled 30 individuals from each occurrence group with reference to a grid (function gridSample, R package dismo (Hijmans et al., 2013)), in which multiple samples could not be drawn from the same cell. Mean pairwise Euclidean distances between all occurrence points of the
30 sampled individuals were calculated a metric of spatial disaggregation. The 30 individuals were partitioned into testing and training datasets with k-fold data partitioning ($k = 5$, function kfold, R package dismo). A composite ENM was constructed from the average of the five k-fold replicate models for Environmental ENMs, and were evaluated with the area under the receiver operating curve (ROC), or AUC. Percent contribution of each predictor to the composite ENMs is reported.

To test ENM performance and comparative metrics, we constructed ENMs from ten “false” groups, using the same methods as described above. Three false groups were constructed of random assortments of landrace accessions (PSEUDO1, PSEUDO2, PSEUDO3), three from random points across Mexico (RANDOM1, RANDOM2, RANDOM3), and four from random points drawn from one of the cardinal direction quadrants across Mexico (RANDOMNW, RANDOMNE, RANDOMSW, RANDOMSE).

In an attempt to assess the role of agriculture on maize landrace niche development, also we constructed ENMs for *mexicana* and *parviglumis* following the methods previously described. We reasoned that distributions of wild relatives would be less influenced by human intervention, and would therefore have a tighter association with bioclimatic variables, resulting in higher AUC and lower niche breadth.

Values of dedicated niche breadth metrics (B1 and B2 (Levins, 1968)) were calculated for each occurrence group. B1 measures the uniformity of the distribution of a group’s occurrence points among available resource states. B1 is equivalent to the reciprocal of the Simpson Diversity Index (Simpson, 1949), ranging from 1 to $n$, where $n$ is the total number of available resource states, and is maximized when occurrence groups are dispersed evenly across resource states. B2 normalizes the value of B1 by the number of available resource states (Hurlbert, 1978).

Though AUC is frequently viewed and used as a metric of model performance at distinguishing presence points from background points, this practice is not recommended (Lobo et al., 2008; Golicher et al., 2012). At the least, AUC should not be used to compare models across disparate studies with differences in magnitude of scale of geographic extent, number of samples, heterogeneity and resolution of environmental predictors, or other key parameters. Rather, the
true utility of AUC (in the context of ENM) is that it provides information about the modeled species/population/occurrence group’s niche. In the context of this analysis, an occurrence group’s mean AUC is taken to represent the model’s capacity to predict the distribution of a group, given environmental data. Therefore, AUC of Environmental ENMs and PCA ENMs (which we refer to as Environmental AUC and PCA AUC, respectively) are here interpreted a relative metrics of adaptation to environmental conditions, reflective of niche breadth.

A fourth measure of Niche Breadth, Species Specificity Index, was calculated from data published by Ruiz Corral et al. (2008). Maize landraces from across Mexico were surveyed and the number of habitat classes (as described by Medina et al. (1998)) that each landrace was found in were counted. Devictor et al. (2008) developed the specialization index of Julliard et al. (2006) for presence/absence occurrence data with the following equation:

\[ SSI = \sqrt{\frac{1}{h/H}} - 1 \]

in which \( h \) is the number of habitat classes a group is found in and \( H \) is the total number of available classes. We used the survey from Corral and colleagues to calculate the SSI of each included landrace.

To determine the impact of spatial disaggregation and altitude on niche breadth, pairwise regressions between spatial disaggregation, altitude, and niche breadth were conducted. Occurrence groups that significantly influenced regression slopes were identified by Studentized residuals, hat values, and Cook’s Distance (R package car (Fox and Weisberg, 2018)). Regressions from false groups were treated as neutral expectations, and difference in slope between the real and false groups was quantified by least square means (R function lstrends, R package lsmeans (Lenth and Lenth, 2018)).

### 4.3.4 Niche Breadth Correlations

Pairwise distance matrices constructed from genetic distance, geographic distance, altitudinal distance, and projected suitability values for Environmental ENMs and PCA ENMs were compared via partial Mantel tests (R package ecodist (Goslee et al., 2007)). Suitability distance matrices
were constructed from the difference in suitability scores of each model at the presence location of every occurrence point. Likewise, altitudinal distance was constructed from difference in altitude between individuals, geographic distance was constructed from pairwise distance in meters, and different aspects of genetic distances were estimated from difference in percent attribution to genetic Groups2, 1, 3, and 3.

4.3.5 Mixed Effect Models

We employed mixed-effect models (function mmer2, R package sommer (Covarrubias-Pazaran, 2016)) to estimate the variance components of genetic relatedness and landrace name in the prediction of an accession’s home environment.

The equation was fit with the following equation:

\[ y = xG + xL + xD1 + xD2 + xD3 \]

in which environmental variable \( y \) is predicted by five instances of an \( n \times n \) symmetric accession matrix \( x \), each of which is granted a different variance-covariance matrix reflecting genetic similarity (\( G \)), landrace identity (\( L \)), and close-range (\( D1 \)), mid-range (\( D2 \)), and far-range (\( D3 \)) distance similarity. This model was used to predict two datasets of environmental variables. The first is the set of predictors used in ENM (which we refer to as the predictor dataset), and the second is environmental conditions independently recorded in the field from the accessions’ home field (which we refer to as the independent dataset). The top five Principal Components of the independent dataset were included as well (Figure 4.3). To test the effect of summarizing genetic similarity into categorical genetic groups, we estimated genetic relatedness (\( G \)) in three ways: 1) a one-hot encoded matrix indicating categorical fastStructure genetic group (\( k = 9 \)), 2) a Radial Basis function kernel (function kernelMatrix, R package kernlab (Karatzoglou et al., 2018)) of fastStructure ancestry coefficients (\( k = 3 \)), or 3) Identical by state (Centered IBS) kinship matrix (created with Tassel5 (Bradbury et al., 2007)). \( D1, D2, \) and \( D3 \) were Radial Basis kernel matrices of latitude and longitude values with sigma values of low, median, and high sigma estimates (function sigest, R
package kernlab (Karatzoglou et al., 2018)) to capture environmental variance due to close-range, mid-range, and long-range distance similarity, respectively.

Variance components and Z-ratios (ratio of variance component to standard error) of genotype and landrace name were estimated for both the predictor environmental dataset and the independent environmental dataset, and are reported separately for each formulation of $G$. Z-ratio values greater than 2 are generally considered significant. Model quality is reported in log-likelihood and AIC. Significance of differences in Z-ratios for genotype and landrace name were tested with paired t-tests.

4.3.6 EAA

Environmental Association Analyses for the 18 environmental rasters in the predictor dataset were conducted with LFMM (Frichot et al., 2013). LFMM is an MCMC algorithm that detects correlations between genetic variants and environmental variables by first modeling $k$ population structure groups as latent factors, lowering the false positive rate. Analyses were run with $k = 3$ ancestral population latent factors with 6000 cycles with a burn-in period of 3000 cycles. SNP $p$-values were adjusted with the Bejamini-Hochberg procedure (Benjamini and Hochberg, 1995) to reduce the false discovery rate. Significant genes were then selected for gene ontology analysis.

4.4 Results

4.4.1 Population Genetic Structure

fastStructure identifies either two (prioritizing marginal likelihood, Figure 4.4(a)) or three (explained population structure, Figure 4.4(b)) populations. Group 2_1 ($k = 2$, blue) includes most individuals from Olotillo, Tuxpeño, and Vandeño, and Group 2_2 ($k = 2$, red) includes most individuals from Chalqueño, Cónico, and Cónico Norteño (Figure 4.4(c)). Bolita, Celaya, Pepitilla, and Tabloncillo show significant admixture between these two groups. This composition changes at $k = 3$ genetic groups (Figure 4.4(d)). Group 3_1 ($k = 3$, green) includes most individuals of
Tabloncillo, while 3.2 corresponds to 2.1 and 3.3 corresponds to 2.2. Bolita, Celaya, and Pepitilla remain largely admixed between the three.

Hierarchical clustering grouped landraces by $F_{ST}$ (Figure 4.4(e)). Chalqueño, Cónico, and Cónico Norteño form one main branch. The second main branch splits Olotillo, Tuxpeño, and Vandeño from Bolita, Celaya, Pepitilla, and Tabloncillo.

Geographic distance between accessions is strongly correlated with difference in percent attribution in Group 2.1 and Group 2.2 ($r = 0.093, p = 0.001$), Group 3.1 ($r = 0.17, p = 0.001$), and Group 3.2 ($r = 0.18, p = 0.001$), but is weakly negatively correlated with Group 3.3 ($r = -0.013, p = 0.17$). When difference due to geographic distance are accounted for, difference in altitude is strongly correlated with difference in percent attribution in Group 2.1 and Group 2.2 ($r = 0.64, p = 0.001$), Group 3.2 ($r = 0.54, p = 0.001$), and Group 3.3 ($r = 0.58, p = 0.001$), but weakly negatively correlated with Group 3.1 ($r = -0.015, p = 0.14$). This indicates latitude/longitude and altitude influence percent attribution at $k = 2$ and $k = 3$, but at $k = 3$, Group 3.1 is not strongly correlated with altitude, and Group 3.3 is not strongly structured across latitude or longitude (i.e., has low signature of isolation by distance).

4.4.2 Ecological Niche Models

4.4.2.1 Niche Characterization

ENMs were constructed for each occurrence group (Figures 4.5 and 4.6), and variable importance of each environmental layer was calculated and compared (Figures 4.7 and 4.8). Altitude was the top predictor for Celaya, Chalqueño, Cónico, Cónico Norteño, and Mexicana. After altitude, these five groups were generally predicted by temperature seasonality, mean temperature of the coldest quarter, mean temperature of the warmest quarter, and/or soil pH. Wind speed was the top predictor for Pepitilla, Tabloncillo, and Parviglumis, though these three groups do not share many other environmental variables consistently in common. Olotillo and Tuxpeño were strongly predicted by annual precipitation. Bolita was very strongly predicted by summer solar radiation
only, and Vandeño was strongly predicted by mean temperature of the coldest quarter, precipitation seasonality, and temperature seasonality.

Unsurprisingly, the top predictors of the genetic groups are approximated by those of the landraces that comprise them. Groups 2_1 and 3_2 were predicted by annual precipitation (like several landraces, but especially Tuxpeño and Olotillo). Group 2_2 was predicted by altitude (like Celaya, Chalqueño, Cónico, Cónico Norteño). However, the corresponding k-group Group 3_3 was mainly predicted by mean temperature of the warmest quarter, which is only of moderate importance in Chalqueño and Cónico. Group 3_1, comprised almost entirely by Tabloncillo and Pepitilla individuals, is predicted mostly by mean wind speed, like Tabloncillo and Pepitilla.

The top predictors were also found for the groups in PCA ENMs. Bolita, Celaya, Chalqueño, Cónico, Cónico Norteño, Group 2_2, and Group 3_3 are mostly predicted by EnvPC2. Olotillo, Tuxpeño, Vandeño, Group 2_1, and Group 3_2 are mostly predicted by EnvPC1. Pepitilla is predicted by EnvPC1 and EnvPC2 in nearly equal measures. Tabloncillo and Group 3_1 are mostly predicted by EnvPC3.

Niche overlap (Shoener’s D) between each landrace and between each genetic group was calculated, and heatmaps and dendrograms produced to illustrate similarity (Figures 4.9 and 4.10). Niche overlap heatmaps were averaged between environmental data and PCA-transformed environmental data. Niche overlap dendrograms group landraces with similar niches. The first branch point separated Pepitilla, Olotillo, Vandeño, Tuxpeño, and Tabloncillo from Cónico, Chalqueño, Bolita, Cónico Norteño, and Celaya. In keeping with other literature (Wallace Jr, 1981; Zaret and Rand, 1971; Mathur, 1977), niche identity of 0.6 or higher was considered evidence of niche identity. Only Celaya and Chalqueño met the 0.6 niche overlap threshold of niche identity.

The corresponding k = 2 and k = 3 genetic groups had high niche overlap. Group 2_2 and Group 3_3 had high niche overlap, as did Group 2_1 and Group 3_2, and these overlap scores met the 0.6 overlap threshold. Group 3_1 clustered less strongly, but paired most closely with Group 2_1 and Group 3_2. A tangle plot (Figure 4.11) indicates high alignment of FS T dendrogram with the niche similarity dendrogram (entanglement coefficient = 0.09)
4.4.2.2 Niche Breadth

To quantify niche breadth, we considered the interconnectivity between altitude (Figure 4.12), spatial disaggregation, AUC and PCA AUC (Figures 4.13 and 4.14), and three other formulations of niche breadth (B1, B2, and SSI). These values are provided in Table 4.1, and the relationship between these values is shown in Figure 4.15.

The distributions of AUC of occurrence groups and mean AUC across classes of occurrence groups (landraces, genetic groups, wild relatives, and random) were calculated (Figure 4.13). On average, landrace AUC (Environmental AUC, 0.93; PCA AUC, 0.91) was higher than genetic group AUC (Environmental AUC, 0.86; PCA AUC, 0.84, Figure 4.13). Both were lower than wild relative mean AUC (Environmental AUC, 0.97; PCA AUC, 0.98) and lower than random group mean AUC (Environmental AUC, 0.71; PCA AUC, 0.72). However, these comparisons do not account for other covariates. Environmental AUC, which is correlated with niche specificity, is positively correlated with altitude ($r^2 = 0.329$, $p = 0.45$, Figure 4.16(a)) and negatively correlated with spatial disaggregation ($r^2 = 0.61$, $p = 0.04$, Figure 4.16(a)). Similar patterns emerge for PCA AUC ($r^2 = 0.30$, $p = 0.001$, Figure 4.16(a); $r^2 = 0.59$, $p = 0.56$, Figure 4.16(a), respectively).

Values of B1 and B2 reflect Environmental AUC and PCA AUC. Wild relatives mean niche breadth (B1 = 0.88, B2 = 0.13) are lower than landrace mean niche breadth (B1 = 0.96, B2 = 0.52), genetic group mean niche breadth (B1 = 0.98, B2 = 0.62), and random groups (B1 = 0.99, B2 = 0.83). Because B1 and B2 are strongly correlated ($r^2 = 0.89$, data not shown), and because B2 is a conceptual improvement on B1, we continue analyses with B2.

Partial Mantel tests find that the PCA ENM suitability distance matrix and Environmental ENM suitability distance matrix are highly correlated but not identical ($r = 0.66$, $p = 0.001$). Suitability distance matrices are most correlated with the altitude distance matrix (Environmental ENM, $r = 0.21$, $p = 0.001$; PCA ENM, $r = 0.45$, $p = 0.001$), followed by genetic distance matrix (Environmental ENM, $r = 0.14$, $p = 0.001$; PCA ENM, $r = 0.14$, $p = 0.001$) and then geographic distance (Environmental ENM, $r = 0.023$, $p = 0.066$; PCA ENM, $r = 0.10$, $p = 0.001$).
Average relative niche breadth (Figure 4.15) indicates that Chalqueño and Cónico have narrow niches, and Tuxpeño and Vandeño have broad niches. Bolita, Celaya, Cónico Norteño, Olotillo, Pepitilla, and Tabloncillo are of intermediate niche breadth, and the wild relatives Mexicana and Parviglumis have very narrow niches. Within genetic groups, Group 2.2 and 3.3 have narrow niches, Group 2.1 and Group 3.2 have broad niches, and Group 3.1 is intermediate. Visually, average relative niche breadth generally tracks with altitude and spatial aggregation. (The additive inverse of) B1 generally underestimates niche specificity, and AUC generally overestimates niche specificity. Bolita, Pepitilla, and Vandeño are groups in which Environmental AUC/PCA AUC and B1/B2 diverge particularly strongly. Altitude tracks with niche narrowness generally well, except with Olotillo and Parviglumis, which have narrow niches despite low altitude.

Influence plots identify occurrence groups with niche breadth values that are outliers from the regression with altitude. Given their altitudes, Group 2.1 and Tuxpeño have lower AUC than expected (Figure 4.17(f)), Groups 2.1, 3.2, and 3.3 have lower PCA AUC than expected (Figure 4.17(i)), Parviglumis has higher PCA AUC (Figure 4.17(l)), and Parviglumis and Olotillo have lower B2 (Figure 4.16(l)). This is evidence that Parviglumis and Olotillo have narrower niches than predicted by a linear regression with mean altitude, and that the opposite is true of Tuxpeño, Group 2.2, Group 3.2, and Group 3.3.

Degree of variability of environments at occurrence locations was quantified. A Euclidean distance dendrogram (R function hclust) of extracted environmental values was compared between Tuxpeño and Cónico, as they represent highland and lowland landrace groups with large sample sizes, as well as group of random individuals from all occurrence groups (Figure 4.18). We see that Tuxpeño has a lower number of branches at greater height and more at low heights, relative to Cónico or RANDOM background data. A consensus of 30 clustering indices (R function NbClust, package NbClust (Charrad et al., 2014)) finds \( k = 3 \) to be optimal for clustering environmental Euclidean distance for all three occurrence groups.
4.4.3 Environment Prediction

When genotype relatedness is estimated with categorical fastStructure group membership \((k = 9)\), genotype explains little of the environmental variance relative to landrace (paired t-test, \(p = 0.0008, df = 22\), Figure 4.19(a)), though there was no significant difference with the independent environmental data set (paired t-test, \(p - value = 0.42, df = 13\), Figure 4.19(b)). At \(k = 2\) and \(k = 3\), relationship matrices are sufficiently uninformative as to be useless for this model (data not shown). At \(k = 9\), the matrix has information enough to explain some variance, but still underperforms landrace name for both predictor environmental dataset and independent environmental dataset. Tables 4.2 and 4.3 provide variable component, Z-ratio, and model fitness metrics for these equations.

The second formulation of genotype relatedness is a transformation of ancestry coefficients with a Radial Basis function kernel. Using this estimation, genotype relatedness explains ENM predictor environmental variance on par with landrace name, but underperforms in the independent environmental data set (paired t-test, \(p = 0.056, df = 13\), Figure 4.19(d)). Tables 4.4 and 4.5 provide variance component, Z-ratio, and model fitness metrics for these equations.

The third formulation, the kinship matrix, captures significantly more ENM predictor environmental variance than does landrace name (paired t-test, \(p = 0.0005, df = 22\), Figure 4.19(e)), though there was no significant difference with the independent environmental data set (paired t-test, \(p = 0.30, df = 13\), Figure 4.19(f)). Genotype had a higher Z-ratio than landrace in 21 of 23 ENM predictor environmental variables (Table 4.6) and 8 out of 14 independent environmental variables, and two variables tied with values of zero (Table 4.7).

4.4.4 EAA

Gene names, associated environments, and gene functions are detailed in Table 4.8.

In all, 543 SNPs out of 19,887 were significant under the Bejamini-Hochberg procedure (Benjamini and Hochberg, 1995) at alpha = 0.20, and 64 of these fell within genic regions. At a lower false discovery rate of alpha = 0.05, 128 SNPs were identified, and 8 fell within genic regions. The
eight genes found at alpha = 0.05 were GRMZM2G165413, Zm00001d006570, GRMZM2G390221, Zm00001d043821, GRMZM2G310410, GRMZM2G353905, GRMZM2G026983, and GRMZM2G024738. Of these, six (GRMZM2G165413, Zm00001d006570, GRMZM2G390221, Zm00001d043821, GRMZM2G353905, GRMZM2G026983) have unknown functions, but GRMZM2G310410 (identified as significantly associated with solar radiation from July-September, soil bulk density, soil organic carbon content, and temperature seasonality) plays a role in protein serine/threonine kinase activity regulation and is involved in response to phosphate deficiency (Gowda, 2017; Gupta et al., 2017), and GRMZM2G024738 (associated with solar radiation in March and precipitation of the driest quarter) is expressed during photosynthesis and plays a role in fungal pathogen resistance (Donaldson, 2014; Kretschmer et al., 2017; Donaldson et al., 2013).

The environmental variable with the greatest number of significantly associated genes is percent soil organic carbon content (20/64), followed by mean temperature of the warmest quarter (11/64), altitude (10/64), percent soil bulk weight (9/64), and solar radiation from July-September (8/64). GRMZM2G390221 was identified as significantly associated with the greatest number of environments (Soil Bulk Density, Mean Temperature of Coldest Quarter, Mean Temperature of Warmest Quarter, Mean Precipitation of Driest Quarter, Precipitation Seasonality, Temperature Seasonality). Though the function of that gene is unknown, we know that it codes for a nucleic acid binding protein, that its expression varies with time exposure to light, and its co-expression network is involved in double-strand break repair and nucleus protein import (Proost and Mutwil, 2018). Several other genes were associated with 4 (GRMZM2G111917, GRMZM2G310410) or 3 (Zm00001d043821, GRMZM2G065461, GRMZM2G026983) environmental variables.

4.5 Discussion

4.5.1 Population Genetic Structure

Chalqueño, Cónico, and Cónico Norteño are consistently grouped within the umbrella group called Cónico (Sánchez and Goodman, 1992), characterized by pubescent plants of short- to medium-height, with high pollen shed and conical ears (Goodman and Bird, 1977). Olotillo, Tuxpeño,
Pepitilla, and Vandeño are consistently grouped with the group called Tropical Dents (Goodman and Bird, 1977; Sánchez G. et al., 2000; Arteaga et al., 2016). Tabloncillo is an eight-rowed group (Sánchez and Goodman, 1992) Caribbean Dent (Goodman and Bird, 1977). Bolita and Celaya belong to either the Caribbean Dents (Goodman and Bird, 1977; Arteaga et al., 2016) or the eight-rowed group (Sánchez and Goodman, 1992). Vandeño and Pepitilla are frequently harder to place. Pepitilla may be categorized either with the Cónicos (Goodman and Bird, 1977) or the Tropical Dents (Goodman and Bird, 1977; Sánchez G. et al., 2000). This is likely due to the hybrid origin of these landraces (Wellhausen et al., 1952) compounded with recent gene flow, which is facilitated by the intermediate elevations at which they are grown in the case of Pepitilla. Wellhausen considered the genealogy of both Olotillo and Pepitilla to be uncertain.

Published racial categorizations frequently disagree on certain landraces. Reif et al. (2006) tested the pedigrees of the modern incipient landraces published by Wellhausen et al. (1952), and found that the allele frequencies of the hybrid races were significantly correlated with expected frequencies. However, Mantel tests between distance matrices built from altitude, geographic location, and allele frequencies was also significant, indicating that these allele frequencies could also be explained by latitude, longitude, and elevation of origin. Reif et al. (2006) conducted STRUCTURE analyses on simple sequence repeat (SSR) markers for 24 landraces. At $k = 3$, the 10 landraces considered in this study were clustered into two groups: (1) Tabloncillo, Olotillo (Blanco), Tuxpeño, Celaya, Bolita, and Pepitilla, and (2) Cónico and Chalqueño. Cónico Norteño could not be clustered, and Vandeño was not included in their study. At $k = 4$, Tuxpeño and Pepitilla split from the rest of their group. We find similar ancestry coefficients, except that we cluster Cónico Norteño with the rest of the high-elevation Mexican conical landraces (despite its “complex pedigree” (Reif et al., 2006)), and we find that the Tabloncillo splits from the rest of the lowland cluster before Tuxpeño.

Many racial categorizations (notably those of Wellhausen et al. (1952)) rely strongly on comparisons of morphological differences. The validity of this approach is supported by $Q_{ST} - F_{ST}$ comparisons which find high morphological distinction between landraces which is maintained by
active selection (Pressoir and Berthaud, 2004). Reif et al. (2006) summarize the efforts of six studies that attempted to cluster like landraces based on morphology, isozyme data, cytological data, and chromosome knob data. Of the six, Pepitilla was clustered with the highland conicals half the time, and with the lowland landraces the other half (though their own analysis breaks the tie in favor of the lowland group). Bolita, Celaya, and Tabloncillo were also clustered with the lowland group, and Cónico, Cónico Norteño, and Chalqueño invariably clustered together in a highland group. (Reif et al., 2006) found Pepitilla and Tabloncillo to cluster with the other lowland landraces, in agreement with Wellhausen 1952, but contrary to Brown and Goodman (Goodman and Brown, 1988) who cluster Pepitilla with the highland conicals.

Our fastStructure results split these 10 landraces into a highland group, Group 2.2/Group 3.3 (Chalqueño, Cónico, Cónico Norteño), which is analogous to the umbrella Cónico group (Goodman and Bird, 1977), and a lowland group, Group 2.1, which breaks into a high-confidence lowland group, Group 3.2 (Olotillo, Tuxpeño, Vandeño), and a low-confidence lowland group, Group 3.1 (Pepitilla, Tabloncillo) at \( k = 3 \), the only fastStructure group that is not significantly structured by altitude. fastStructure finds Bolita and Celaya to be highly admixed at both \( k = 2 \) and \( k = 3 \), and are perhaps best characterized as mid-elevation landraces. Indeed, Bolita and Celaya are found at somewhat lower elevations than the rest of the highland group (Figure 4.12) and Celaya is reported to be highly productive at mid-elevation (Wellhausen et al., 1952).

Pairwise \( F_{ST} \) likewise clusters Cónico, Chalqueño, and Cónico Norteño, as in Group 2.2/Group 3.3, and Olotillo, Tuxpeño, and Vandeño, as in Group 3.2. On a sister branch from the landraces of Group 3.2, pairwise \( F_{ST} \) groups Bolita, Celaya, Pepitilla, and Tabloncillo, the landraces with high admixture or high attribution from Group 3.1. The low \( F_{ST} \) between Bolita and Celaya is explained by their common derivation from Tabloncillo, along with Zapalote Chico and Tuxpeño, respectively (Wellhausen et al., 1952).

The degree to which disparity between published phylogenies and those presented here can be explained by ongoing gene flow and evolution is unknown, but hybridization and criollization between improved and traditional maize varieties is common and intentional (Dyer and Taylor,
In one account of agricultural practices in Chiapas, discrimination of local maize landraces was “open and fairly pragmatically determined” (Dempsey, 1996), in that cornfields with plant traits that approximate the farmer’s idea of the landrace were incorporated into that landrace, with little regard for geographic origin. With this in mind, we would rightly expect phylogenetic trees to differ significantly depending on the location, the year, and even the field at which representative individuals are accessed for comparison, even given the same metrics of comparison, be they genomic, morphological, or cytological.

4.5.2 Ecological Niche Structure

While there are several valid definitions of ecological niche (Levin et al., 2012; McInerny and Etienne, 2012), we here define and consider niche in terms of abiotic environment tolerance, in keeping with Hutchinson (1957), while recognizing that abiotic variables are only a component of the suite of environmental conditions into which a niche must fit. This may be particularly true for cultivated and domesticated species, whose distribution and reproduction are strongly influenced by human industry as well as natural factors.

We are not the first to model the ecological niches of contrived genetic groups. Valdivia-Carrillo et al. (2017) also used Maxent to model the ecological niches, distributions, and niche overlap of $k$-groups produced by STRUCTURE. Several publications have set out to characterize the ecological niches of maize landraces. Ruiz Corral et al. (2008) split Mexican maize landraces into several niche categories, which frequently but imperfectly reflect phylogenetic categories. By their delineations, Bolita, Celaya, Chalqueño, Cónico, and Cónico Norteño are found in what they call Group 1, which has temperate to semi-hot temperatures and low precipitation. Pepitilla and Tabloncillo are placed in Group 2, and especially a subgroup of Group 2, which is semi-hot to hot and moderately wet. The authors also note that Bolita and Celaya trend towards Group 2 in PC space, due to their preference for higher precipitation. Thirdly, their Group 3 consists of Olotillo, Tuxpeño, and Vandeño, and is characterized by very hot and wet environments.
Niche similarity shows that Group 3.1 is most related to the lowland Group 2.1, which aligns with fastStructure results. Landrace niche overlap results between Environmental ENMs and PCA ENMs differ somewhat. In both, Cónico Norteño is highly similar to Celaya, Chalqueño is highly similar to Cónico, and Tabloncillo is highly similar to Tuxpeño. Pepitilla and Celaya cluster in the main branch with the rest of the highland group, though their closest relatives are variable. Olotillo clusters with Vandeño in one model and Pepitilla in the other, and Bolita’s place in the niche overlap dendrograms varies greatly between the two models. These results reinforce the previously defined highland and lowland groups, and lend support to placing Celaya and Pepitilla in the “highland” group, despite their high genetic admixture.

The major node in our composite niche overlap dendrogram separates Corral’s Group 1 landraces from Group 2 landraces (Figure 4.11). Broadly, we find that these ten landraces are split into two niche groups (Bolita/Celaya/Chalqueño/Cónico/Cónico Norteño/Pepitilla and Olotillo/Tabloncillo/Tuxpeño/Vandeño). This niche division places Vandeño within the group of Caribbean Dents and Pepitilla with the tropical group, though its high attribution of Group 3.1 indicates that it and Tabloncillo share some other sort of structure within the group. While we find Pepitilla to cluster with the other highland landraces, closest to Cónico and Chalqueño, Corral and colleagues place Pepitilla with the lowland landraces, closest to Tabloncillo. Pepitilla is characterized as a member of the Cónico group, but due to its history of being grown at more moderate altitudes, has experienced greater gene flow with other groups (Goodman and Bird, 1977), likely influencing its location in our admixture plots, genetic similarity dendrogram, and niche similarity heatmaps and dendrograms.

Like Corral, we find altitude to be the strongest environmental determinant of landrace niche determination. After altitude, annual precipitation, precipitation seasonality, mean temperature of the warmest and coldest quarters, temperature seasonality, and average wind speed were among the most important environmental variables.
4.5.3 Comparing Population Genetic and Niche Structure

We expect that degree of genetic relatedness is positively correlated with ecological similarity. This expectation is borne out in a survey of 32 plant species (Burns and Strauss, 2011).

We find that dendrograms built from pairwise $F_{ST}$ and from niche overlap are similar but not identical (Figure 4.11). While certain pairs or trios of landrace are consistently clustered, the positions of others are more sensitive to the similarity metric being used.

On the other hand, we find highly significant correlation between suitability distance and genetic distance (partial Mantel test, $r = 0.14$, $p = 0.001$). These results accounted for correlation due to geographic and altitudinal distance, and were similarly significant for both PCA ENMs and Environmental ENMs.

4.5.4 Ecological Niche Modeling Considerations and Caveats

As ecological niche modeling has grown in popularity, so too has the literature surrounding the proper implementation of this approach. One of the topics receiving the most attention is the proper interpretation of AUC. Many have shown AUC to be highly sensitive to the parameters of the model, including sample size (Lobo et al., 2008; Costa et al., 2010), number of absence or pseudoabsence or background points (Soberon and Peterson, 2005; Wisz et al., 2008; Golicher et al., 2012), the spatial extent of either (Soberon and Peterson, 2005; Golicher et al., 2012; Boria et al., 2014), the number of environmental variables included as predictors (Warren et al., 2014), the degree of correlation between predictors (Warren et al., 2014), the degree of autocorrelation within predictor(s) (Lennon, 2000; Golicher et al., 2012), and incomplete spatial sampling (also known as the Wallacean Shortfall (Bini et al., 2006)). Because AUC is also influenced by spatial aggregation, this sampling was conducted with reference to a grid, in which the sample’s distribution is as representative as possible of the full group’s distribution. Maxent is less sensitive to small sample size than other predictive algorithms (Baldwin, 2009; Costa et al., 2010), sometimes using only five occurrence points (Hernandez et al., 2006; Pearson et al., 2007; Costa et al., 2010), but others report that predictive models generally perform poorly under $n = 30$ samples (Wisz et al., 2008), and AUC
scores that have 95% confidence intervals that do not include a slope of 1 are good indicators that sample size is sufficient (Hanczar et al., 2010). Also, Maxent avoids complications associated with absence and pseudo-absence by using only absence data, and in doing so, has higher performance and dependability (Baldwin, 2009). Instead, Maxent uses background data, which does not make assumptions regarding the incapacity of a group to survive at locations where the group is not reported to be observed.

A wealth of literature has built up around the issue of quantifying niche breadth. One important caveat to estimation of niche breadth is the positive relationship between niche breadth and geographical range (Slatyer et al., 2013). Geography also ties in to niche overlap estimation. Sobel et al. (2010) elucidates the theoretical framework which distinguishes effective geographic isolation and ecogeographic isolation. Though the geographic isolation of two populations may be complete, their ecogeographic isolation (estimated either by suitability projections via ENMs or tested empirically with reciprocal transplantations) may be complete, null, or anywhere in between. The number and types of resources and/or environmental pressures considered greatly influence niche breadth estimations. A population may be a generalist with regards to one environmental cline, but a specialist at surviving in another.

Furthermore, crop systems present additional logistic challenges for ENM. Few studies have attempted to commit ENM methodology to cultivated crops. Ureta et al. (2012) in 2012 suggested that they might have been the first (though it would seem that Ruiz Corral et al. (2008) may be more deserving of such a title, as they applied GLMs to find correlations between occurrence data and environmental layers). The role of climate in a crop group’s geographical distribution can be difficult to disentangle from the critical role of human intervention (Ureta et al., 2012; Dyer and López-Feldman, 2013; Orozco-Ramírez et al., 2016), though new tools and datasets are emerging which allow for the modeling of anthropogenic selection forces across evolutionary time scales (e.g., Used Planet: A Global History (Ellis et al., 2013)). Farmers consciously select primarily for ear characteristics (kernel filling, large ears, varietal consistency (Louette and Smale, 2000)) and plant characteristics, in response to market forces (seed price, culinary preferences, etc. (McLean-
Rodríguez et al., 2019)) and other cultural considerations. These traits may have little or nothing to do with plant survival or net reproduction, and in fact may actually decrease plant fitness in the field. Further complicating niche modeling in crops, farmers understand the environmental conditions in their fields and how they affect their crops (Hernández Xolocotzi, 1985), and modify environment with inputs to permit crop cultivation in environments where they otherwise would be replaced by other crops or land uses (Pearsall, 1978). Unless land use data can capture the ways in which farmer inputs modify environment, these dynamics will lead to error in our niche models. Farmers and their crops may fruitfully be conceived of as highly complex communities, wherein landraces compete for farmer’s choice in either positive or negative frequency-dependent selection, where both landraces and farmers engage in niche construction, and non-random mating is directed by farmers (even in open-pollinated fields, intentional seed transfer is used to introduce desired traits into populations). While this may be true of any number of species which regularly interact with humanity, this interaction is almost certainly stronger in domesticated species, whose survival is almost entirely dependent on farmer intervention. A full accounting of a landrace’s niche is impossible without recognizing the integral role of humanity. At the same time, the abiotic environment still selects for plant survival and reproduction (Cleveland and Soleri, 2007). Together, the net aggregate of these at times conflicting selective factors comprise the agroecosystem to which landraces evolve (Villa et al., 2005; Caldu-Primo et al., 2017), and reducing a crop’s niche to its association with abiotic environment is therefore at best a useful and informative approximation of elements important to its ecological niche. We stress the importance of the role of farmer decision in landrace distribution, while yet maintaining that abiotic environment plays an influential role as well, both directly on the survival of the plant and indirectly via influencing farmer decision.

We lack a robust analytic framework to model all interactions between crop system, agriculturalist, and environment in ENM, but we include highland and lowland wild relatives in these analyses to enable the comparison of ENMs of cultivated and non-cultivated related species.

fastStructure found $k = 2$ and $k = 3$ to optimally explain the population structure within these 10 landraces. If we had chosen a $k$ value equal to the number of landrace groups, each
genetic group would have lower spatial disaggregation, lower variance in altitude, higher AUC and lower niche breadth (though groups with lower mean elevation would have broader niches than groups with higher mean elevation). Indeed, this trend would likely continue for \( k \)-values beyond the number of included landraces. Groups would consist of smaller and smaller sample sizes of like individuals, achieving higher and higher AUC and lower and lower niche breadth, until ENM integrity is compromised by small sample size. It seems unlikely that there exists a threshold \( k \)-value beyond which ENM algorithms would be unable to produce niche models with “acceptable” AUC scores to which biological interpretations could not be levied. For this reason, we return to and expand our previous notes of caution regarding the relativity of these descriptive niche metrics. Niche breadth and AUC may not be useful in selecting the optimal degree of population sub-structure, but they can be used describe the niche breadth of a genetic group relative to other groups at a given level of population sub-structure.

4.5.5 Local Adaptation in Landraces and Genetic Groups

The strongest determinant of niche breadth is mean elevation. On average, landraces from higher elevations have narrower niches. Of secondary importance is the degree of geospatial aggregation of the group.

Perhaps confusingly, Ruiz Corral et al. (2008) refer to the capacity of a landrace group to persist in a wide variety of habitat types as “adaptability.” Terms like “tolerance” or “plasticity” may be more appropriate for the kind of habitat flexibility here described. We would prefer to save the word “adaptability” to refer to the capacity for a population to evolve adaptations to its environment. A locally-adapted population that specializes in one or a few habitat types may be “adapted” but have low “adaptability,” as it has undergone natural selection and has purged genetic diversity that would allow local adaption to other habitats. Reciprocally, a generalist population with large effective population size and low genetic load might be considered to have high “adaptability” precisely because it has not undergone local adaptation. By their metric, Tuxpeño is the most adaptable, followed by Olotillo, Pepitilla, and Celaya (19, 13, 12, and 11 climate types, respectively).
They note Tuxpeño’s high adaptability as the reason for its successful incorporation into modern agronomic lines. They also note that Tabloncillo, Tuxpeño, and Olotillo have the highest variability in altitudes at which they are cultivated. We likewise find that Tuxpeño has a high niche breadth, and that Tabloncillo, Tuxpeño, and Olotillo have the lowest altitudes of the landraces included here.

Ureta et al. (2012) provide AUC scores for ENMs for Mexican maize landraces. They find that the mean maize AUC is 0.924. Relative to this mean value, Cónico, Chalqueño, Bolita, and Cónico Norteño to have narrow niches, Olotillo and Tabloncillo have average niche breadth, and Celaya, Tuxpeño, and Venden have low niche breadth. Again, these results track tightly with niche breadth patterns calculated here.

Tuxpeño carries all the hallmarks of a generalist landrace. It has the lowest mean AUC, the widest geographic distribution, the lowest SSI, and moderately high B2, and as expected of generalist landraces, Tuxpeño is found at low elevations (Table 4.1). In general, specialists are adapted to a single stable environment, while generalists tolerate a variety of habitats and have an advantage in marginal, temporally variable, and unpredictable environments (Smith and McKelvey, 1986). Devictor et al. (2008) describe generalists as inhabiting disturbed and dispersed habitats. Both of these traits might be used to describe the fields across lowland central Mexico in which modern agriculture is practiced, and in which Tuxpeño has been observed to be better suited than other landraces (Ruiz Corral et al., 2008).

An alternative hypothesis regarding high niche breadth landraces is that within-landrace structure clusters may be uniquely adapting to different local environments (Bolnick et al., 2002; Arteaga et al., 2016). In this case, landraces like Tuxpeño would be umbrella groups that are made of smaller genetic groups with narrow but distinct niches. Dendrograms clustering environments experienced by individuals of Cónico, Tuxpeño, and random landrace individuals from any landrace (all subsampled to equal sample size) reveals branching at low and high distances, signifying deep structure and shallow structure in environmental differentiation respectively. Tuxpeño, Cónico, and random individuals have comparable intra-race genetic structure (fastStructure \( k = 3 \)) and environmental
clustering \((k = 3)\), but a weak signal emerges of deeper branching in Cónico and random group than in Tuxpeño, suggesting that environments experienced by Tuxpeño are more similar than those experienced by Cónico, or than expected at random (Figure 4.18), contrary to expectations. This level of intra-race structure is greater than that of the other landraces \((k = 1)\), but is due only to larger sample size; When all landraces were subsampled down to 30 individuals, all landraces returned genetic structure of \(k = 1\) (data not shown). We therefore fail to find support for adaptive sub-groups or elevated between-individual niche differentiation within Tuxpeño or Cónico, but we encourage further investigation into these possibilities.

Conversely, the landraces with the narrowest niches are Chalqueño, Cónico, and Olotillo, when considering multiple breadth metrics. The niche breadth of Pepitilla is harder to gauge, as Environmental AUC and PCA AUC diverge strongly from B1 and B2 for this landrace. These metrics diverge strongly in Vandeño as well, but all still trend lower than in Pepitilla (except for the weakest indicator, SSI, which is higher in Vandeño than in Pepitilla). Arteaga et al. (2016) found that though landraces do not cluster succinctly in PCA space, of the seven genetic clusters in their panel, Cónico was the least dispersed. This finding is in accordance with our results regarding the included members of the Cónico group (Chalqueño, Cónico, and Cónico Norteño), which have relatively narrow niches by most metrics, and have fastStructure ancestry coefficient values that are relatively stable between tested \(k\) values. Marginal high-elevation conditions exert more severe abiotic selective pressure, and because of the lower genetic diversity of the highland material due to elevational selection and range expansion bottlenecks, the members of the highland Cónico group have a narrower niche, a more similar genetic base, and are more phenotypically homogenous.

### 4.5.6 Targets of Selection

Neither genetic nor morphological clustering methods concisely cluster landraces into discrete groups (Van Heerwaarden et al., 2011; Arteaga et al., 2016; Caldu-Primo et al., 2017). Research generally finds low genetic divergence between landraces, and what structure is found generally correlates with latitudinal and elevational clines (Doebley et al., 1985; Van Heerwaarden et al.,
2011; Arteaga et al., 2016) or social structure (Orozco-Ramírez et al., 2016). One reason for low genetic divergence is high gene flow between populations, either due to pollen drift or intentional seed trade between farmers (Pressoir and Berthaud, 2004; Dyer and Taylor, 2008; Orozco-Ramírez et al., 2016; Arteaga et al., 2016). Farmers select for only a few traits (at least consciously (Louette and Smale, 2000; Bellon et al., 2003)), and these selective regimes can vary from region to region (Pressoir and Berthaud, 2004). Because of rapid linkage disequilibrium decay in maize (Remington et al., 2001; Takuno et al., 2015), selection can be efficient. Therefore, while most of the genome can be affected by broad-scale selection to bioclimatic variables, altitude, and latitude, as well as gene flow with neighboring landrace populations, the few key genes of large and/or pleiotropic effect that are responsible for the defining traits of landraces (ear morphology, etc.) can be retained within populations (Pressoir and Berthaud, 2004; Arteaga et al., 2016; Caldu-Primo et al., 2017).

GRMZM2G310410 is one of the eight genes which was identified at the more stringent FDR alpha = 0.05, identified as correlated with Soil Bulk Density, Soil Organic Carbon Content, Solar Radiation (July-September), and Temperature Seasonality. While the precise function of GRMZM2G310410 is unknown, it is known to be a protein serine/threonine kinase, which is inhibited by the microRNA zma-miR395c/l/m, possibly in relation to low phosphate stress (Gupta et al., 2017). Protein kinase superfamily proteins engage in phosphorylation, often in association with post-translational modification, which is important in regulating response to a variety of stressors, including high temperature (Zhang et al., 2019a) and nitrate stress (Trevisan et al., 2012), making GRMZM2G310410 a prime candidate for further investigation.

Though four genes were associated with wind speed, three are uncharacterized and the other is a heat shock protein (GRMZM2G002131, HSF transcription factor 1, resistance to heat stress (Renaud, 2015)) also identified as associated with Soil Bulk Density.

Soil type and quality variables were associated with several genes. Soil Cation Exchange Capacity was associated with GRMZM2G383122, which functions in leaf blight resistance (Li et al., 2018), and GRMZM2G037630, which is associated with signalling in response to waterlogging stress (Zhang et al., 2019b; Rajhi et al., 2011; Liu et al., 2014). Soil Organic Carbon Content was associated with
GRMZM2G054050 (low phosphate stress response (Svistoonoff et al., 2007; Calderón-Vázquez et al., 2011; Zhang et al., 2014; Schlüter et al., 2012)), GRMZM2G100403 (heat stress (Frey et al., 2015)), GRMZM2G167865 (drought stress response (Goyal et al., 2005)), and GRMZM2G310410 (discussed earlier). Soil Bulk Density was associated with GRMZM2G464510 (Emp602, critical gene in kernel development (Ren et al., 2019)), GRMZM2G002131 (HSF transcription factor 1, resistance to heat stress (Renaud, 2015)), and GRMZM2G310410 (discussed earlier). Soil Sand Percentage was associated with two different genes that are connected with drought resistance (GRMZM2G405064, ZmPYL12 (Peng et al., 2017); GRMZM2G015132, dihydrolipoamide S-acetyltransferase1 (Xin et al., 2018)). Soil Silt Percentage was associated with GRMZM2G556131 (DHDPSI, lysine biosynthesis and stress response (Liu et al., 2016)) and GRMZM2G064725 (SNF1-related protein kinase regulatory subunit beta-1, regulates signal transduction and metabolism (Li et al., 2009)).

This analysis does not detect structural variants, copy number variants, or dominance or epistatic effects, which are likely also important to crop adaptation (Gaut et al., 2018). Therefore, any adaptation we detect by looking at SNPs may be just the tip of the iceberg.

4.5.7 Landrace Name vs. Genotype

Landrace names are an attempt to place hard lines around and between dynamic systems, and while these names reflect biological realities to varying degrees, they are by their nature generalizations. As adaptations to environment have genetic bases, we expect genetic similarity to better reflect the environments to which they are adapted. When genetic similarity is reduced to categorical genetic group clusters via fastStructure or PCA (Arteaga et al., 2016), landraces do not cluster succinctly, and clusters comprise individuals from multiple landraces. Categorical clusters of this nature predict environmental variance more poorly than categorical landrace groups. Even continuous genetic similarity calculated from the Radial Basis function kernel of ancestry coefficients underperforms relative to categorical landrace name. Only when genetic similarity is estimated with a kinship matrix does genetic similarity outperform landrace name. The equation used to construct the Centered IBS kinship matrix is more effective at capturing additive genetic variance.
(Endelman and Jannink, 2012) than clustering algorithms and likely gives a better estimate of the truer estimate of environmental variance explained.

Of the Environmental ENM predictor environmental variables, Precipitation Seasonality is the only in which the landrace Z-ratio was both greater than the genotype Z-ratio and greater than 1. For the independent environmental dataset, the above criteria are met by Monthly Average Daily Mean Temperature, Potential Evapotranspiration, envPC2 (which mostly captures variation in Average Daily Mean Temperature and Altitude), and envPC4 (which mostly captures variation in Frost Day Frequency, in which landrace Z-ratio is nearly 1 and genotype Z-ratio is 0). It is not entirely clear as to why landrace would be a better predictor of environment than the kinship matrix, except that additive genetic variance is not the only kind of genetic variance that contributes to adaptation.

Paired t-tests found that genotype Z-ratios were significantly higher than landrace Z-ratios in the predictor environmental dataset, but not the independent set. This may be because the first environment dataset is a better representation of the environmental factors to which these maize landraces have adapted. The predictor environmental dataset takes average values from over many years, and the independent dataset records environmental data from a single growing season, which may be skewed by particular weather events not reflective of longstanding patterns of environmental variance. This result emphasizes the value of studies that span multiple growing seasons.

4.5.8 Landrace Conservation

Landraces are vulnerable to replacement by modern agronomic hybrids, and some landraces are more at risk than others. Though socioeconomic factors contribute significantly to landrace vulnerability (Steinberg, 1999; Dyer and Taylor, 2008; de la Barrera and Martínez, 2018; Bellon et al., 2018; McLean-Rodríguez et al., 2019), the two factors that explain the greatest variance in vulnerability are potential geographic distribution and the projected response of potential distribution to climate change (de la Barrera and Martínez, 2018). The 10 landraces included in our study, in order of decreasing vulnerability, are as follows: Pepitilla, Tabloncillo, Celaya, ConNor, Olotillo,
Chalqueño, Cónico, Vandeño, Bolita, and least vulnerable, Tuxpeño. This ranking does not closely mirror our rankings of either niche breadth, altitude, or SDM range (number of cells with suitability scores greater than 0.5, data not shown), highlighting the influence of socioeconomic factors on niche vulnerability.

We do not find evidence that fastStructure-informed sampling regimens would fruitfully assist in guiding seed germplasm conservation efforts. ENMs constructed from either grouping method (landrace name or fastStructure group) have comparable niche breadths, relative to covariates. The greatest number of different landraces can be found at temperate subtropical climates in Mexico (Ruiz Corral et al., 2008), but landrace diversity per se is not the goal of conservation. As the top priority of conservation should be sub-populations with unique, unaccessed adaptations to challenging environments, and highly vulnerable landraces in particular, efforts should be focused on landraces maintained by secluded indigenous populations that grow maize in marginal conditions. Not only are such landraces more likely to harbor unique adaptations, but such communities are more likely to switch to modern varieties, either due to economic development and/or climate change. Whether landraces designations within other crop systems are as reflective of meaningful adaptations and genetic population structure is unknown. Crops with poorly characterized or fluid landrace populations may benefit more from inclusion of genotyping in prioritization of conservation focus.

Traditional knowledge of landraces and farming techniques is also at risk of loss. Farmers are generally knowledgable about their seeds and the conditions in their field and surrounding environment (Ruiz Corral et al., 2008), though the degree of knowledge about field conditions and adaptations appears to be highly variable from area to area (Steinberg, 1999). Though some fear the loss of traditional knowledge of these landrace resources and the farming practices to utilize them, others counter that such knowledge and practices evolve in situ along with the changing socio-economic landscape, and with the landraces themselves (Bellon, 1991; McLean-Rodríguez et al., 2019).
The biological reality and stability of maize landrace designations, despite high gene flow (Arteaga et al., 2016) and seed abandonment (Dyer and Taylor, 2008; McLean-Rodríguez et al., 2019), is remarkable, yet perhaps explained in part by the deep cultural significance of maize varieties to the communities that maintain them. No other crop is as culturally important to the indigenous people of Mexico as is maize (Steinberg, 1999). While other crops like beans and rice are grown as cash crops, maize is embedded into the culture, involved in medicine, mythology, religious ceremony, and community social status. However, social change loosens these ties, and leads to replacement by modern varieties (Steinberg, 1999). When locally-grown maize is sold rather than used by the farmer or the community for subsistence and cultural applications, market forces select for the more easily sold modern varieties rather than landrace varieties, which have more traditional or ceremonial uses and values. Landraces that are grown for their association with indigenous religious symbolism and ceremony are no longer grown when those religious beliefs are replaced. The influx of Protestant evangelical Christianity into Mayan Mopan communities lead to the abandonment of local varieties, though some Maya beliefs were admixed with the introduced Catholic faith, preserving those traditions and preventing further abandonment of modern varieties (Steinberg, 1999). Though farmers also function in seed improvement and diffusion, a farmer’s goal is to gain value through their crop, be it sustenance, money, or products of cultural/ritual significance. Seed improvement and diffusion are emergent functions of the broader system (Dyer and Taylor, 2008). Although alleviation of economic marginalization and isolation leads to greater infrastructure and elevated standards of living, it also increases the risk of landrace abandonment (Byerlee, 1996; Arteaga et al., 2016). Participatory breeding efforts can serve the best interests of both farming communities and landrace conservation (see Dempsey (1996) and Bellon et al. (2003) for examples). In order to preserve maize landrace diversity in the fields of Mexico, the farmers responsible for managing that diversity must be empowered to maintain the traditional agricultural methods that gave rise to that diversity (Arteaga et al., 2016; Caldu-Primo et al., 2017).
4.6 Conclusions

Maize landraces are an agronomic resource of high but as of yet unquantified agronomic value. When considered at broad geographic scales, we find evidence of local adaptation across ten landraces, with highland landraces aligning with specialist strategies and lowland landraces aligning with generalist strategies. Moreover, landrace name is a fairly good predictor of both genetic diversity and ecological adaptation, in many respects on par with genetic clusters attainable via genetic clustering algorithms (though additive genetic variance is still a better predictor of environment than is landrace name). Finally, we identify SNPs correlated with environmental variables that strongly constrain landrace geographic distribution, signalling critical roles in abiotic adaptation.

4.7 References


Figure 4.1 Loadings for the first five environmental PCA components.
Figure 4.2 Pearson Correlation between environmental predictors. Correlations between all predictors (a) resulted in sixteen clusters (less than 30% correlated). We then combined monthly values into seasonal values for solar radiation and water vapor, and combined all wind values into an annual average. In the second correlation (b), eighteen unique environmental layer clusters were found (less than 15% correlated), and within clusters, a single layer was selected.
Figure 4.3 Loadings for the first five Principal Components from the independent environmental dataset.
Figure 4.4  Population genetic structure.  (a), (b) Each vertical line represents a genotyped individual. Individuals are grouped by landrace identity, and are colored by membership into genetic groups. Genetic structure complexity was best explained at either $k = 2$ or $k = 3$.  (c), (d) Ancestry component of each landrace.  (e) Hierarchical clustering dendrogram constructed from mean pairwise $F_{ST}$ values.
Figure 4.5  Maxent model suitability score predictions based on environmental layers.
Figure 4.6  Maxent model suitability score predictions based on PCA-transformed environmental layers.
Figure 4.7 Percent contribution of each predictor to each occurrence group.
Figure 4.8 Percent contribution of each predictor to each occurrence group.
Figure 4.9  Heatmaps and dendrograms of niche overlap (Schoener’s D) between landraces and genetic groups.
Figure 4.10  Heatmaps and dendrograms of niche overlap (Schoener’s D) between landraces and genetic groups, averaged between landrace and genetic group models, and with applied threshold $D = 0.6$. 
Figure 4.11 Tangle plot of alignment of dendrograms demonstrating genetic similarity (mean pairwise $F_{ST}$) and niche similarity (Schroener’s D, entanglement coefficient = 0.09).
Figure 4.12 Elevation of sampled landrace accessions and wild relatives.
Figure 4.13  Distribution of AUC values for each occurrence group. Red lines indicate the value of the combined mean of the AUC for all ENM replicates for each category of occurrence groups (landrace, genetic group, wild relative, and random). (a) AUC from models built from environmental data. (b) AUC from models built from PCA-transformed environmental data.
Figure 4.14 Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.
Figure 4.15  Relative metrics of niche breadth and correlated environmental variables. High values indicate high niche specificity (narrow niche breadth). The pink line is the average of Normalized AUC, Normalized AUC (PCA), 1 - Normalized B1, 1 - Normalized B2, and Normalized SSI. All variables are normalized for comparison.

Table 4.1  Niche breadth and correlated values

<table>
<thead>
<tr>
<th>Variable</th>
<th>Spatial Disaggregation (km)</th>
<th>Altitude (m)</th>
<th>Mean AUC</th>
<th>Mean AUC (PCA)</th>
<th>B1</th>
<th>B2</th>
<th>SSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOLITA</td>
<td>561.81</td>
<td>1735</td>
<td>0.97</td>
<td>0.89</td>
<td>0.98</td>
<td>0.67</td>
<td>1.91</td>
</tr>
<tr>
<td>CELAYA</td>
<td>285.32</td>
<td>1729</td>
<td>0.91</td>
<td>0.86</td>
<td>0.97</td>
<td>0.55</td>
<td>1.24</td>
</tr>
<tr>
<td>CHALQO</td>
<td>335.85</td>
<td>2184</td>
<td>0.97</td>
<td>0.96</td>
<td>0.91</td>
<td>0.18</td>
<td>1.34</td>
</tr>
<tr>
<td>CONICO</td>
<td>212.62</td>
<td>2157</td>
<td>0.98</td>
<td>0.98</td>
<td>0.94</td>
<td>0.31</td>
<td>1.45</td>
</tr>
<tr>
<td>CONNOR</td>
<td>331.27</td>
<td>1912</td>
<td>0.89</td>
<td>0.90</td>
<td>0.97</td>
<td>0.50</td>
<td>1.45</td>
</tr>
<tr>
<td>OLOTITL</td>
<td>501.68</td>
<td>580</td>
<td>0.94</td>
<td>0.92</td>
<td>0.95</td>
<td>0.36</td>
<td>1.07</td>
</tr>
<tr>
<td>PEPITI</td>
<td>157.41</td>
<td>1466</td>
<td>0.97</td>
<td>0.93</td>
<td>0.98</td>
<td>0.69</td>
<td>1.15</td>
</tr>
<tr>
<td>TABLON</td>
<td>493.07</td>
<td>893</td>
<td>0.93</td>
<td>0.92</td>
<td>0.96</td>
<td>0.42</td>
<td>1.58</td>
</tr>
<tr>
<td>TUXPEN</td>
<td>782.84</td>
<td>610</td>
<td>0.79</td>
<td>0.81</td>
<td>0.98</td>
<td>0.71</td>
<td>0.69</td>
</tr>
<tr>
<td>VANDEN</td>
<td>458.48</td>
<td>740</td>
<td>0.94</td>
<td>0.91</td>
<td>0.99</td>
<td>0.85</td>
<td>1.58</td>
</tr>
<tr>
<td>group2_1</td>
<td>723.55</td>
<td>619</td>
<td>0.74</td>
<td>0.77</td>
<td>0.99</td>
<td>0.74</td>
<td>-</td>
</tr>
<tr>
<td>group2_2</td>
<td>402.72</td>
<td>2199</td>
<td>0.93</td>
<td>0.93</td>
<td>0.97</td>
<td>0.51</td>
<td>-</td>
</tr>
<tr>
<td>group3_1</td>
<td>600.22</td>
<td>843</td>
<td>0.90</td>
<td>0.87</td>
<td>0.98</td>
<td>0.69</td>
<td>-</td>
</tr>
<tr>
<td>group3_2</td>
<td>727.00</td>
<td>504</td>
<td>0.82</td>
<td>0.79</td>
<td>0.99</td>
<td>0.74</td>
<td>-</td>
</tr>
<tr>
<td>group3_3</td>
<td>255.19</td>
<td>2268</td>
<td>0.92</td>
<td>0.87</td>
<td>0.96</td>
<td>0.40</td>
<td>-</td>
</tr>
<tr>
<td>MEXICANA</td>
<td>247.32</td>
<td>2156</td>
<td>0.98</td>
<td>0.98</td>
<td>0.89</td>
<td>0.14</td>
<td>-</td>
</tr>
<tr>
<td>PARVIGLUMIS</td>
<td>236.64</td>
<td>1122</td>
<td>0.97</td>
<td>0.97</td>
<td>0.88</td>
<td>0.12</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 4.16 Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.
Figure 4.17 Difference in correlation between AUC values from environmental variables and PCA-transformed environmental variables. (a) Difference in linear regression between real data and null model. (b) Linear trend in residual from null model. (c) Influence plot identifying occurrence groups that most heavily influence the linear regression.
Figure 4.18  Environmental heterogeneity and structure within landrace groups. Dendrograms of individual similarity based on Euclidean distance of 18 scaled environmental variables at location. TUXPEN branches slightly less at higher height and more at lower height, relative to CONICO and RANDOM.
Figure 4.19 Z-ratios of genotype and landrace name designation at predicting two different environmental datasets. Genotype similarity matrices estimated by categorical fastStructure group membership \((k = 9, \text{(a), (b)})\), Radial Basis function kernel of fastStructure ancestry coefficients \((k = 3, \text{(c), (d)})\), and kinship matrix \((\text{(e), (f)})\).
Table 4.2  Predictor environmental dataset prediction with categorical fastStructure group membership ($k = 9$, Figure 4.19(a))

<table>
<thead>
<tr>
<th>Var</th>
<th>genotype_VarComp</th>
<th>genotype_Zratio</th>
<th>landrace_VarComp</th>
<th>landrace_Zratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>0.000</td>
<td>0.569</td>
<td>0.000</td>
<td>0.536</td>
<td>1067.21</td>
<td>-2132.42</td>
</tr>
<tr>
<td>Annual.Mean.Diurnal.Range</td>
<td>0.000</td>
<td>0.857</td>
<td>0.000</td>
<td>0.991</td>
<td>834.47</td>
<td>-1666.94</td>
</tr>
<tr>
<td>Temperature.Seasonality</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.629</td>
<td>1859.06</td>
<td>-3716.12</td>
</tr>
<tr>
<td>Mean.Temperature.of.Warmst Quart</td>
<td>0.000</td>
<td>0.289</td>
<td>0.000</td>
<td>0.965</td>
<td>897.94</td>
<td>-1793.88</td>
</tr>
<tr>
<td>Mean.Temperature.of.Coldest Quart</td>
<td>0.000</td>
<td>0.349</td>
<td>0.000</td>
<td>0.969</td>
<td>980.90</td>
<td>-1959.81</td>
</tr>
<tr>
<td>Annual.Precipitation</td>
<td>0.000</td>
<td>0.347</td>
<td>0.000</td>
<td>0.294</td>
<td>1147.05</td>
<td>-2292.11</td>
</tr>
<tr>
<td>Bulk.Density..mean.</td>
<td>0.002</td>
<td>1.161</td>
<td>0.004</td>
<td>1.626</td>
<td>318.94</td>
<td>-635.88</td>
</tr>
<tr>
<td>Cation.Exchange.Capacity.of.Soil..mean.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.795</td>
<td>39.99</td>
<td>-77.99</td>
</tr>
<tr>
<td>Weight.Percentage.of.Clay.Particles..mean.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.789</td>
<td>-17.20</td>
<td>36.39</td>
</tr>
<tr>
<td>Soil.Organic.Carbon.Content..mean.</td>
<td>0.000</td>
<td>0.153</td>
<td>0.002</td>
<td>0.995</td>
<td>170.93</td>
<td>-339.87</td>
</tr>
<tr>
<td>pH.Index.Measured.in.Water.Solution..mean.</td>
<td>0.002</td>
<td>0.839</td>
<td>0.006</td>
<td>1.579</td>
<td>229.61</td>
<td>-457.23</td>
</tr>
<tr>
<td>Weight.Percentage.of.the.Silt.Particles..mean.</td>
<td>0.002</td>
<td>0.786</td>
<td>0.002</td>
<td>0.816</td>
<td>-50.49</td>
<td>102.98</td>
</tr>
<tr>
<td>Weight.Percentage.of.the.Sand.Particles..mean.</td>
<td>0.001</td>
<td>0.345</td>
<td>0.002</td>
<td>0.815</td>
<td>-109.49</td>
<td>220.98</td>
</tr>
<tr>
<td>Precipitation.Seasonality</td>
<td>0.000</td>
<td>0.339</td>
<td>0.000</td>
<td>1.431</td>
<td>1427.04</td>
<td>-2852.07</td>
</tr>
<tr>
<td>Precipitation.of.Driest.Quarter</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1270.42</td>
<td>-2538.84</td>
</tr>
<tr>
<td>Solar.Radiation..kJ.m.2.day.1...March.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.196</td>
<td>1507.94</td>
<td>-3013.89</td>
</tr>
<tr>
<td>Solar.Radiation..kJ.m.2.day.1...July.September.</td>
<td>0.000</td>
<td>1.148</td>
<td>0.000</td>
<td>0.863</td>
<td>1116.11</td>
<td>-2230.23</td>
</tr>
<tr>
<td>Wind.Speed..m.s.1...Annual.Mean.</td>
<td>0.000</td>
<td>0.833</td>
<td>0.000</td>
<td>0.782</td>
<td>1008.95</td>
<td>-2015.90</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.000</td>
<td>0.581</td>
<td>0.000</td>
<td>0.000</td>
<td>1409.33</td>
<td>-2816.67</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1440.52</td>
<td>-2879.04</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.004</td>
<td>0.933</td>
<td>0.004</td>
<td>1.174</td>
<td>704.16</td>
<td>-1406.32</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.474</td>
<td>1166.39</td>
<td>-2330.78</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.604</td>
<td>823.70</td>
<td>-1645.40</td>
</tr>
</tbody>
</table>
Table 4.3 Independent environmental dataset prediction with categorical fastStructure
group membership ($k = 9$, Figure 4.19(b))

<table>
<thead>
<tr>
<th></th>
<th>genotype_VarComp</th>
<th>genotype_Zratio</th>
<th>landrace_VarComp</th>
<th>landrace_Zratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Cover</td>
<td>0.080</td>
<td>0.249</td>
<td>0.000</td>
<td>0.000</td>
<td>1490.51</td>
<td>-2979.02</td>
</tr>
<tr>
<td>Diurnal Temperature Range</td>
<td>0.938</td>
<td>1.408</td>
<td>0.033</td>
<td>0.368</td>
<td>805.64</td>
<td>-1609.28</td>
</tr>
<tr>
<td>Frost Day Frequency</td>
<td>84.930</td>
<td>0.942</td>
<td>41.061</td>
<td>0.807</td>
<td>477.11</td>
<td>-952.21</td>
</tr>
<tr>
<td>Monthly Average Daily Mean Temperature</td>
<td>0.234</td>
<td>0.258</td>
<td>3.835</td>
<td>1.397</td>
<td>697.25</td>
<td>-1392.50</td>
</tr>
<tr>
<td>Wet Day Frequency</td>
<td>84.323</td>
<td>0.952</td>
<td>0.000</td>
<td>0.000</td>
<td>1110.68</td>
<td>-2219.36</td>
</tr>
<tr>
<td>Potential Evapotranspiration</td>
<td>0.005</td>
<td>0.381</td>
<td>0.022</td>
<td>1.101</td>
<td>1114.95</td>
<td>-2227.90</td>
</tr>
<tr>
<td>Precipitation</td>
<td>40.745</td>
<td>0.202</td>
<td>0.000</td>
<td>0.000</td>
<td>646.43</td>
<td>-1290.87</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>0.000</td>
<td>0.358</td>
<td>0.000</td>
<td>0.069</td>
<td>569.58</td>
<td>-1137.16</td>
</tr>
<tr>
<td>Soil pH</td>
<td>0.000</td>
<td>0.000</td>
<td>0.024</td>
<td>1.601</td>
<td>-208.14</td>
<td>418.27</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.002</td>
<td>0.826</td>
<td>0.000</td>
<td>0.000</td>
<td>985.89</td>
<td>-1969.78</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.005</td>
<td>1.128</td>
<td>0.005</td>
<td>1.418</td>
<td>844.12</td>
<td>-1686.24</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.027</td>
<td>1.590</td>
<td>-115.34</td>
<td>232.67</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.001</td>
<td>0.761</td>
<td>0.001</td>
<td>1.059</td>
<td>481.62</td>
<td>-961.24</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.000</td>
<td>0.344</td>
<td>0.003</td>
<td>1.142</td>
<td>90.73</td>
<td>-179.46</td>
</tr>
</tbody>
</table>
Table 4.4 Predictor environmental dataset prediction with Radial Basis function kernel ancestry \((k = 3, \text{Figure 4.19(c)})\)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>genotype Variance Comp</th>
<th>genotype Z Ratio</th>
<th>landrace Variance Comp</th>
<th>landrace Z Ratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>0.002</td>
<td>0.790</td>
<td>0.000</td>
<td>0.233</td>
<td>1068.61</td>
<td>-2135.23</td>
</tr>
<tr>
<td>Annual Mean Diurnal Range</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.936</td>
<td>834.10</td>
<td>-1666.20</td>
</tr>
<tr>
<td>Temperature Seasonality</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.628</td>
<td>1859.06</td>
<td>-3716.12</td>
</tr>
<tr>
<td>Mean Temperature of Warmest Quarter</td>
<td>0.002</td>
<td>0.591</td>
<td>0.000</td>
<td>0.891</td>
<td>898.30</td>
<td>-1794.61</td>
</tr>
<tr>
<td>Mean Temperature of Coldest Quarter</td>
<td>0.001</td>
<td>0.715</td>
<td>0.000</td>
<td>0.877</td>
<td>981.59</td>
<td>-1961.18</td>
</tr>
<tr>
<td>Annual Precipitation</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.226</td>
<td>1146.99</td>
<td>-2291.97</td>
</tr>
<tr>
<td>Bulk Density Mean</td>
<td>0.022</td>
<td>0.887</td>
<td>0.004</td>
<td>1.650</td>
<td>320.00</td>
<td>-638.00</td>
</tr>
<tr>
<td>Cation Exchange Capacity of Soil Mean</td>
<td>0.004</td>
<td>0.318</td>
<td>0.001</td>
<td>0.812</td>
<td>40.07</td>
<td>-78.13</td>
</tr>
<tr>
<td>Weight Percentage of Clay Particles Mean</td>
<td>0.004</td>
<td>0.251</td>
<td>0.002</td>
<td>0.810</td>
<td>-17.18</td>
<td>36.35</td>
</tr>
<tr>
<td>Soil Organic Carbon Content Mean</td>
<td>0.077</td>
<td>0.962</td>
<td>0.000</td>
<td>0.146</td>
<td>174.82</td>
<td>-347.64</td>
</tr>
<tr>
<td>pH Index Measured in Water Solution Mean</td>
<td>0.039</td>
<td>0.888</td>
<td>0.003</td>
<td>1.308</td>
<td>231.70</td>
<td>-461.40</td>
</tr>
<tr>
<td>Weight Percentage of the Silt Particles Mean</td>
<td>0.048</td>
<td>0.847</td>
<td>0.001</td>
<td>0.736</td>
<td>-49.12</td>
<td>100.24</td>
</tr>
<tr>
<td>Weight Percentage of the Sand Particles Mean</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.843</td>
<td>-109.60</td>
<td>221.20</td>
</tr>
<tr>
<td>Precipitation Seasonality</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.442</td>
<td>1426.90</td>
<td>-2851.81</td>
</tr>
<tr>
<td>Precipitation of Driest Quarter</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1270.42</td>
<td>-2538.84</td>
</tr>
<tr>
<td>Solar Radiation kJ/m² day 1...March</td>
<td>0.002</td>
<td>0.756</td>
<td>0.000</td>
<td>0.308</td>
<td>1507.75</td>
<td>-3013.50</td>
</tr>
<tr>
<td>Solar Radiation kJ/m² day 1...July, September</td>
<td>0.009</td>
<td>1.005</td>
<td>0.000</td>
<td>0.479</td>
<td>1119.20</td>
<td>-2236.41</td>
</tr>
<tr>
<td>Wind Speed m/s 1...Annual Mean</td>
<td>0.008</td>
<td>1.065</td>
<td>0.000</td>
<td>0.512</td>
<td>1012.15</td>
<td>-2022.29</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.010</td>
<td>0.796</td>
<td>0.000</td>
<td>0.000</td>
<td>1410.63</td>
<td>-2819.26</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1440.52</td>
<td>-2879.04</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.109</td>
<td>0.952</td>
<td>0.003</td>
<td>1.030</td>
<td>706.10</td>
<td>-1410.21</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.001</td>
<td>0.270</td>
<td>0.000</td>
<td>0.406</td>
<td>1166.47</td>
<td>-2330.94</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.603</td>
<td>823.70</td>
<td>-1645.40</td>
</tr>
</tbody>
</table>
Table 4.5  Independent environmental dataset prediction with Radial Basis function kernel ancestry ($k = 3$, Figure 4.19(d))

<table>
<thead>
<tr>
<th></th>
<th>genotype_VarComp</th>
<th>genotype_Zratio</th>
<th>landrace_VarComp</th>
<th>landrace_Zratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Cover</td>
<td>0.569</td>
<td>0.202</td>
<td>0.000</td>
<td>0.000</td>
<td>1490.47</td>
<td>-2978.93</td>
</tr>
<tr>
<td>Diurnal Temperature Range</td>
<td>0.000</td>
<td>0.000</td>
<td>0.024</td>
<td>0.290</td>
<td>801.84</td>
<td>-1601.68</td>
</tr>
<tr>
<td>Frost Day Frequency</td>
<td>1143.390</td>
<td>0.831</td>
<td>58.586</td>
<td>0.952</td>
<td>477.26</td>
<td>-952.53</td>
</tr>
<tr>
<td>Monthly Average Daily Mean Temperature</td>
<td>19.309</td>
<td>0.745</td>
<td>4.155</td>
<td>1.418</td>
<td>698.67</td>
<td>-1395.35</td>
</tr>
<tr>
<td>Wet Day Frequency</td>
<td>699.238</td>
<td>0.742</td>
<td>9.670</td>
<td>0.314</td>
<td>1108.96</td>
<td>-2215.93</td>
</tr>
<tr>
<td>Potential Evapotranspiration</td>
<td>0.000</td>
<td>0.000</td>
<td>0.022</td>
<td>1.096</td>
<td>646.41</td>
<td>-1290.82</td>
</tr>
<tr>
<td>Precipitation</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>646.41</td>
<td>-1290.82</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>0.001</td>
<td>0.438</td>
<td>0.000</td>
<td>0.002</td>
<td>569.64</td>
<td>-1137.29</td>
</tr>
<tr>
<td>Soil pH</td>
<td>0.000</td>
<td>0.000</td>
<td>0.024</td>
<td>1.600</td>
<td>-208.14</td>
<td>418.27</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.010</td>
<td>0.558</td>
<td>0.000</td>
<td>0.000</td>
<td>984.96</td>
<td>-1967.92</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.045</td>
<td>0.855</td>
<td>0.005</td>
<td>1.411</td>
<td>844.75</td>
<td>-1687.50</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.027</td>
<td>1.589</td>
<td>-115.34</td>
<td>232.67</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.004</td>
<td>0.414</td>
<td>0.002</td>
<td>1.081</td>
<td>481.08</td>
<td>-960.17</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.011</td>
<td>0.577</td>
<td>0.003</td>
<td>1.146</td>
<td>90.80</td>
<td>-179.59</td>
</tr>
</tbody>
</table>
Table 4.6  Predictor environmental dataset prediction with kinship matrix (Figure 4.19(e))

<table>
<thead>
<tr>
<th></th>
<th>genotype_VarComp</th>
<th>genotype_Zratio</th>
<th>landrace_VarComp</th>
<th>landrace_Zratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>0.002</td>
<td>2.646</td>
<td>0.000</td>
<td>0.590</td>
<td>1070.71</td>
<td>-2139.42</td>
</tr>
<tr>
<td>Annual.Mean.Diurnal.Range</td>
<td>0.003</td>
<td>1.200</td>
<td>0.000</td>
<td>0.939</td>
<td>834.59</td>
<td>-1667.19</td>
</tr>
<tr>
<td>Temperature.Seasonality</td>
<td>0.000</td>
<td>1.663</td>
<td>0.000</td>
<td>0.559</td>
<td>1860.40</td>
<td>-3718.81</td>
</tr>
<tr>
<td>Mean.Temperature.of.Warmest.Quarter</td>
<td>0.004</td>
<td>2.601</td>
<td>0.000</td>
<td>1.134</td>
<td>900.98</td>
<td>-1799.95</td>
</tr>
<tr>
<td>Mean.Temperature.of.Coldest.Quarter</td>
<td>0.001</td>
<td>2.325</td>
<td>0.000</td>
<td>1.145</td>
<td>983.54</td>
<td>-1965.08</td>
</tr>
<tr>
<td>Annual.Precipitation</td>
<td>0.001</td>
<td>2.862</td>
<td>0.000</td>
<td>0.144</td>
<td>1150.48</td>
<td>-2298.96</td>
</tr>
<tr>
<td>Bulk.Density..mean.</td>
<td>0.016</td>
<td>2.822</td>
<td>0.005</td>
<td>1.684</td>
<td>320.83</td>
<td>-639.66</td>
</tr>
<tr>
<td>Cation.Exchange.Capacity.of.Soil..mean.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.795</td>
<td>39.99</td>
<td>-77.99</td>
</tr>
<tr>
<td>Weight.Percentage.of.Clay.Particles..mean.</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.789</td>
<td>-17.20</td>
<td>36.39</td>
</tr>
<tr>
<td>Soil.Organic.Carbon.Content..mean.</td>
<td>0.026</td>
<td>1.825</td>
<td>0.001</td>
<td>0.613</td>
<td>173.86</td>
<td>-345.72</td>
</tr>
<tr>
<td>pH.Index.Measured.in.Water.Solution..mean.</td>
<td>0.021</td>
<td>2.059</td>
<td>0.005</td>
<td>1.536</td>
<td>232.26</td>
<td>-462.51</td>
</tr>
<tr>
<td>Weight.Percentage.of.the.Silt.Particles..mean.</td>
<td>0.014</td>
<td>0.875</td>
<td>0.002</td>
<td>0.906</td>
<td>-50.83</td>
<td>103.65</td>
</tr>
<tr>
<td>Weight.Percentage.of.the.Sand.Particles..mean.</td>
<td>0.036</td>
<td>1.937</td>
<td>0.001</td>
<td>0.703</td>
<td>-107.81</td>
<td>217.63</td>
</tr>
<tr>
<td>Precipitation.Seasonality</td>
<td>0.000</td>
<td>0.296</td>
<td>0.000</td>
<td>1.442</td>
<td>1426.93</td>
<td>-2851.86</td>
</tr>
<tr>
<td>Precipitation.of.Driest.Quarter</td>
<td>0.001</td>
<td>1.519</td>
<td>0.000</td>
<td>0.000</td>
<td>1271.46</td>
<td>-2540.93</td>
</tr>
<tr>
<td>Solar.Radiation..kJ.m.2.day.1..March.</td>
<td>0.002</td>
<td>2.000</td>
<td>0.000</td>
<td>0.419</td>
<td>1510.07</td>
<td>-3018.14</td>
</tr>
<tr>
<td>Solar.Radiation..kJ.m.2.day.1..July.September.</td>
<td>0.005</td>
<td>3.699</td>
<td>0.000</td>
<td>0.652</td>
<td>1123.33</td>
<td>-2244.66</td>
</tr>
<tr>
<td>Wind.Speed..m.s.1..Annual.Mean.</td>
<td>0.002</td>
<td>2.225</td>
<td>0.000</td>
<td>0.850</td>
<td>1010.69</td>
<td>-2019.38</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.011</td>
<td>2.861</td>
<td>0.000</td>
<td>0.621</td>
<td>1413.47</td>
<td>-2824.94</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.001</td>
<td>0.366</td>
<td>0.000</td>
<td>0.000</td>
<td>1440.45</td>
<td>-2878.91</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.065</td>
<td>3.624</td>
<td>0.006</td>
<td>1.292</td>
<td>709.54</td>
<td>-1417.08</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.002</td>
<td>0.565</td>
<td>0.000</td>
<td>0.380</td>
<td>1166.64</td>
<td>-2331.29</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.000</td>
<td>0.105</td>
<td>0.000</td>
<td>0.598</td>
<td>823.71</td>
<td>-1645.41</td>
</tr>
</tbody>
</table>
Table 4.7 Independent environmental dataset prediction with kinship matrix (Figure 4.19(f))

<table>
<thead>
<tr>
<th></th>
<th>genotype_VarComp</th>
<th>genotype_Zratio</th>
<th>landrace_VarComp</th>
<th>landrace_Zratio</th>
<th>logLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Cover</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1490.45</td>
<td>-2978.89</td>
</tr>
<tr>
<td>Diurnal Temperature Range</td>
<td>2.007</td>
<td>1.774</td>
<td>0.027</td>
<td>0.315</td>
<td>803.32</td>
<td>-1604.64</td>
</tr>
<tr>
<td>Frost Day Frequency</td>
<td>0.000</td>
<td>0.000</td>
<td>51.707</td>
<td>0.914</td>
<td>475.07</td>
<td>-948.13</td>
</tr>
<tr>
<td>Monthly Average Daily Mean Temperature</td>
<td>15.856</td>
<td>1.543</td>
<td>3.749</td>
<td>1.379</td>
<td>698.86</td>
<td>-1395.72</td>
</tr>
<tr>
<td>Wet Day Frequency</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1107.96</td>
<td>-2213.91</td>
</tr>
<tr>
<td>Potential Evapotranspiration</td>
<td>0.018</td>
<td>0.174</td>
<td>0.021</td>
<td>1.085</td>
<td>1114.77</td>
<td>-2227.54</td>
</tr>
<tr>
<td>Precipitation</td>
<td>2321.693</td>
<td>0.998</td>
<td>0.000</td>
<td>0.000</td>
<td>646.78</td>
<td>-1291.55</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>0.003</td>
<td>1.230</td>
<td>0.000</td>
<td>0.035</td>
<td>570.18</td>
<td>-1138.35</td>
</tr>
<tr>
<td>Soil pH</td>
<td>0.085</td>
<td>2.064</td>
<td>0.022</td>
<td>1.553</td>
<td>-206.06</td>
<td>414.12</td>
</tr>
<tr>
<td>envPC1</td>
<td>0.005</td>
<td>0.373</td>
<td>0.000</td>
<td>0.000</td>
<td>984.70</td>
<td>-1967.41</td>
</tr>
<tr>
<td>envPC2</td>
<td>0.019</td>
<td>1.454</td>
<td>0.005</td>
<td>1.419</td>
<td>843.89</td>
<td>-1685.79</td>
</tr>
<tr>
<td>envPC3</td>
<td>0.099</td>
<td>2.051</td>
<td>0.025</td>
<td>1.548</td>
<td>-113.34</td>
<td>228.67</td>
</tr>
<tr>
<td>envPC4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>1.096</td>
<td>480.92</td>
<td>-959.84</td>
</tr>
<tr>
<td>envPC5</td>
<td>0.026</td>
<td>1.991</td>
<td>0.003</td>
<td>1.134</td>
<td>92.55</td>
<td>-183.10</td>
</tr>
</tbody>
</table>
Table 4.8 LFMM results table

<table>
<thead>
<tr>
<th>GENE</th>
<th>GRMZ</th>
<th>CHROM</th>
<th>Gene Name(s)</th>
<th>ENV</th>
<th>Gene Function</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zm00001d027546</td>
<td>GRMZM2G173372</td>
<td>1</td>
<td>bhlh35 - bHLH-transcription factor 35</td>
<td>Soil pH</td>
<td>Anthocyanin, lignin biosynthesis</td>
<td>Li et al., 2019; Barrière et al., 2016; Xie et al., 2012</td>
</tr>
<tr>
<td>Zm00001d027723</td>
<td>GRMZM2G813886</td>
<td>1</td>
<td>ZmWBC11</td>
<td>Solar Radiation (July-September)</td>
<td>Pathogen resistance, lipase secretion</td>
<td>Pang et al., 2013</td>
</tr>
<tr>
<td>Zm00001d028046</td>
<td>GRMZM2G364510</td>
<td>1</td>
<td>Emp602 - empty pericarp602</td>
<td>Soil Bulk Density</td>
<td>Kernel development</td>
<td>Ren et al., 2019</td>
</tr>
<tr>
<td>Zm00001d029489</td>
<td>GRMZM2G037630</td>
<td>1</td>
<td>ca2p3 - CCAAT-HAP2-transcription factor 23, nuclear transcription factor Y subunit A-7</td>
<td>Soil Cation Exchange Capacity</td>
<td>Starch biosynthesis, response to waterlogging, general stress response</td>
<td>Zhang et al., 2019b; Rajhi et al., 2011; Liu et al., 2014</td>
</tr>
<tr>
<td>Zm00001d031543</td>
<td>GRMZM2G132712</td>
<td>1</td>
<td>pco101694</td>
<td>Soil Cation Exchange Capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d032717</td>
<td>GRMZM2G053413</td>
<td>1</td>
<td></td>
<td>Soil Bulk Density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d033902</td>
<td>GRMZM2G091964</td>
<td>1</td>
<td>ca2p6 - CCAAT-HAP2-transcription factor 21b, ZmNF-YA6</td>
<td>Temperature of Coldest Quarter</td>
<td>Pathogen resistance, drought stress response, low nitrate stress response</td>
<td>Luan et al., 2015; Zhao et al., 2013; Xu et al., 2011; Luo et al., 2015</td>
</tr>
<tr>
<td>Zm00001d034517</td>
<td>GRMZM2G062394</td>
<td>1</td>
<td>putative serine/threonine protein phosphatase superfamily protein</td>
<td>Soil Organic Carbon Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d034192</td>
<td>GRMZM2G062354</td>
<td>2</td>
<td>pco141842</td>
<td>Temperature of Warmest Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d032003</td>
<td>GRMZM2G045049</td>
<td>2</td>
<td>probable phosphoribosylformylglycinamidine synthase</td>
<td>Soil Organic Carbon Content, Soil Sand Percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d002277</td>
<td>GRMZM2G383122</td>
<td>2</td>
<td>BAG family molecular chaperone regulator 1</td>
<td>Soil Cation Exchange Capacity</td>
<td>Pathogen resistance</td>
<td>Li et al., 2018</td>
</tr>
<tr>
<td>Zm00001d002533</td>
<td>GRMZM2G050564</td>
<td>2</td>
<td>pto12 - pyrabactin resistance-like protein12</td>
<td>Soil Sand Percentage</td>
<td>Drought resistance</td>
<td>Peng et al., 2017</td>
</tr>
<tr>
<td>Zm00001d002602</td>
<td>GRMZM2G556131</td>
<td>2</td>
<td>DHDPS1</td>
<td>Soil Silt Percentage</td>
<td>Lysine biosynthesis, general stress response</td>
<td>Liu et al., 2016</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Name</td>
<td>Type</td>
<td>Description</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
<td>------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d003439</td>
<td>GRMZM2G393762</td>
<td>2</td>
<td>Altitude, Temperature of Warmest Quarter</td>
<td>(Muraya et al., 2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d004445</td>
<td>GRMZM2G1111917</td>
<td>2</td>
<td>Altitude, Temperature of Warmest Quarter, Precipitation Seasonality, Soil Clay Percentage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d004559</td>
<td>GRMZM2G165597</td>
<td>2</td>
<td>probable LRR receptor-like serine/threonine-protein kinase At1g06840</td>
<td>Soil Organic Carbon Content (Coleman, 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d005888</td>
<td>GRMZM2G092131</td>
<td>2</td>
<td>heatH - HSP-transcription factor 1</td>
<td>Soil Bulk Density, Wind Speed Heat shock protein (Benaud, 2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d006205</td>
<td>GRMZM2G312501</td>
<td>2</td>
<td>BRD1</td>
<td>Soil Bulk Density, Soil Organic Carbon Content Brassinosteroid biosynthesis (Besbrugge et al., 2018; Makarevitch et al., 2012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d006303</td>
<td></td>
<td>2</td>
<td>Soil Bulk Density, Soil pH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d006570</td>
<td>pco092737</td>
<td>2</td>
<td>Solar Radiation (March), Temperature Seasonality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d006712</td>
<td></td>
<td>2</td>
<td>Wind Speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d009341</td>
<td>GRMZM2G068217</td>
<td>3</td>
<td>ein2 - ethylene insensitive 2</td>
<td>Solar Radiation (March) Pathogen resistance, low Phosphate stress response, general stress response (WU et al., 2015; Alonso et al., 1999; SU et al., 2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d009455</td>
<td>GRMZM2G050890</td>
<td>3</td>
<td>prenylated rab acceptor family protein</td>
<td>Precipitation Seasonality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d009910</td>
<td>GRMZM2G320521</td>
<td>3</td>
<td>Protein ASPARTIC PROTEASE IN GUARD CELL 1</td>
<td>Precipitation of Driest Quarter Precipitation Seasonality Drought stress response (Yao et al., 2012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d040034</td>
<td>GRMZM2G054050</td>
<td>3</td>
<td>pca02427, Multicopper oxidase Like LPR2</td>
<td>Soil Organic Carbon Content Low Phosphate stress response (Svistoonoff et al., 2007; Calderón-Valenzuela et al., 2011; Zhang et al., 2014; Schlüter et al., 2012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d040142</td>
<td>GRMZM2G051822</td>
<td>3</td>
<td>pca0103509</td>
<td>Altitude, Temperature of Warmest Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Expression Level</td>
<td>Function/Response</td>
<td>Solar Radiation (Month)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d040166</td>
<td>GRMZM2G425751, AC244199, LG0603, GRMZM2G425747, GRMZM2G464723, GRMZM2G306661</td>
<td>3</td>
<td>atm1 - ataxia-telangiectasia mutated, serine/threonine-protein kinase ATM</td>
<td>Solar Radiation (March)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d043168</td>
<td>GRMZM2G402092</td>
<td>3</td>
<td>Soil Cation Exchange Capacity, Precipitation of Driest Quarter</td>
<td>Soil Cation Exchange Capacity, Precipitation of Driest Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d043294</td>
<td>GRMZM5G851026</td>
<td>3</td>
<td>Soil Organic Carbon Content</td>
<td>Drought stress response (Li et al., 2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d043347</td>
<td>GRMZM2G390221</td>
<td>3</td>
<td>Soil Bulk Density, Altitude, Temperature of Coldest Quarter, Temperature of Warmest Quarter, Precipitation of Driest Quarter, Temperature Seasonality</td>
<td>Soil Organic Carbon Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d0435221</td>
<td>GRMZM2G379327</td>
<td>4</td>
<td>Soil Organic Carbon Content</td>
<td>Soil Organic Carbon Content, Temperature of Warmest Quarter, Temperature Seasonality</td>
<td>Low phosphate stress response (Gowda, 2017; Gupta et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Zm00001d051399</td>
<td>GRMZM2G304410</td>
<td>4</td>
<td>Soil Bulk Density, Soil Organic Carbon Content, Solar Radiation (July-September), Temperature Seasonality</td>
<td>Soil Organic Carbon Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d052074</td>
<td>GRMZM2G044670</td>
<td>4</td>
<td>Soil Organic Carbon Content</td>
<td>Soil Organic Carbon Content, Temperature of Warmest Quarter, Solar Radiation (July-September)</td>
<td>Vascular, pericarp development (Zhang et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Zm00001d052130</td>
<td>GRMZM2G149150</td>
<td>4</td>
<td>bZIP43, pco105355</td>
<td>Altitude, Temperature Diurnal Range</td>
<td>Vascular, pericarp development (Zhang et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Zm00001d013783</td>
<td>GRMZM2G178767</td>
<td>5</td>
<td>dof12 - Ca2+ - Dof transcription factor 12, ZmDof10</td>
<td>Altitude</td>
<td>Salt stress response, pollen development</td>
<td></td>
</tr>
<tr>
<td>Zm00001d015846</td>
<td>GRMZM2G029979</td>
<td>5</td>
<td>bzip28 - bZIP transcription factor 28, transcription factor TGA19</td>
<td>Soil Sand Percentage</td>
<td>(Jiang et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Zm00001d016385</td>
<td>GRMZM2G015112</td>
<td>5</td>
<td>dla1 - dihydrolipoamide S-acyltransferase1</td>
<td>Soil Sand Percentage</td>
<td>Drought stress response</td>
<td>(Xin et al., 2018)</td>
</tr>
<tr>
<td>Zm00001d015991</td>
<td>GRMZM2G74823</td>
<td>6</td>
<td>kinesin-like protein KIN-1</td>
<td>Soil Organic Carbon Content</td>
<td>Pathogen resistance</td>
<td>(Majumdar et al., 2017)</td>
</tr>
<tr>
<td>Zm00001d015841</td>
<td>GRMZM2G353905</td>
<td>6</td>
<td>Altitude, Temperature of Warmest Quarter</td>
<td>Soil Organic Carbon Content</td>
<td>Pathogen resistance</td>
<td>(Jiang et al., 2012)</td>
</tr>
<tr>
<td>Zm00001d036510</td>
<td>GRMZM2G30671</td>
<td>6</td>
<td>hs81 - Homeobox transcription factor 81, homeobox-leucine zipper protein HOX18</td>
<td>Soil Organic Carbon Content</td>
<td>(Jiang et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>Zm00001d036751</td>
<td>GRMZM2G169333</td>
<td>6</td>
<td>TIDP2935</td>
<td>Solar Radiation (July-September)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d018957</td>
<td>GRMZM2G100803</td>
<td>7</td>
<td>60S ribosomal protein L4-1</td>
<td>Soil Organic Carbon Content</td>
<td>Heat stress response</td>
<td>(Frey et al., 2015)</td>
</tr>
<tr>
<td>Zm00001d021206</td>
<td>GRMZM2G065461</td>
<td>7</td>
<td>s1940087602</td>
<td>Altitude, Temperature of Coldest Quarter, Temperature of Warmest Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d022169</td>
<td>GRMZM2G381395</td>
<td>7</td>
<td>rpoT1 - RNA polymerase T phase-like 1</td>
<td>Soil pH, Soil Organic Carbon Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d010098</td>
<td>GRMZM2G311386</td>
<td>8</td>
<td>Wind Speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zm00001d011386</td>
<td>GRMZM2G024738</td>
<td>8</td>
<td>uroporphyrinogen decarboxylase</td>
<td>Precipitation of Driest Quarter, Solar Radiation (March)</td>
<td>Pathogen resistance, chlorophyll biosynthesis</td>
<td>(Donaldson, 2014; Kretschmer et al., 2017; Donaldson et al., 2013)</td>
</tr>
<tr>
<td>Zm00001d011793</td>
<td>GRMZM2G167365</td>
<td>8</td>
<td>late embryogenesis abundant protein-related / LEA protein-related</td>
<td>Soil Organic Carbon Content</td>
<td>Drought stress response</td>
<td>(Goyal et al., 2005)</td>
</tr>
<tr>
<td>Zm00001d011828</td>
<td>GRMZM2G133926</td>
<td>8</td>
<td>SR452</td>
<td>Temperature of Warmest Quarter</td>
<td>Drought stress response</td>
<td>(Li et al., 2014)</td>
</tr>
<tr>
<td>Zm00001d035519</td>
<td>GRMZM2G255515</td>
<td>9</td>
<td>oligopeptide transporter 9</td>
<td>Altitude, Temperature of Warmest Quarter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4.8 (continued)

| Zm00001d047566 | GRMZM2G475360 | 9 | si01402104b | Wind Speed |
| Zm00001d047594 | GRMZM2G064725 | 9 | SNP1-related protein kinase regulatory subunit beta-1 | Soil Silt Percentage |
| Zm00001d047631 | GRMZM2G007590 | 9 | umc2342, spliceosomal protein | Solar Radiation (July-September) |
| Zm00001d047699 | GRMZM2G047894 | 9 | pco123655 | Soil Organic Carbon Content |
| Zm00001d047739 | GRMZM2G147534 | 9 | S-adenosyl-L-methionine-dependent methyltransferase superfamily protein | Soil pH, Solar Radiation (July-September) |
| Zm00001d047904 | GRMZM2G147534 | 9 | S-adenosyl-L-methionine-dependent methyltransferase superfamily protein | Soil pH, Solar Radiation (July-September) |
| Zm00001d048051 | AC211178.3 | 9 | Precipitation Seasonality |
| Zm00001d023910 | GRMZM2G082167 | 10 | adaptor complex medium subunit family protein | Soil Cation Exchange Capacity |
| Zm00001d025538 | GRMZM2G097599 | 10 | Soil pH |
| Zm00001d025965 | GRMZM2G109648 | 10 | AY108476, pco136451a | Soil Organic Carbon Content |
| Zm00001d025803 | GRMZM2G047894 | 10 | Soil Cation Exchange Capacity |
| Zm00001d025992 | GRMZM2G025992 | 10 | Soil Organic Carbon Content |
| Zm00001d026476 | GRMZM2G143489 | 10 | 6,7-dimethyl-8-ribityllumazine synthase | Soil Bulk Density, Soil Organic Carbon Content |

Notes: (Li et al., 2009) (Renaud, 2015) (Geniaux et al., 2017) (Joshi and Chang, 1998; Moffatt and Weretilnyk, 2001)
CHAPTER 5. LOCAL ADAPTATION AMONG LANDRACES WITH COMPLEX POPULATION STRUCTURE

Authors: Garrett M Janzen\textsuperscript{1}, John Nason\textsuperscript{1}, Matthew B. Hufford\textsuperscript{1}

\textsuperscript{1}Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

5.1 Abstract

In my third chapter, I use genotypic data collected in the reciprocal transplant experiment and further investigate the partitioning of adaptive genetic variation into groups across geographic space. Landrace genetic population structure is largely shaped by geographic distance, latitude, and elevation, and is consistent with demographic historical migration and bottleneck events. To incorporate these complex patterns of relatedness across the Americas, we looked for phenotypic trends across the axes of variation of an eigen-decomposition of a genetic distance matrix which capture elements of genetic structure between populations that correlate to latitude and elevation. Traits that varied more or less than expected given neutral genetic variance were identified as under divergent or stabilizing selection, receptively. I also identified SNPs with high $F_{ST}$ between highland and lowland populations in either Mexico or South America, or both, to find genes involved in highland adaptation unique to one continent or the other or parallel highland/lowland adaptation.
5.2 Introduction

One of the main foci in the study of local adaptation is the identification of divergent selection pressures that lead to local adaptation. Divergent selection is frequently evinced by subpopulations displaying diverged phenotypes or allele frequencies. Common garden experiments are designed for the purpose of reducing environmental effects to facilitate estimation of phenotypic differences between populations that is due to genetic variation. Divergent phenotypes and genotypes are particularly indicative of local adaptation when their divergence is in concert with environmental or ecological clines of conceivable impact on survival and/or reproduction.

However, the central challenge inherent in the identification of divergent selection is in distinguishing it from the neutral effects of genetic drift or the historical effects of demography that are unrelated to divergent selective pressures, which can by chance produce phenotypic and genotypic patterns that resemble those produced by local adaptation. One popular method of investigating whether subpopulations with divergent phenotypes are truly under divergent selective pressures is to compare the genetic divergence of a trait \( Q_{ST} \) to the degree of genetic divergence due to neutral stochastic effects \( F_{ST} \). This comparison, known as \( Q_{ST} - F_{ST} \), has successfully identified numerous instances of local adaptation (McKay and Latta, 2002), but is not without conceptual limitations. Firstly, \( Q_{ST} - F_{ST} \) requires the \textit{a priori} delineation of subpopulations, and presumes equal relation between them. This is a difficult and reductive task in metapopulations with complex and dynamic population structure. Second, estimation of \( Q_{ST} \) requires estimates of additive genetic variance between and within populations, which are typically gained from controlled breeding experiments. Some studies replace \( Q_{ST} \) with phenotypic variation \( P_{ST} \), though environmental effects may inflate \( P_{ST} \) and lead to false positives, especially when data are collected from organisms reared outside of experimentally controlled environments (common gardens). Though \( Q_{ST} - F_{ST} \) (and \( P_{ST} - F_{ST} \)) comparisons are built on the assumption that \( Q_{ST} \) and \( F_{ST} \) are equal for traits that are not under selection, meta-analyses show that \( Q_{ST} \) is frequently greater than \( F_{ST} \), and the two are sometimes poorly correlated (McKay and Latta, 2002).
In addition to identifying traits under divergent selection, we seek to identify alleles contributing to adaptation. A suite of methods have risen up around the task of identifying alleles under divergent selection (Narum and Hess, 2011). Most are conceptually related to the identification of SNPs that are $F_{ST}$ outliers or that are significantly associated with environmental variables, and some account for the confounding effects of genetic drift and isolation by distance.

The study of adaptation in crop populations is interesting for economic as well as academic reasons. Maize was domesticated in the lowlands of the Balsas River Valley in Mexico from the teosinte taxon *Zea mays* subsp. *parviglumis* roughly 9000 years BP (Matsuoka et al., 2002). From there, maize was carried across across North America into South America as early as 6000 years BP (Grobman et al., 2012; Bush et al., 1989), north into the present-day United States by about 4500 years BP (Merrill et al., 2009), and around the world (Tenaillon and Charcosset, 2011; Van Heerwaarden et al., 2011a). The demographic history of geographic range and population size expansion has introduced a series of serial founder effects, decreasing neutral genetic diversity in populations further from the center of domestication (particularly populations from South America (Wang et al., 2017)). Population structure among maize landraces is shaped broadly by altitude, latitude, geographic distance, and introgression with *Zea mays* subsp. *mexicana* (Arteaga et al., 2016; Vigouroux et al., 2008; Hufford et al., 2013). Though overall genetic differentiation between landraces is quite low (Caldu-Primo et al., 2017), landraces yet retain numerous highly divergent phenotypes between them (Wellhausen et al., 1952; Eagles and Lothrop, 1994). Because of the complex population structure found between maize landraces, defining subpopulations between which divergent selection has acted is a nontrivial concern in the identification of maize landrace adaptations.

We defined partitioned 120 New World maize landraces into four populations (Mexican highland, Mexican lowland, South American highland, South American lowland) which I show to capture meaningful phenotypic divergence. We here characterize the population structure of maize landraces across the Americas with greater resolution, and incorporate more nuanced approaches to detecting phenotypic and genotypic divergence, contrasting them to traditional methods. We evaluated
phenotypic divergence with traditional $P_{ST} - F_{ST}$ and with $Q_{PC}$ (Josephs et al., 2019), which detects phenotypic divergence across numerous axes of population structure. We use pairwise $F_{ST}$ values between populations defined a priori and between populations defined by sNMF, which flexibly models population structure in populations with high admixture and complex ancestry (Martins et al., 2016).

5.3 Methods

5.3.1 Population Structure

5.3.1.1 Subpopulations

Accessions were selected based on elevation and latitude with the goal of producing four populations (Mexican highland, Mexican lowland, South American highland, South American lowland). To further test the affect of latitude within these populations, we split each population into northern and southern populations, resulting in a total of eight subpopulations comprised of levels of continent, elevation, and latitude.

One plant from 92 families were genotyped (DArTseq) in each garden (184 individual plants total). In total, 19,450 high-quality SNP markers were used. Pairwise Euclidean allele frequency distances between the four original populations and between the eight continent/elevation/latitude subpopulations were calculated (function gl.dist.pop, R package dartr (Gruber et al., 2018)). I used population graphs (Dyer and Nason, 2004) to provide a representation of genetic covariance evolutionary history between the eight subpopulations described above.

Cross-entropy validation of ancestral gene pools (function snmf, R package LEA (Frichot and François, 2015)) and Bayesian Information Criterion (BIC, function find.clusters, R package adegenet (Jombart and Ahmed, 2011)) were used independently to identify the optimal level of population complexity. Both functions converged on an optimal population structure complexity of $k = 3$ groups (Figure 5.1), which roughly correspond to Mexican Highland, South American Highland,
and Lowland groups. To facilitate hierarchical population structure comparisons above and below this threshold, $k$ values from 2 to 5 were additionally considered for specific analyses.

Population structure was partitioned into $k = 2 - 5$ ancestral population groups and plotted over geographic space (TESS3, (Caye et al., 2016), Figures 5.2, 5.3, 5.4, 5.5). Sparse Non-negative Matrix Factorization (sNMF) uses allele frequencies to define ancestral populations, between which $F_{ST}$ is calculated.

Isolation by distance regressions were conducted at the level of individuals and continent/elevation/latitude subpopulations. Pairwise genetic distance for each combination of genotyped individuals was calculated and regressed to pairwise Euclidean geographic distance. A Mantel test was used to determine significance of correlation between genetic and geographic distance matrices.

Genetic similarity of individuals was demonstrated in a Principal Coordinates Analysis (PCoA) plot (R function gl.pcoa, package dartr (Gruber et al., 2018)). Color groups overlaying these PCoA plots demonstrate the various clustering strategies.

### 5.3.1.2 Discriminant Analysis of Principal Components

Discriminant Analysis of Principal Components (DAPC (Jombart and Collins, 2015)) synthesizes discriminant functions which optimize between-group variance and minimize within-group variance (contrary to PCoA, which describes global genetic variance between individuals). I used DAPC to assess the relationship and separation of genetic clusters identified at the levels specified above ($k = 3$ sNMF genetic groups, 4 hypothesized populations, 8 latitudinally sub-divided subpopulations).

I also tested the ability of these discriminant functions to predict group membership of unknown grouping. Individuals were randomly partitioned into training ($n = 100$) and testing ($n = 80$) datasets, and clusters were evaluated by the percent success at which they accurately predict individual membership.
5.3.2 Phenotypic Divergence

5.3.2.1 $P_{ST} - F_{ST}$ Comparison

Phenotypic data were collected from a reciprocal transplant across an elevational gradient in Mexico (described in Chapter 3). A linear mixed effects model was used to partition phenotypic variance between population, landrace accession line, and garden/block.

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

Within-population and between-population variances were calculated with the R function VarCorr (R package lme4 (Bates et al., 2014)), and were used to calculate $P_{ST}$ as below:

$$F_{ST} = \frac{\sigma_{GB}^2}{\sigma_{GB}^2 + \sigma_{GW}^2}$$

$$P_{ST} = \frac{\sigma_{GB}^2}{\sigma_{GB}^2 + 2\sigma_{GW}^2}$$

$F_{ST}$ was calculated with the R function fst.each.snp.hudson (R package dartR, (Gruber et al., 2018)), and distributions of $F_{ST}$ were taken between highland and lowland populations from the same continent.

5.3.2.2 $Q_{PC}$

I used R package calcQpc (Josephs et al., 2019) to identify phenotypes with greater diverge than expected given neutral expectations. I modified the calcQpc function to provide $p$-values for both upper and lower tails of the F distribution (corresponding to diversifying and stabilizing selection, respectively), and ran this analysis in three batches; lowland garden phenotype data, highland garden phenotype data, and combined phenotype data.

5.3.3 Genotypic Divergence

$F_{ST}$ values for all SNPs were calculated between populations. In particular, we compared elevational population pairs (Mexican Highland vs. Mexica Lowland, South American Highland vs. South American Lowland, and all Highland vs. all Lowland). As in (Caldu-Primo et al., 2017),
the top 1\% highest $F_{ST}$ SNPs were compared between these populations to find SNPs involved in highland/lowland adaptation in Mexico, or South America, or in both.

$F_{ST}$ outliers were also identified with reference to sNMF ancestry coefficients at $k = 2 − 5$. Statistical significance of population differentiation was determined with Benjamini Hochberg method (false discovery rate $= 0.001$ (Benjamini and Hochberg, 1995)).

5.4 Results

5.4.0.1 Population Distances

Euclidean allele frequency distance was computed between populations (Table 5.1) and subpopulations (Table 5.2).

There is a strong positive relationship between geographic and genetic distance (Figure 5.6). The Mantel test between geographic and genetic distance matrices is significant at the continent/elevation/latitude population level ($r = 0.5527, p = 0.015$) and the individual level ($r = 0.427, p = 0.001$), indicating that these distance matrices are correlated. Popgraph finds that subpopulation distances correspond both to post-domestication range expansion and latitude (Figure 5.7). Mexican subpopulations are more connected across latitude and elevation than are South American subpopulations.

Relationships between populations, subpopulations, and sNMF genetic clusters are demonstrated in PCoA plots in Figure 5.8.

5.4.0.2 DAPC

DAPC was used to discriminate groups based on genetic cluster, population, and subpopulation. Discriminant functions were 100\% effective at predicting genetic cluster (Figure 5.9(b)), though were less effective at predicting population (86.3\%, Figure 5.9(d)) and subpopulation (55.0\%, Figure 5.9(f)).

Individuals predicted into the wrong population/subpopulation are more common in South America than in Mexican populations. For populations, incorrect group predictions mainly result
in individuals being placed in groups of the correct continent but wrong elevation division. For subpopulations, incorrect group predictions mainly result in individuals being placed in groups of the correct continent and elevation but wrong latitude division, demonstrating the relatively weak latitudinal effect within these populations.

5.4.1 $P_{ST}/F_{ST}$ Comparison

$P_{ST}$ values for quantitative traits were plotted against the distribution of $F_{ST}$ values (Figure 5.10). Four-way $P_{ST}$ comparisons between all populations, as well as two-way comparisons between Mexican Highland and Lowland and between South American Highland and Lowland, were conducted.

Most quantitative traits measured have $P_{ST}$ values far greater than expected given the distribution of $F_{ST}$ values (mean $F_{ST} = 0.004$). δ$^{13}$C (d13C), days to anthesis (DTA), days to silking (DTS), ear height (EH), leaf sheath macrohair density (M_DENsolid), plant height (PH), and tassel branch number (TBN) have very high $P_{ST}$ values for all sets of comparisons. Barrenness (BRN) and tassel length (TL) have high $P_{ST}$ when comparing only Mexican populations, and uniform leaf sheath pigment extent (P_EXTsolid) and intensity (P_INTsolid) and non-uniform pattern leaf sheath pigment extent (P_EXTspot) are significant when considering only South American populations. The traits with the strongest $P_{ST}$ between all four populations were sheath macrohair density, tassel branch number, flowering time (days to silking, days to anthesis), and ear height.

5.4.2 $Q_{PC}$

Traits were plotted against PCs of genetic relatedness (Figure 5.11). $Q_{PC}$ finds strongly divergent phenotypic traits across several axes of population differentiation (Figure 5.12). Population structure (population groups and sNMF $k = 3$ groups) across the first six axes is demonstrated in Figures 5.13, 5.14, and 5.15. Phenotypic divergence is most notable across PC3, which is linearly correlated with altitude, annual mean precipitation, and annual mean temperature, and PC6 (Table 5.3). Ear height, tassel branch number, days to silking and anthesis, and solid-pattern macrohair
density demonstrate highly significant trait divergence ($p < 0.001$). Plant height is also strongly diverged ($p < 0.01$).

Though several traits were stable across PCs of relatedness, no traits show strong statistical evidence of stabilizing selection (Figure 5.12(d), 5.12(e), 5.12(f)).

### 5.4.3 Candidate Genes

Of $n = 19,450$ total SNPs, we took the top 1% of SNPs with the highest $F_{ST}$ between pairs of populations. Though convergent evolution at the genetic level is reported to be rare in maize (Takuno et al., 2015), we found 20 genes to have high $F_{ST}$ (top 1% tail) between highland and lowland populations in both Mexico and South America. These 20 genes are candidates for highland adaptation common to maize landraces from across the Americas. Of these 20 SNPs, 9 are found within characterized genes.

The most promising highland adaptation candidate in this subset is PPDK1 (GRMZM2G306345), a key gene involved in Carbon fixation during the PEP-regeneration phase during C4 photosynthesis (Chastain et al., 2011; Yan et al., 2018; Dong et al., 2018; Lappe et al., 2018) and endosperm development (Lappe et al., 2018). PPDK is highly expressed in leaf mesophyll (Chastain et al., 2011), and expression is elevated in response to cold, dry, high-light highland conditions (in the C4 grass Miscanthus lutarioriparius (Xing et al., 2016)), hypoxic conditions (Mustroph et al., 2014), and to a lesser extent, root flooding conditions (Rajhi et al., 2011). Together, these patterns are consistent with adaptations to highland conditions by regulating photosynthetic rate and metabolism to save energy and improve water use efficiency (Mustroph et al., 2014; Lappe et al., 2018).

GRMZM2G019746 is listed as either 4-coumarate–CoA ligase-like 5 (Bosch et al., 2011) or AMP-dependent synthetase and ligase (ADS) (Wen et al., 2014). 4-coumarate–CoA ligase-like 5 plays a role in phenylpropanoids in maize stalk internodes (Bosch et al., 2011), and ADS is active in flavinoid biosynthesis in maize kernels (Wen et al., 2014). What is clear is that this gene is involved in biosynthesis of phenylpropanoids or their derivatives, such as flavonoids. Flavonoids function in response to many biotic and abiotic stresses (Dixon and Paiva, 1995; Winkel-Shirley,
but also play a role in human use traits of interest, such as nutrition, flavor, and medicinal properties (Tungmunnithum et al., 2018).

GRMZM2G083076 has predicted function as 26S proteasome non-ATPase regulatory subunit 13. The 26S proteasome degrades targeted proteins, a function of transcription, signal transduction, and cell cycle regulation (Tanaka and Tsurumi, 1997; Ferrell et al., 2000).

GRMZM2G015542 codes for a putative cytochrome P450 superfamily protein. This superfamily of proteins performs many functions, and their expression patterns suggest roles in plant defense to biotic and abiotic environmental pressures (Narusaka et al., 2004).

Zm00001d021613 codes for an ankyrin repeat domain-containing protein 13C, part of the ankyrin repeats (ANK) gene family. ANKs function in cell growth, signal transduction, and regulation of cell cycle, and some function in stress response (Jiang et al., 2013).

sNMF identified SNPs that have high \( F_{ST} \) between ancestral populations. Statistical significance is determined with Benjamini Hochberg method (false discovery rate = 0.001). The number of SNPs with significant \( p \)-values increases as greater number of comparisons between \( k \) ancestral populations are considered. For \( k = 2 − 5 \), 1.23%, 2.88%, 3.16%, and 5.14% of SNPs were significantly diverged between subpopulations, respectively.

5.5 Discussion

5.5.1 Population Structure

By contrasting \( k \) values above and below the optimal \( k \) value, we can better elucidate the hierarchical population structure present in the maize landrace metapopulation. More general population divisions are found at the lower \( k \) values, more specific signals of sub-structure are demonstrated at higher \( k \) values, and the optimal \( k \) value \( k = 3 \) demonstrates the degree of population structure that is best predicts masked genotypes. The broadest signal of population \( (k = 2) \) mostly separates North America from South America, although half of South American lowland (mostly northern) groups with North America. The optimal \( k = 3 \) separates Mexican maize by high- and low- elevation groups. The next strongest structural division at \( k = 4 \) mostly divides
high-elevation Mexican maize into central and north-western groups, which roughly correspond to previously-defined Mexican Highland and West Mexico, respectively (Van Heerwaarden et al., 2011b). Neither $k = 3$ nor $k = 4$ address additional structure within South America. At $k = 5$, an additional group in South America is found, between the lowland and southern South American groups. This group roughly corresponds to either Coastal Brazil (Van Heerwaarden et al., 2011b), Chaco Area or Amazonian Zone (Bedoya et al., 2017), or Middle South America (Vigouroux et al., 2008), and is represented by few individuals.

DAPC also arrives at an optimal $k = 3$. At this level of population structure, discriminant functions effectively partition individuals into their respective subpopulations. At $k = 8$, there exists greater overlap between clusters, and the discriminant functions are moderately capable of predicting individual group membership. Individuals from South American Lowland South are particularly poorly predicted, and many individuals are incorrectly predicted as South American Low North. Aside from a few Mexican Low South predicted as South American Low North, no individuals are incorrectly predicted to be members of subpopulations from the opposite continent.

Choice of $k$ should be interpreted with caution. Clustering of continuous genetic variation into groups is by its very nature reductive, and results may be strongly influenced by genotyping efforts and choice of accession inclusion. Our finding of optimal $k = 3$ is therefore not necessarily a biological reality, though other surveys of maize landrace population structure arrive at this same conclusion (Vigouroux et al., 2008; Bedoya et al., 2017). Our panel of landrace accessions was largely devoid of samples from Central America, northern South America, the southern Andean range, and the Caribbean, so TESS3 ancestry projections across those regions are more speculative. Other studies (Van Heerwaarden et al., 2011b) find that Central American, northern South America, and the Caribbean are genetically intermediate between the rest of Mesoamerica and South America (Van Heerwaarden et al., 2011b), though still distinguishable (Van Heerwaarden et al., 2011b; Vigouroux et al., 2008). The southern Andean range is influenced heavily by US-imported varieties (Van Heerwaarden et al., 2011b; Vigouroux et al., 2008) and therefore likely would not group with the northern Andean varieties as shown in our map plots.
5.5.2 Phenotypic Divergence

In systems like maize landraces, relatedness between populations and sub-populations is complicated due to frequent and ongoing gene flow, multiple migration and bottleneck events, within-population structure, and introgression between domesticated maize and crop wild relatives (predominantly in the Mexican highlands). These complications make drawing population boundaries around like individuals difficult. $Q_{PC}$ circumvents this complication by PCA-transforming a genetic similarity matrix and using the most informative (low-order) PCs to capture between-population structure and the least informative (high-order) PCs to capture within-population variation and to estimate $V_A$. Rather than looking for divergence between pre-defined populations, $Q_{PC}$ looks for divergence between major axes of genetic relatedness.

As expected, these results show that $P_{ST} - F_{ST}$ comparisons find much greater phenotypic divergence than does $Q_{PC}$. $P_{ST} - F_{ST}$ seems to overestimate additive genetic variance between populations. $Q_{PC}$, on the other hand, finds strong divergence only for a few traits. Both $P_{ST} - F_{ST}$ and $Q_{PC}$ identify ear height, tassel branch number, days to silking and anthesis, and solid-pattern macrohair density as the the traits with the highest phenotypic divergence. These traits diverge most strongly across PC3, and to a lesser extent, PC6. PC3 situates South American Highland near 0, positive values for Mexican Highland, and negative values for Mexican and South American Lowlands. PC6 also situates Highland populations near 0, and lowland populations dispersed. Among the top 6 PCs, PC3 captures the strongest variation in altitude, annual mean precipitation, and temperature (Table 5.3). The population structure that is captured by PC6 is less obvious, but seems to capture meaningful differentiation within the Mexican and South American Lowlands that does not correspond to elevation, precipitation, or temperature. These axes of relatedness with unclear interpretation are the types of complex population structure that $Q_{PC}$ is designed to uncover. Further investigation into the nature of PC6 may yield insights into population dynamics across maize landraces, perhaps especially among lowland landraces.

No traits show strong signals of stabilizing selection. In the Highland garden, three fitness-related traits (ear-producing stand count, anthesis-silking interval, tassel branch number) are
marginally significant, as are two metrics of anthocyanin expression (solid-pattern extent and spot-pattern intensity) in the lowland garden and when averaged between garden. We would expect traits such traits to be under stabilizing selection between populations, as they are directly correlated (either positively or negatively) with fitness.

Simulations by Josephs and colleagues (Josephs et al., 2019) show that environmental effects reduce statistical power in \(Q_{PC}\). Our implementation of this method finds that phenotype data pooled from multiple common garden experiments can still reveal strong population differentiation. However, traits with strong environmental and genetic effects may be missed by averaging in this way. Flowering time (days to anthesis and days to silking), though strongly influenced by garden site, are also strongly diverged between highland and lowland populations, both in our study and in published literature (Mercer et al., 2008). This signal is detected when the scope is restricted to either the highland garden or the lowland garden, but when these two datasets are pooled, it is lost. On the other hand, if a trait value is not environmentally plastic, then pooling common garden data can increase statistical power. Macrophair density is shown in Chapter 3 to have low plasticity between these two common garden sites. Though macrohair density divergence is found across several axes of population structure based on data collected at either garden, but combining these gardens, macrohair density is revealed to be diverged across a greater number of such axes. Therefore, knowledge of the biological system and of the plasticity of the trait in question will inform the utility of pooling data as described here.

5.5.3 \(F_{ST}\) Outliers

Most modern genomic selection scans are capable of finding statistically significant population divergence. It is beyond the scope of this study to quantify the statistical power and accuracy of these methods. Rather, we see that sNMF allows for the detection of divergent allele frequencies between subpopulations that would otherwise be difficult to define. Genomic scans for candidate SNPs across hierarchical levels of population structure can reveal highly detailed and focused associations between allele frequencies and unique, biologically meaningful components of population structure.
With this method, we identify SNPs with strongly divergent allele frequencies between higher- and lower-order population structure. Further investigation into the functional consequences of these SNPs is encouraged.

5.6 Conclusions

Complex population structure is important to take into account when attempting to detect phenotypic and genotypic signals of local adaptation. Though some delineations between diverged populations may be apparent, cryptic and significant population structure can be uncovered using a number of approaches. We show that $P_{ST} - F_{ST}$ finds many more significantly divergent phenotypes than $Q_{PC}$, probably due to an elevated false discovery rate due to conflation of phenotypic divergence with quantitative trait divergence. Also, the detection of genotypes involved in local adaptation between diverged populations can be assisted by better modeling those populations into a hierarchical framework.

5.7 Acknowledgements

This research is made possible by a grant from the National Science Foundation (The Genetics of Highland Adaptation in Maize, Award Number 1546719).

5.8 References


Figure 5.1 Optimal $k$ values of population structure.  (a) sNMF cross-entropy for $k = 1 - 15$. $k = 3$ is optimal, $k = 2 - 5$ are considered in further comparisons.  (b) Bayesian Information Criterion (BIC) for $k = 1 - 40$. $k = 3$ is optimal.
Figure 5.2  Population genetic structure across North and South America, given \( k = 2 \) subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of \( p \)-values of population differentiation (\( F_{ST} \)) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate = 0.001) are in blue.
Figure 5.3  Population genetic structure across North and South America, given $k = 3$ subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of $p$-values of population differentiation ($F_{ST}$) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate = 0.001) are in blue.
Figure 5.4  Population genetic structure across North and South America, given $k = 4$ subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of $p$-values of population differentiation ($F_{ST}$) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate $= 0.001$) are in blue.
Figure 5.5  Population genetic structure across North and South America, given $k = 5$ subpopulations. (a) Barplot representing individual ancestral coefficients. (b) Ancestry coefficients interpolated over geography. (c) Histogram and Manhattan plot of $p$-values of population differentiation ($F_{ST}$) of all SNPs. Significantly differentiated SNPs (Benjamini Hochberg false discovery rate = 0.001) are in blue.
Figure 5.6  Isolation by distance. (a) Distance between all individuals. (b) Distance between all continent/elevation/latitude subpopulations.
Figure 5.7  Relationship between eight continent/elevation/latitude subpopulations.  (a) Neighbor-joining tree.  (b) Popgraph.  (c) Heatmap of isolation by graph distance.
Figure 5.8  Principal Components Analysis demonstrating genetic population structure within Mexican and South American maize landrace accessions. (a) 4 populations. (b) 8 subpopulations. (c) sNMF $k = 2$. (d) sNMF $k = 3$. (e) sNMF $k = 4$. (f) sNMF $k = 5$. 
Figure 5.9  DAPC analysis and group membership prediction. Scatter plots (left) represent relationships between clusters across discriminant functions. Circles represent training data, squares represent testing data. Contingency tables (right) show assignment of individuals into groups. Columns represent the true group of the individual, and rows represent group predicted by the discriminant functions. (a), (b) Population clusters determined by DAPC at optimal $k = 3$. (c), (d) 4 continent/elevation subpopulations. (e), (f) 8 continent/elevation/latitude subpopulations.
Figure 5.10 $P_{ST}$-$F_{ST}$ comparison between highland and lowland Mexican populations (blue), between highland and lowland South American populations (green), and between all four populations (black).
Figure 5.11  Ear height divergence across PCs of relatedness. Each point represents the trait value of a genotyped individual, and color represents its membership in a population. Solid purple lines show the linear regression of trait values across PCs. Dashed blue lines represent 95% confidence intervals (used only for visualization purposes). Significant phenotypic divergence is found when the linear regression is greater or lower than the confidence interval. (a), (b), (c) Ear height of plants from both common gardens plotted against PC2, PC3, and PC6. (d), (e), (f) Ear height of plants the lowland garden (PV) plotted against PC2, PC3, and PC6. (g), (h), (i) Ear height of plants the highland garden (MT) plotted against PC2, PC3, and PC6.
Figure 5.12  Q-values of significantly higher ((a), (b), (c)) or lower ((d), (e), (f)) phenotypic divergence than expected given axes of population structure captured in the top 11 PCs. Comparisons are taken with the lowland garden (PV, (a), (d)), the highland garden (MT (b), (e)), or from both gardens ((c), (f)). Black = 0-0.001, Red = 0.001-0.01, Orange = 0.01-0.05, Yellow = 0.05-0.1, White = 0.1-1.
Figure 5.13 Population structure across PC1 (a) and PC2 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density ($k$ groups labeled MexHigh, Low, and SAHigh) and population groups (MexHigh, MexLow, SAHigh, and SALow).
Figure 5.14  Population structure across PC3 (a) and PC4 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density ($k$ groups labeled MexHigh, Low, and SAHigh) and population groups (MexHigh, MexLow, SAHigh, and SALow).
Figure 5.15  Population structure across PC5 (a) and PC6 (b) of the genetic similarity matrix. In each subfigure, the top panel demonstrates sNMF group density ($k$ groups labeled MexHigh, Low, and SAHigh) and population groups (MexHigh, MexLow, SAHigh, and SALow).
Table 5.1  Euclidean allele frequency distance between continent/elevation populations

<table>
<thead>
<tr>
<th></th>
<th>MexHigh</th>
<th>MexLow</th>
<th>SAHigh</th>
<th>SALow</th>
</tr>
</thead>
<tbody>
<tr>
<td>MexHigh</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MexLow</td>
<td>11.73</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAHigh</td>
<td>19.53</td>
<td>18.10</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>SALow</td>
<td>16.60</td>
<td>12.42</td>
<td>11.85</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5.2  Euclidean allele frequency distance between continent/elevation/latitude subpopulations

<table>
<thead>
<tr>
<th></th>
<th>MexHighNorth</th>
<th>MexHighSouth</th>
<th>MexLowNorth</th>
<th>MexLowSouth</th>
<th>SAHighNorth</th>
<th>SAHighSouth</th>
<th>SALowNorth</th>
<th>SALowSouth</th>
</tr>
</thead>
<tbody>
<tr>
<td>MexHighNorth</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MexHighSouth</td>
<td>10.48</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MexLowNorth</td>
<td>13.65</td>
<td>14.15</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MexLowSouth</td>
<td>13.67</td>
<td>13.78</td>
<td>10.20</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAHighNorth</td>
<td>20.47</td>
<td>20.21</td>
<td>18.79</td>
<td>19.09</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAHighSouth</td>
<td>21.05</td>
<td>20.83</td>
<td>19.48</td>
<td>19.78</td>
<td>8.66</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SALowNorth</td>
<td>17.40</td>
<td>17.30</td>
<td>13.06</td>
<td>12.90</td>
<td>14.45</td>
<td>15.08</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>SALowSouth</td>
<td>18.82</td>
<td>18.70</td>
<td>15.41</td>
<td>15.50</td>
<td>12.20</td>
<td>12.30</td>
<td>9.87</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 5.3 Relationships between the first six vectors of the eigen decomposition of the inverted genetic distance matrix and geospatial variables: elevation, latitude, longitude, annual mean precipitation, and annual mean temperature. Absolute values of slopes indicate how much variance is explained by the geospatial variable. Bold values represent the highest in the column.

<table>
<thead>
<tr>
<th>PC</th>
<th>Alt</th>
<th>m Alt</th>
<th>ρ Alt</th>
<th>R² Alt</th>
<th>ρ Lat</th>
<th>m Lat</th>
<th>R² Lat</th>
<th>ρ Long</th>
<th>m Long</th>
<th>R² Long</th>
<th>ρ Precip</th>
<th>m Precip</th>
<th>R² Precip</th>
<th>ρ Temp</th>
<th>m Temp</th>
<th>R² Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>-0.00000</td>
<td>0.00493</td>
<td>-0.07019</td>
<td>0.00071</td>
<td>0.4950</td>
<td>0.70361</td>
<td>-0.00078</td>
<td>0.45300</td>
<td>-0.67328</td>
<td>0.00000</td>
<td>0.00037</td>
<td>0.01917</td>
<td>0.0002</td>
<td>0.0020</td>
<td>0.0446</td>
<td></td>
</tr>
<tr>
<td>PC2</td>
<td>0.00004</td>
<td>0.13000</td>
<td>-0.36080</td>
<td>0.00287</td>
<td>0.5090</td>
<td>0.71360</td>
<td>-0.00305</td>
<td>0.43900</td>
<td>-0.66268</td>
<td>0.00002</td>
<td>0.04660</td>
<td>0.21589</td>
<td>0.0047</td>
<td>0.1160</td>
<td>0.3404</td>
<td></td>
</tr>
<tr>
<td>PC3</td>
<td>0.00004</td>
<td>0.35400</td>
<td>0.59489</td>
<td>0.00145</td>
<td>0.1270</td>
<td>0.35668</td>
<td>-0.00193</td>
<td>0.17100</td>
<td>-0.41409</td>
<td>-0.00004</td>
<td>0.20300</td>
<td>-0.45015</td>
<td>-0.00089</td>
<td>0.4090</td>
<td>-0.6398</td>
<td></td>
</tr>
<tr>
<td>PC4</td>
<td>0.00000</td>
<td>0.00571</td>
<td>0.07559</td>
<td>0.00107</td>
<td>0.0686</td>
<td>0.26201</td>
<td>-0.00159</td>
<td>0.11500</td>
<td>-0.33930</td>
<td>-0.00003</td>
<td>0.07050</td>
<td>-0.26435</td>
<td>0.00701</td>
<td>0.00105</td>
<td>0.0984</td>
<td></td>
</tr>
<tr>
<td>PC5</td>
<td>0.00000</td>
<td>0.09550</td>
<td>-0.30903</td>
<td>0.00072</td>
<td>0.0313</td>
<td>0.17700</td>
<td>-0.00073</td>
<td>0.02450</td>
<td>-0.15655</td>
<td>0.00000</td>
<td>0.00015</td>
<td>-0.01233</td>
<td>0.0038</td>
<td>0.0731</td>
<td>0.2703</td>
<td></td>
</tr>
<tr>
<td>PC6</td>
<td>0.00001</td>
<td>0.01120</td>
<td>0.10595</td>
<td>-0.00103</td>
<td>0.0628</td>
<td>-0.25067</td>
<td>0.00106</td>
<td>0.05120</td>
<td>0.22632</td>
<td>0.00001</td>
<td>0.01570</td>
<td>0.12527</td>
<td>-0.0004</td>
<td>0.0007</td>
<td>-0.0267</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 6. GENERAL CONCLUSION

6.1 Dissertation Objectives

I have addressed my dissertation objectives in the following ways:

1. Characterize the strength and breadth of New World maize landrace adaptations to local abiotic environments.

The reciprocal transplant experiment shows clearly that New World maize landraces are locally adapted, both to elevation and other factors specific to either Mexico or South America. Landraces have higher fitness when grown at their native elevation, and out-compete landraces from other elevations. Temperature, precipitation, soil type, and other variables are important components of abiotic environmental heterogeneity to which maize landraces show strong and distinct correlations. Due to the greater challenges of highland conditions (lower temperature, shorter growing seasons, lower air pressure, and greater solar radiation), highland adaptation requires adaptive trade-offs, which result in highland-adapted landraces exhibiting narrower niches. Maize adaptation to highlands occurred separately in North and South America, yet certain phenotypic convergences are evident, and some genes appeared to be important in both cases.

This research was limited to landraces from Mexico and South America. We speculate that patterns here found may be generalized to landraces from the rest of Mesoamerica and the United States, and perhaps to landraces on other continents, in so far as some characteristics of highland conditions are universal.
2. Demonstrate the adaptive nature of the putatively highland adaptive traits anthocyanin pigmentation and leaf sheath pilosity.

The expression of leaf sheath anthocyanin pigmentation increases at higher elevation. Regardless of garden site, highland elevations are more strongly pigmented than lowland populations, demonstrating strong effects of genotype and environment but very little genotype-by-environment interaction. Macrohair density is much less plastic, but Mexican highland maize exhibits much higher pilosity than the other populations.

In all tested maize landrace populations, anthocyanin pigmentation is more positively correlated with fitness at high elevations than at low elevations. Macrohair density, however, is more correlated with fitness in the highland only with Mexican landraces. Furthermore, distinct developmental patterns of both traits have different fitness associations. Solid- or uniform-pattern anthocyanin pigmentation is more associated with highland fitness than is spotty pigmentation, and leaf sheath macrohair density is more associated with highland fitness (among Mexican landraces) than is sheath margin macrohair density.

Note that these correlations do not constitute a controlled experiment in which highland adaptation traits were manipulated to determine the fitness consequences. Past literature has focused on establishing the mechanisms by which these traits mediate the relationship between the plant and the environment. Rather, this research builds on this foundation and establishes that these traits are relevant for local adaptation across maize broadly, and to such a degree that agronomic output could reasonably be improved by breeding these traits into lines that are grown in cooler, dryer, and/or high elevation environments.

3. Map the relationships between environmental variable, adaptive phenotype, and genetic architecture.

Though altitude is clearly the gradient across which a great deal of adaptation occurs, numerous environmental variables which co-vary with altitude contribute to this pattern. Highland landrace distributions are largely defined by associations with high altitude and high temperature season-
ality, whereas lowland landraces are associated with a larger number of factors, including annual precipitation, mean winter temperatures, temperature seasonality, and annual wind speed.

Numerous phenotypes distinguish highland and lowland landrace populations, notably traits related to plant size (plant height, ear height, tassel length, tassel branch number), plant phenology (days to anthesis, days to silking), and highland adaptation traits (anthocyanin pigmentation and macrohair density). Smaller plants and faster maturation are beneficial traits in highland conditions, where high seasonality and cooler temperatures shorten the growing season. Anthocyanin and macrohairs help the plant cope with cooler temperatures and dry conditions, respectively.

Population genetic structure on the geographic scale reflects division due to elevation, latitude, and distance, and perhaps historical demography. Landraces retain agronomic distinctions as well as local adaptations, despite high gene flow and seed turnover. Neutral genetic divergence between populations is lower than phenotypic divergence for most traits, a pattern characteristic of divergent selection. However, several SNPs show very strong correlations with key environmental variables that are shown to constrain landrace geographic distribution. Though many of these SNPs are uncharacterized, others are located within gene bodies of genes with functions in temperature, precipitation, and nutrient stress response.

4. **Determine the degree to which maize landrace names capture adaptive genetic variation.**

Though local adaptation necessitates a genetic foundation, landrace name can be a fairly accurate heuristic of adaptation group. Landrace name predicts the environment at which a plant grows with higher fidelity than does categorical genetic group produced by Bayesian clustering algorithms, though additive genetic variance estimated by a kinship matrix outperforms landrace name. Landrace ENMs have narrower niches than genetic groups when accounting for significant differences due to elevation and spatial disaggregation. For these reasons, for the purpose of selecting units of demarcation with which to capture adaptive genetic diversity within discrete categorical groups, landrace name is as suitable as those constructed explicitly from population genetic structure, if not
more so. The implication of this result is that landrace conservation efforts may be justified in the continued to use landrace name as a criterion in seed conservation and preservation prioritization calculations. However, a more nuanced estimation of population genetic structure that does not necessitate drawing fixed lines around dynamic systems would be preferable, and is a direction that is supported by results here.