Plasticity and genetic adaptation as contributors to the evolutionary history of cultivated maize and its wild relatives

Anne Lorant

université PARIS-SACLAY

NNT : 2018SACLS087

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Plasticity and genetic adaptation as contributors to the evolutionary history of cultivated maize and its wild relatives

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud
GQE-Le Moulon

École doctorale n°567 sciences du végétal :
du gène à l'écosystème (SDV)

Spécialité de doctorat : Biologie

Thèse présentée et soutenue à Gif-sur-Yvette, le 28/03/2018, par

Anne Lorant

Composition du Jury :

Alain Charcosset DR INRA, GQE-Le Moulon, Gif-sur-Yvette	Président
Violaine Llaurens CR CNRS, Muséum National d'Histoire Naturelle, Paris	Rapporteure
Yves Vigouroux DR IRD, IRD, Montpellier	Rapporteur
Patricia Gibert DR CNRS, Université Claude Bernard Lyon 1, Lyon	Examinatrice
Antoine Branca MdC, Université Paris-Sud, Paris	Examinateur
Maud Tenaillon DR CNRS, GQE-Le Moulon, Gif-sur-Yvette	Directrice de thèse
Jeffrey Ross-Ibarra Professeur, UC Davis, California	Co-Directeur de thèse, invité

ACKNOWLEDGEMENTS

I thank Maud Tenaillon and Jeffrey Ross-Ibarra for their help, support and great mentoring throughout my PhD. A particular shootout to Jeff for giving me the opportunity to move to the US even though my English was broken and for not teaching me how to hold a pipette. Another important one goes to Maud for keeping a positive attitude even when I forced her to read my first drafts ☺. Without you, I would have never been able to write this thesis.

I would like to thank Yves Vigouroux and Violaine Llaurens for accepting to act as reviewers of my thesis. Big thanks also to Patricia Gibert, Antoine Branca and Alain Charcosset for accepting to be my examiners.

I would like to thank Dolores Piperno and Matthew Hufford to allow me to be part of their projects, to give me great feedbacks and comments.

I would like to thank the RILAB. To Tim Beissinger, I did not build a canoe and I did not get 2 PhDs… I did what I could… I hope you are not too disappointed! Thanks Marcus Stetter but please be quiet, the office is not the same since you joined. Thanks to Luis Avila and his unicorn who knows how to throw a party. Thanks to Randi Famula for helping me manage all this madness. Thanks to Dianne Velasco to always looking out for me. Thanks to Paul Bilinski, I miss you men, it is not the same here without you. Thanks to Anna O'Brien for your smiles and your happiness. Thanks to Kate Crosby for your numerous advices about work and life.

Thanks to the whole team at the Moulon and especially to Zeineb and Natalia for welcoming me during my time in Paris. I really appreciated spending time with you girls.

I would like to thank my family. First, my sister Isabelle Lorant and my brother Marc Lorant who are always there for me. My parents who have always been behind me in all my

enterprises, even when I told them that I was going to leave to the other side of the world. I am where I am because of you and I am grateful for it.

Thanks to Mirko, you survived to the worse of myself, well I guess is your turn now to finish your PhD.

No Michelle, I did not forget you, but I kept the best for the end. Without your help and comforts, I would never be where I am. Also thank you so much for teaching me the rope of sociability.

Table of Contents

Synthèse en français

Le taxa *Zea* inclus des espèces sauvages et cultivées, offrant une occasion unique de disséquer les déterminants de l'adaptation à la fois à court (depuis la domestication) et à long terme (sélection naturelle dans les populations sauvages). De plus, le processus de sélection au cours de la domestication du maïs a été largement étudié. Le maïs a été domestiqué à partir de *Zea mays* subsp. *parviglumis* au cours d'un seul événement, il y a environ 9 000 ans, dans la vallée de la rivière Balsas au Mexique (Piperno et Flannery 2001, Matsuoka et al., 2002, van Heerwaarden et al., 2011). La domestication du maïs a entraîné un goulot d'étranglement initial (réduction de la taille de la population) dû à la sélection artificielle humaine suivie d'une récupération (Eyre-Walker et al., 1998, Tenaillon et al., 2004, Wright et al., 2005; Beissinger et al., 2016). Ces événements ont légèrement diminué la diversité génétique du maïs par rapport à *parviglumis* (~ 20%) (Matsuoka et al., 2002; Hufford, Xu et al., 2012). Cependant, son histoire démographique est probablement plus complexe qu'un seul goulot d'étranglement suivi d'une expansion. En effet, le maïs s'est rapidement développé depuis le centre de domestication à travers les Amériques, mais la croissance démographique s'est ralentie lorsqu'il s'est adapté aux latitudes et aux altitudes plus élevées (Da Fonseca et al 2015, Swarts et al., 2017). De plus, le flux génétique de *Zea mays* subsp. *mexicana* dans le maïs des hautes élévations, qui a contribué à son adaptation aux hautes terres, complique davantage ces patternes démographiques (van Heerwaarden et al., 2011, Hufford et al., 2013).

Dans le premier chapitre, nous avons mis en évidence l'effet de la démographie durant la période suivant la domestication et l'introgression des téosintes dans des variétés locales de maïs couvrant une grande partie de sa distribution précolombienne. Nous avons constaté que les changements

historiques dans la taille des populations associés à la domestication, ainsi que l'expansion du maïs dans les Amériques, ont augmenté le nombre d'allèles délétères (c'est-à-dire sa charge génétique). Au cours de ces goulots d'étranglement, le maïs a connu une dépression de consanguinité due à une réduction significative de la taille de la population. Une telle réduction est connue pour réduire l'efficacité de sélection à l'échelle du génome, permettant ainsi d'augmenter la fréquence des variant légèrement délétères. Cette charge génétique était particulièrement prononcée chez le maïs andin, qui a subi un effet fondateur plus fort. Il est intéressant de noter que le flux génétique du *mexicana* dans les maïs des hautes terres a contribué de manière significative à la restauration de la diversité génétique. Notamment, cette introgression n'était pas possible dans le maïs andin, car la distribution de *mexicana* est limitée au Mexique. Ensemble, ces observations renforcent l'importance du flux génétique et de la démographie en tant qu'acteurs clés de l'adaptation durant la post-domestication du maïs.

Dans le deuxième chapitre, nous avons étudié un aspect sous-estimé de l'évolution des caractères nouveaux au cours de la domestication, qui est la plasticité. La plasticité est définie comme la capacité d'un génotype à exprimer différents phénotypes en réponse à des conditions environnementales variées. Les populations avec des génotypes plastiques générant des phénotypes flexibles devraient mieux faire face aux changements environnementaux, coloniser des niches plus larges et présenter un plus grand potentiel d'expansion (Wennersten et Forsman 2012). La flexibilité phénotypique permise par la plasticité est un processus particulièrement important pour les plantes, car elles sont fixées dans un endroit spécifique et ont des capacités limitées à se protéger de leur environnement (Des Marais, Hernandez et Juenger 2013). Dans ce contexte, nous avons étudié la plasticité de l'expression génique du maïs et de *parviglumis* entre le début de l'Holocène et les conditions actuelles. Par rapport aux niveaux actuels, le plus faible taux

de CO_2 et les températures plus froides induit une réponse plastique dans de multiples phénotypes de *parviglumis*, certains typiques du maïs (par exemple des changements dans la sexualité de l'inflorescence et l'architecture végétative). Il est intéressant de noter que ces caractères ont été canalisés dans le maïs, aucune variation de la ramification ou de la sexualité des fleurs n'ayant été observée dans les mêmes conditions. En plus de ces changements de morphologie, nous avons observé des changements substantiels dans les réseaux de co-expression et plus de 2000 gènes qui présentaient une expression différentielle uniquement dans la parviglumis. Ces résultats suggèrent que pendant la domestication du maïs, il a perdu une plasticité substantielle.

Dans le dernier chapitre, nous avons identifié des signatures génomiques d'adaptation locale dans six populations naturelles de la sous-espèce *parviglumis*. *Parviglumis* est une graminée sauvage annuelle originaire du Mexique qui a subi une adaptation locale importante (Aguirre-Liguori et al 2017, Fustier et al 2017, Pyhäjärvi et al., 2013). Nous avons spécifiquement recherché des traces de sélection directionnelle récente, telles que des « hard and soft sweeps », ainsi que les histoires démographiques pour ces populations de téosinte. Nous avons observé que l'expansion post-glaciaire des téosintes a engendré une réduction différentielle de Ne parmi les populations, bien que tous aient étendu leurs aires de répartition. Le fait que les populations ne partagent que quelques sweeps confirme une forte adaptation locale chez *parviglumis*. Notre observation de l'expansion des populations est en accord avec les analyses de teneur en pollen des carottes de sédiments du Mexique qui suggèrent que les graminées se sont développées au cours des 10 000 dernières années en raison des conditions environnementales changeantes pendant la période du holocène et la gestion humaine de la végétation (Piperno et al., 2007 , Correa-Metrio et al., 2012). De plus, la variation absolue et relative du nombre de sweeps sélectifs dur et doux au sein de ces

populations indique une interaction étroite entre la démographie et la sélection, ce qui a une incidence importante sur le potentiel d'adaptation des populations individuelles et des espèces.

En résumé, 1) l'adaptation génétique à une petite échelle géographique ; 2) des introgressions répétées entre les formes sauvages et domestiques; et 3) l'assimilation génétique qui "cimente" les phénotypes domestiqués initialement induits par les réponses plastiques à l'environnement, sont tous des mécanismes impliqués dans l'adaptation de *Zea mays*. La caractérisation complète des relations entre la démographie, le flux génétique et la sélection fournira de nouvelles pistes pour comprendre comment l'histoire a influencé la trajectoire évolutive actuelle d'une espèce. En outre, en raison de ses liens étroits avec le maïs, les connaissances récoltées sur le téosinte peuvent être traduites au maïs pour améliorer sa culture. L'introgression des téosintes dans le maïs n'est pas rare et a aidé le maïs à s'adapter aux hautes altitudes, dans la mesure où l'environnement est plus sec et plus froid comparé au centre de domestication (Takuno et al., 2015). En d'autres termes, les allèles du téosinte pourraient être utilisés pour améliorer le germplasme du maïs pour les processus importants qui ont un impact sur la sélection du maïs, tels que la sécheresse ou la tolérance au froid. L'introgression d'allèles sauvages a été utilisée dans d'autres programmes de sélection, par exemple pour améliorer le rendement des tomates (Tanksley et McCouch, 1997).

Références

Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. “Connecting Genomic Patterns of Local Adaptation and Niche Suitability in Teosintes.” *Molecular Ecology* 26 (16):4226–40. https://doi.org/10.1111/mec.14203.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, and J. Ross-Ibarra. 2016. “Recent Demography Drives Changes in Linked Selection across the Maize Genome.” *Nature Plants* 2 (7):16084. https://doi.org/10.1038/nplants.2016.84.

Correa-Metrio, A., S. Lozano-García, S. Xelhuantzi-López, S. Sosa-Nájera, and S. E. Metcalfe. 2012. “Vegetation in Western Central Mexico during the Last 50 000 Years: Modern Analogs and Climate in the Zacapu Basin.” *Journal of Quaternary Science* 27 (5):509–18. https://doi.org/10.1002/jqs.2540.

Eyre-Walker, a, R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. “Investigation of the Bottleneck Leading to the Domestication of Maize.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (April):4441–46. https://doi.org/10.1073/pnas.95.8.4441.

Fonseca, R. R. Da, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. Alfredo Samaniego, et al. 2015. “The Origin and Evolution of Maize in the Southwestern United States.” *Nature Plants* 1. https://doi.org/10.1038/nplants.2014.3.

Fustier, M. A., J. T. Brandenburg, S. Boitard, J. Lapeyronnie, L. E. Eguiarte, Y. Vigouroux, D. Manicacci, and M. I. Tenaillon. 2017. “Signatures of Local Adaptation in Lowland and Highland Teosintes from Whole-Genome Sequencing of Pooled Samples.” *Molecular Ecology* 26 (10):2738–56. https://doi.org/10.1111/mec.14082.

Heerwaarden, J. van, J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. de Jesus Sanchez Gonzalez, and J. Ross-Ibarra. 2011. “Genetic Signals of Origin, Spread, and Introgression in a Large Sample of Maize Landraces.” *Proceedings of the National Academy of Sciences* 108 (3):1088–92. https://doi.org/10.1073/pnas.1013011108.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. “The Genomic Signature of Crop-Wild Introgression in Maize.” *PLoS Genetics* 9 (5). https://doi.org/10.1371/journal.pgen.1003477.

Hufford, M. B, X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J-M. Chia, R. A. Cartwright, R. J Elshire, et al. 2012. “Comparative Population Genomics of Maize Domestication and Improvement.” *Nat Genet* 44 (7):808–11. https://doi.org/10.1038/ng.2309.

Marais, D. L Des, K. M Hernandez, and T. E Juenger. 2013. “Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment.” *Annual Review of Ecology, Evolution, and Systematics* 44. Annual Reviews:5–29.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. S. Buckler, and J. Doebley. 2002. “A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping.” *Proc Natl Acad Sci U S A* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

Piperno, D. R., and K. V. Flannery. 2001. “The Earliest Archaeological Maize (Zea Mays L.) from Highland Mexico: New Accelerator Mass Spectrometry Dates and Their Implications.” *Proc Natl Acad Sci U S A* 98 (4):2101–3. https://doi.org/10.1073/pnas.98.4.2101.

Piperno, D. R., J. E. Moreno, J. Iriarte, I. Holst, M. Lachniet, J. G. Jones, A. J. Ranere, and R. Castanzo. 2007. "Late Pleistocene and Holocene Environmental History of the Iguala Valley, Central Balsas Watershed of Mexico." *Proceedings of the National Academy of Sciences* 104 (29). National Acad Sciences:11874–81.

Pyhäjärvi, T., M. B Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biol Evol* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski, J. Holland, M. Kruse-Peeples, et al. 2017. "Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America." *Science* 357 (6350):512–15. https://doi.org/10.1126/science.aam9425.

Takuno, S., P. Ralph, K. Swart, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. 2015. "Independent Molecular Basis of Convergent Highland Adaptation in Maize." *Genetics* 200 (4):1297–1312. https://doi.org/10.1534/genetics.115.178327.

Tanksley, S. D., and S. R. McCouch. 1997. "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild." *Science* 277 (5329):1063–66. https://doi.org/10.1126/science.277.5329.1063.

Tenaillon, M. I., J. U'Ren, O. Tenaillon, and B. S. Gaut. 2004. "Selection versus Demography: A Multilocus Investigation of the Domestication Process in Maize." *Molecular Biology and Evolution* 21 (7):1214–25. https://doi.org/10.1093/molbev/msh102.

Vigouroux, Y., J. C. Glaubitz, Yoshihiro Matsuoka, Major M. Goodman, Jesús Sánchez G., and John Doebley. 2008. "Population Structure and Genetic Diversity of New World Maize Races Assessed by DNA Microsatellites." *American Journal of Botany* 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097.

Wennersten, L., and A. Forsman. 2012. "Population-Level Consequences of Polymorphism, Plasticity and Randomized Phenotype Switching: A Review of Predictions." *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2012.00231.x.

West-Eberhard, M. J.. 2003. *Developmental Plasticity and Evolution*. Oxford University Press.

Wright, S.I, I. V. Bi, S. G Schroeder, M. Yamasaki, J.ohn F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. "The Effects of Artificial Selection on the Maize Genome." *Science (New York, N.Y.)* 308 (5726):1310–14. https://doi.org/10.1126/science.1107891.

Yang, J., S. Mezmouk, A. Baumgarten, E.S. Buckler, K. E. Guill, M.l D. McMullen, R. H. Mumm, and J. Ross-Ibarra. 2017. "Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize." *PLoS Genetics* 13 (9). https://doi.org/10.1371/journal.pgen.1007019.

Zhou, Y., M. Massonnet, J. S. Sanjak, D. Cantu, and B. S. Gaut. 2017. “Evolutionary Genomics of Grape (Vitis Vinifera Ssp. Vinifera) Domestication.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (44):11715–20. https://doi.org/10.1073/pnas.1709257114.

I. OBJECTIVES

The overarching goal of my PhD thesis was to study the genetic bases of adaptation in the *Zea* taxa. This taxon is a particularly good model to study the determinants of adaptation as it features both wild populations and a domesticated relative. This allows us to apply state-of-the-art population genomic tools to dissect adaptation within both short- (since domestication) and long-time scales (natural selection in wild populations). Moreover, because of the central role of maize as a food, fuel and fiber source, it is an extensively studied plant and features a complete and well annotated reference genome as well as extended phenotypic information measured across numerous panels.

Maize is known to be highly adaptable and is therefore an appropriate model to study plant adaptation. Indeed, the geographic range of maize has rapidly exceeded the range of teosintes, with documented routes of diffusion northward and southward of Mexico (Da Fonseca et al. 2015; Vigouroux et al. 2008) as well as to the European continent, and adapted to various environmental conditions (Brandenburg et al. 2017). To the current days, maize have a wider distribution than any other important crop (Hake and Ross-Ibarra 2015). As for teosintes, during the last ten years, *Zea mays* ssp. *parviglumis* and ssp. *mexicana* have gradually emerged as excellent models for dissecting long-term adaptation to natural selection (Hufford, Bilinski, et al. 2012). Compared to maize, their range distribution is limited geographically to Mexico, yet they span numerous environmental conditions with varying temperatures, precipitation levels and elevations. This characteristic, combined with a reduced migration rate, have enhanced the effect of local adaptation (Pyhäjärvi et al. 2013).

Although human-driven selection in maize has been extensively studied, with famous reported examples of beneficial mutations and sweeps (Doebley, Stec, and Gustus 1995) that illustrated the genetic basis of short term adaptation, the genetic bases of natural variation are still poorly understood. This is mainly due to traits driving local adaptation being mostly quantitative (Savolainen, Lascoux, and Merilä 2013). This complex determinism may involve numerous, but not necessarily substantial, allele frequency changes.

Genetic adaptation can proceed rapidly by direct phenotypic adjustments without genetic alterations, a mechanism called phenotypic plasticity (Moczek et al. 2011). Alternatively, the fitness of the individuals can increase over time through selection changing the frequency of beneficial alleles underlying selected traits. From an applied perspective, understanding the genetic bases of local adaptation and the contribution of plasticity will provide us with new tools to both better understand and mitigate the effect of climate changes as well as the fate of populations impacted by those. Indeed, the ongoing global warming is predicted to be accompanied with a significant decline in rainfall and shifts of pests and diseases which will severely impact current yields (Lobell, Schlenker, and Costa-Roberts 2011). Artificial selection during modern crop breeding increased drastically yields and produced hybrids optimized to their growing environments. However, recent studies suggest these hybrids lack in plastic alleles (Gage et al. 2017). Crop uniformity, across sometimes entire continents, makes cultivars highly susceptible to biotic and abiotic stresses. Although average changes to growing conditions, due to global warming, are unlikely to affect these cultivars, extreme weather conditions can impact yield considerably. Under such scenario, the plasticity of a plant might mitigate these effects to some extent. Regardless, we have to move toward crops that are adapted at a more local scale. There is,

therefore, a pressing need to better understand the dynamics and genomic basis of adaptation which has the potential to help restore diversity in cultivated species as well as improve current yields.

The introduction of my thesis is in the form of a review that will constitute a chapter to be published in 2018 in a book entitled "Statistical Population Genomics". In this review, we reported empirical evidences of short- and long-term adaptation in maize and teosintes, discussed the role of phenotypic plasticity, local adaptation and convergence in adaptation. We then discussed possible drivers of adaptation by summarizing results from selection experiments in maize. This introduction is followed by three chapters for which goals, major results and my contributions are summarized below.

In the first chapter, we investigated the demographic history of maize using high-coverage re-sequencing data from 31 landraces spanning its pre-Colombian distribution. This included a study of the demographic history of different populations of maize and its effect on their genomes. My contributions were to grow the material, collect the samples, prepare the sequencing libraries and supervise the sequencing process. However, my contributions were not limited to the generation of the data, but I also provided deep scientific inputs to the demographic analyses. This chapter is presented under the form of a peer-reviewed published article in Genome Biology.

In the second chapter, I described our results on the effect of plastic phenotypic changes on gene expression and their impact during maize domestication using RNA sequencing. For this, we compared maize and teosinte transcriptomes from plants grown in climatic conditions simulating environmental conditions found during the Holocene period, i.e. at low temperature and low CO_2 level. Our experiment was based on observations made by Dolores Piperno (Piperno et al. 2015) showing that exposure of teosinte to environments mimicking those found during early domestication resulted in a plastic induction of domesticated phenotypes in teosinte. This

suggested that early agriculturalists may have selected for genetic mechanisms that cemented domestication phenotypes initially induced by a plastic response of teosinte to environment, a process known as genetic assimilation. My contribution was to prepare the libraries, supervise the sequencing process and analyze the data. This chapter is presented under the form of a peer-reviewed published article in Plos One.

In the third chapter, I investigated the genomic signatures of local adaptation using 6 natural populations of *parviglumis* (60 individuals) sequenced at high depth (20-25x). This is the first study of local adaptation in teosintes using high coverage sequencing of single individuals. This chapter is presented in the form of preliminary results investigating the relative contribution of hard and soft sweeps during local adaptation. My contribution here was to prepare the libraries, supervise the sequencing process and analyze the data.

Finally, I discuss the main results of the thesis, the methods used and future perspectives.

References

Brandenburg, J. T., T. Mary-Huard, G. Rigaill, S. J. Hearne, H. Corti, J. Joets, C. Vitte, A. Charcosset, S. D. Nicolas, and M. I. Tenaillon. 2017. "Independent Introductions and Admixtures Have Contributed to Adaptation of European Maize and Its American Counterparts." *PLoS Genetics* 13 (3). https://doi.org/10.1371/journal.pgen.1006666.

Doebley, J., A. Stec, and C. Gustus. 1995. "Teosinte branched1 and the Origin of Maize: Evidence for Epistasis and the Evolution of Dominance." *Genetics* 141 (1):333–46.

Fonseca, R. Da, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. A. Samaniego, et al. 2015. "The Origin and Evolution of Maize in the Southwestern United States." *Nature Plants* 1. https://doi.org/10.1038/nplants.2014.3.

Gage, J. L., D. Jarquin, C. Romay, A. Lorenz, E. S. Buckler, S. Kaeppler, N. Alkhalifah, et al. 2017. "The Effect of Artificial Selection on Phenotypic Plasticity in Maize." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01450-2.

Hake, S., and J. Ross-Ibarra. 2015. "Genetic, Evolutionary and Plant Breeding Insights from the

Domestication of Maize." *eLife*. https://doi.org/10.7554/eLife.05861.

Hufford, M. B., P. Bilinski, T. Pyhäjärvi, and J. Ross-Ibarra. 2012. "Teosinte as a Model System for Population and Ecological Genomics." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.08.004.

Lobell, D. B, W. Schlenker, and J. Costa-Roberts. 2011. "Climate Trends and Global Crop Production since 1980." *Science (New York, N.Y.)* 333 (6042):616–20. https://doi.org/10.1126/science.1204531.

Moczek, A. P., S. Sultan, S. Foster, C. Ledón-Rettig, I. Dworkin, H. F. Nijhout, E. Abouheif, and D. W. Pfennig. 2011. "The Role of Developmental Plasticity in Evolutionary Innovation." *Proc Biol Sci* 278 (1719):2705–13. https://doi.org/10.1098/rspb.2011.0971.

Piperno, D. R., I. Holst, K. Winter, and O. McMillan. 2015. "Teosinte before Domestication: Experimental Study of Growth and Phenotypic Variability in Late Pleistocene and Early Holocene Environments." *Quaternary International* 363. Elsevier:65–77.

Pyhäjärvi, T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biol Evol* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Savolainen, O., M. Lascoux, and J. Merilä. 2013. "Ecological Genomics of Local Adaptation." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3522.

Vigouroux, Y., J. C. Glaubitz, Y. Matsuoka, M. M. Goodman, J. Sánchez G., and J. Doebley. 2008. "Population Structure and Genetic Diversity of New World Maize Races Assessed by DNA Microsatellites." *American Journal of Botany* 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097.

II. GENOMICS OF LONG- AND SHORT- TERM ADAPTATION IN MAIZE AND TEOSINTES

Anne Lorant[1,4], Jeffrey Ross-Ibarra[1,2,3], Maud Tenaillon[4,*]

[1] : Dept. of Plant Sciences, University of California, Davis, CA 95616, USA.

[2] : Center for Population Biology, University of California, Davis, CA 95616, USA.

[3] : Genome Center, University of California, Davis, CA 95616, USA.

[4] : Génétique Quantitative et Evolution – Le Moulon, Institut National de la Recherche agronomique, Université Paris-Sud, Centre National de la Recherche Scientifique, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, 91190, France.

* Corresponding author

E-mail: maud.tenaillon@inra.fr

II.1 Introduction

A combination of archeobotanical records and genetic data has established that maize (*Zea mays* ssp. *mays*) was domesticated around 9 000 years ago in the Balsas river valley of Mexico from the wild teosinte *Zea mays* ssp. *parviglumis* (Piperno and Flannery 2001; Matsuoka et al. 2002; van Heerwaarden et al. 2011). Unlike complex domestication scenarios involving multiple domestication events in the common bean (*Phaseolus vulgaris* L.) and lima bean (*Phaseolus. lunatus* L.) (Kwak et al. 2012) or multiple progenitors from different regions in barley (*Hordeum vulgare*; Poets et al. 2015), maize stands as a relatively simple scenario involving only a single domestication event resulting in a moderate decrease of genetic diversity of roughly 20% (Matsuoka et al. 2002; Hufford, Xu, et al. 2012).

With the rise of coalescent simulation tools since the late 1990's (Hudson 2002), researchers have repeatedly attempted to establish demographic scenarios of maize domestication. All concur with a simple bottleneck model, i.e. a reduction of effective population size (N_e), with <10% of the teosinte population contributing to the maize gene pool (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005; Beissinger et al. 2016). A recent investigation indicates that this bottleneck was followed by a major expansion resulting in an N_e for modern maize much larger than that of teosinte (Beissinger et al. 2016). However, the complexity of the forces acting to shape diversity at a genome-wide scale makes it difficult to disentangle them. On one hand, domestication has likely promoted strong positive selection at ~2% to 4% of loci (Wright et al. 2005) producing one of the most famous text-book example of selective sweeps at *tb1*, a gene responsible for the reduced branching phenotype in maize (Doebley, Stec, and Gustus 1995). On the other hand, purifying selection has also reduced neutral genetic diversity (Beissinger et al.

2016). Such selection may lead to an excess of rare variants, a footprint easily confounded with both positive selection and population expansion (Cvijovic, Good, and Desai 2017).

After its initial domestication, the geographic range of maize has rapidly exceeded that of its wild relatives, with documented routes of diffusion northward and southward of Mexico (Da Fonseca et al. 2015; Vigouroux et al. 2008) and to the European continent (Brandenburg et al. 2017). Today the maize gene pool worldwide consists of locally adapted/selected open-pollinated populations (landraces) as well as modern inbred lines, derived from landraces, that are used in hybrid production for modern breeding. Such spatial movement has exerted a diversity of selective pressures, triggering changes in the phenology of individuals that ultimately determines the completion of the annual cycle and individual fitness (Chuine 2010; Swarts et al. 2017).

Populations respond to environmental pressures in three ways: (1) by migration to environments whose conditions are similar to their original conditions; (2) by genetic adaptation through the recruitment of pre-existing or new alleles that increase the fitness of individuals carrying them; or (3) by immediate phenotypic adjustments without genetic alterations, a mechanism called phenotypic plasticity. While migration is more prevalent in animals than in plants (for instance Hill, Thomas, and Huntley 1999; Thomas and Lennon 1999; Lundy, Montgomery, and Russ 2010; Bebber, Ramotowski, and Gurr 2013), plants are capable of long distance expansion from their native range (Clark et al. 1998; Davis and Shaw 2001; Williams et al. 2004). For instance, recent shifts driven by global warming have been reported in tree species distributed in California, Oregon and Washington. In most of them, seedlings have experienced an average shift in ranges of about 27m in altitude and 11kms northwards, towards colder environments, compared to mature trees (Monleon and Lintz 2015). Likewise, rising temperatures have likely caused the upslope migration reported for vascular plants species across European

boreal-to-temperate mountains (Pauli et al. 2012). The range of annual teosinte, in contrast, appears to be quite similar to what it was at the time of domestication (Hufford, Martínez-Meyer, et al. 2012; Ureta et al. 2012). In fact, relatively minor shifts are predicted to have occurred even over the dramatic changes of the last glacial maximum, suggesting that migration has not been the dominant strategy in response to environmental change in teosintes.

In contrast, pioneering studies on the genetics of human-driven selection in maize have provided novel insights into our understanding of short-term adaptive processes. In the last decade, the annual teosintes *Zea mays* ssp. *parviglumis* and ssp. *mexicana* have also emerged as models for dissecting long-term adaptation to natural selection (Hufford, Bilinski, et al. 2012). While their distribution is rather limited geographically, teosintes span extremely various environmental conditions in terms of temperatures, precipitations and elevations. Migration is also somewhat limited by the complex landscape of Mexico (Pyhäjärvi et al. 2013). Together, these conditions set the stage for extensive local adaptation. The two subspecies of teosinte diverged ~60,000 years ago and they currently occupy distinct ecological niches (Ross-Ibarra, Tenaillon, and Gaut 2009). While *parviglumis* grows at low elevations (<1800 m), *mexicana* has colonized the cooler and drier Central Plateau of Mexico from 1500 to 2800 m of altitude (Fukunaga et al. 2005; Hufford, Martinez-Meyer, et al. 2012). Both teosinte taxa display a high level of nucleotide diversity (Fukunaga et al. 2005; Ross-Ibarra, Tenaillon, and Gaut 2009) consistent with large estimates of effective population sizes from 120k to 160k (Ross-Ibarra, Tenaillon, and Gaut 2009).

Here we review empirical reports of short- and long-term adaptation in maize and teosintes, and discuss the role of phenotypic plasticity.

II.2 How to explore adaptation?

Genetic adaptation can be defined as the modulation of allele frequencies through natural and/or artificial selection. Natural selection is imposed by changes in environmental conditions, and artificial selection by humans, consciously or unconsciously. Identification of adaptive loci (Fig. 1A-B) and/or traits (Fig. 1C-D) uses spatial or temporal variation in conjunction with quantification of traits in native environments (Fig. 1F) or in common gardens (Fig. 1G) (Anderson and Geber 2010; Fournier-Level et al. 2011; Savolainen, Lascoux, and Merilä 2013). While the temporal approach includes retrospective studies that follow the phenotypic and genetic composition of populations through time (for instance Thompson et al. 2013) to infer past selective events (Tiffin and Ross-Ibarra 2014), the spatial approach relies on samples of populations that are geographically separated.

In *Zea*, experimental approaches have been coupled with genotyping of sampled/evolved populations to identify the genomic bases of observed phenotypic changes. More often, however, studies have focused only on species-wide population genomic analyses tracing patterns of variation. These include searches for (1) spatial associations of allele frequencies with environmental factors or phenotypes (Fig. 1A); (2); shifts in allele frequencies across genetic groups (e.g. comparing wild and cultivated samples) using genome scans (Fig. 1B); and (3) differential gene expressions related to population/subspecies differentiation. An increasingly popular approach that was initiated in 2003 by Jaenicke-Despres (Jaenicke-Despres 2003) is the use of ancient DNA, as maize cobs are often well preserved making them an attractive source for ancient DNA studies.

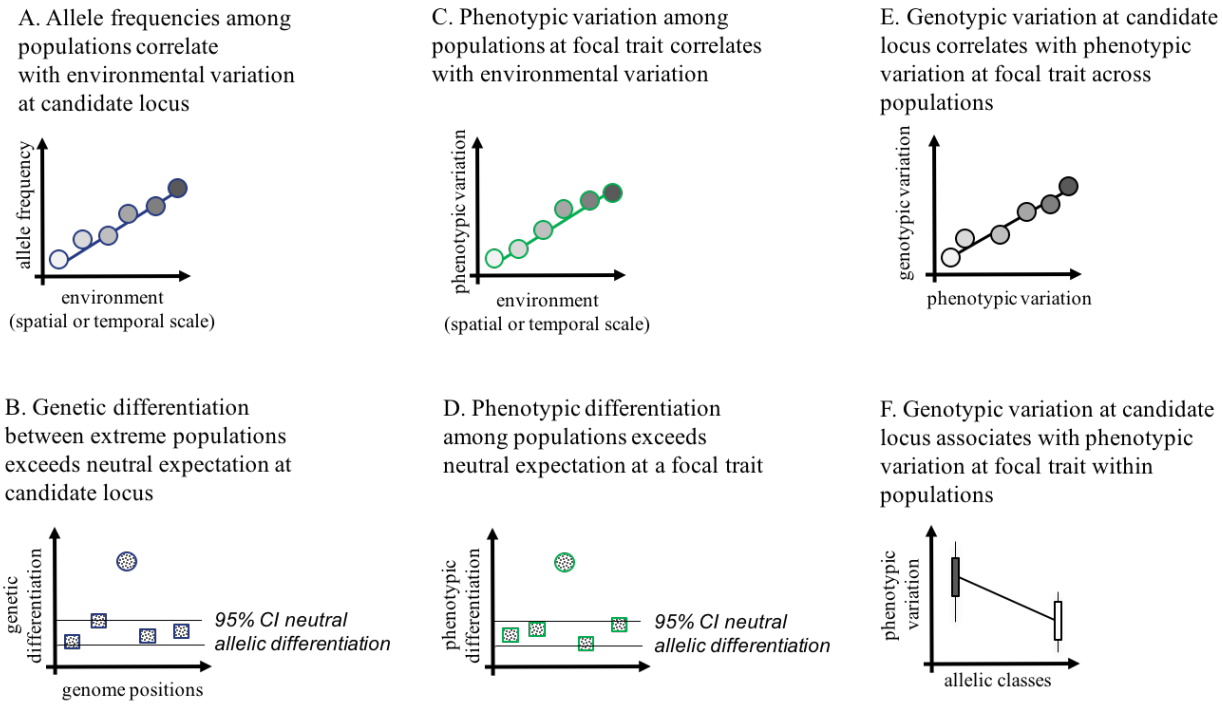


Figure 1: Experimental approaches to detect potentially adaptive polymorphisms and traits using population genetic (A-B) or phenotypic (C-D) data, or combining both (E-F).

A candidate polymorphism whose allele frequency among populations varies with spatial or temporal variation, can be detected using correlation-based methods (A) or genome-wide scans, where it displays an elevated differentiation of allele frequencies compared with neutral (squares) loci (B). A candidate trait that co-varies with spatial or temporal variation among populations can be detected using correlation-based methods (C) or when phenotypic differentiation measured in common environment(s) exceeds genotypic differentiation at neutral (squares) loci (D). A link between candidate loci and traits can be established by correlating genotypic and phenotypic variation measures in common environment(s) across populations (E), and within populations (F).

II.3 Local adaptation in maize and teosintes

Strictly defined, a genotype can be considered locally adapted if it has a higher fitness at its native site than any other non-native genotypes (Kawecki and Ebert 2004). Locally adapted alleles can be either neutral or deleterious in other environments. Two models depict those situations, namely conditional neutrality and antagonistic pleiotropy (Anderson et al. 2013).

Despite numerous studies, the genetic processes underlying local adaptation in natural populations are still poorly understood. Studies showed that highlands maize landraces outperform lowland maize populations in their native environment but perform worse than any other population at lower elevation sites (Mercer, Martínez-Vásquez, and Perales 2008), suggesting strong adaptation for high altitude. Interestingly, an ancient DNA study shows that, by 4000 years ago, maize was already largely cultivated in the lowlands of southwestern United States but the adaptation to the highland of Colorado Plateau took an extra 2000 years. This delay is probably the result of a long time to adapt to local conditions (Swarts et al. 2017).

Natural selection acts on phenotypic traits, changing the frequency of underlying alleles and shifting population phenotypes toward local optima. Since these optima rely on local conditions, genes ecologically important usually differ between sub-populations in heterogeneous environments, which results in divergence in allele frequencies over time. This characteristic has been utilized in genome scans to mine correlations between allele frequencies and environmental variables (Fig. 1A). Such studies have revealed that, in teosintes, loci associated with environmental variables impact flowering time and adaptation to soil composition (Aguirre-Liguori et al. 2017; Fustier et al. 2017; Pyhäjärvi et al. 2013). Flowering time was also a key component of maize's local adaptation to higher latitudes during post-domestication. Maize evolved a reduced sensitivity to photoperiod, in part due to a CACTA-like TE insertion in the promoter region of the *ZmCCT* gene that drives photoperiod response in early flowering maizes (Hung et al. 2012; Yang et al. 2013). An example of adaptation driven by soil interactions is the tolerance of maize and teosinte to aluminum in highly acidic soils. In these lines, the adaptation is linked to tandem duplications of the *MATE1* gene involved in the extrusion of toxic compounds (Maron et al. 2013).

Numerous other biotic and abiotic factors are likely involved in adaptation in maize and teosinte, including predation, parasitism, moisture and herbicide (Linhart and Grant 1996; Valverde 2007). For example, a study on *parviglumis* has shown that in response to herbivory, immunity genes involved in the inhibition of insect' digestive proteases experienced a recent selective sweep in a region of Mexico, probably reflecting their local adaptation (Moeller and Tiffin 2008). These measures are however more difficult to gather making their study less common.

Interestingly, three large inversion polymorphisms seem to play an important role in local adaptation. Among them, a 50Mb inversion on chromosome 1 is found at high frequency in *parviglumis* (20-90%), low frequency in *mexicana* (10%) and is absent in maize. This inversion is highly correlated with altitude and significantly associated with temperature and precipitation (Fang et al. 2012; Pyhäjärvi et al. 2013). Inversions on chromosomes 4 and 9 also displayed environmental association in teosintes (Pyhäjärvi et al. 2013). Local adaptation to different habitats or niches is a gradual process that can promote divergence and, in the long run, ecological speciation (Schluter 2009). Genotyping of a broad sample of 49 populations covering the entire geographic range of teosintes has recently provided some evidence of this. Aguirre-Liguori et al. 2017 showed that both within *parviglumis* and *mexicana*, populations distributed at the edge of the ecological niche, but not the range edge, experience stronger local adaptation. This suggests that local adaptation may have contributed to divergence between these two subspecies.

II.4 What is the role of phenotypic plasticity?

Phenotypic plasticity is defined as the capacity of a genotype to produce a range of expressed phenotypes in distinct environments. This is achieved through differential developmental pathways in response to changing conditions (Beldade, Mateus, and Keller 2011; Gilbert and Epel 2009). Studies have shown that plasticity is an important process for the evolution of novel traits during adaptation. Indeed, populations with flexible phenotypes are predicted to better cope with environmental changes, to colonize broader niches, and to display a greater potential for expansion (Wennersten and Forsman 2012). This process is particularly important for plants as they are fixed in a specific location and not sheltered from the environment (Des Marais, Hernandez, and Juenger 2013b).

When the environment changes, the phenotypic optimum of a population is likely altered as well. As a result, individuals that show a plastic response in the direction of the new optimum, will have a fitness advantage. In contrast, individuals that exhibit no plasticity or that produce phenotypes too far from this optimum, will be selected against.

However, plasticity has some limits and may entail a fitness cost. For instance, compared to developmentally fixed phenotypes, plastic individuals in constant environments may display lower fitness or produce a less adapted phenotype. Possible reasons include sensory mechanisms that have a high energetic cost, the epistatic effects of regulatory genes involved in the plastic response, lag time between the perception and the phenotypic response and genetic correlations among traits (Auld, Agrawal, and Relyea 2010; DeWitt, Sih, and Wilson 1998; Nicotra et al. 2010).

Phenotypic plasticity is difficult to study as it arises from genetic and environmental interactions which are often hard to disentangle. Moreover, phenotypic plasticity is fundamentally

intertwined with genetic adaptation, furthering the difficulty of determining the causality of a phenotype. The difference between genetic adaptation and phenotypic plasticity is that for the former the phenotype is genetically determined, whereas plastic phenotypes plasticity may be heritable, the phenotype is largely determined by the environment (Kawecki and Ebert 2004; van Kleunen and Fischer 2005). In addition, plasticity can be lost, and the phenotype constitutively expressed by the fixation of genetic variation after a number of generations of constant selection, a process called genetic assimilation (Diggle and Miller 2013; Kuzawa and Bragg 2012; Standen, Du, and Larsson 2014). Hence an initially plastic phenotype may become a genetic adaptation after genetic assimilation. Some examples of plastic responses are well documented in plants, for example, the response to vernalization in *Arabidopsis* regulating flowering time in some ecotypes (Nicotra et al. 2010). Another example is the change in seed dormancy in response to the environment which prevents germination when conditions are unlikely to lead to the survival of the plant (Nicotra et al. 2010).

The *Zea* taxa are good models to investigate plasticity as maize is grown worldwide and adapted to a diversity of environments. In addition, studies of teosinte allow comparison to ancestral levels of plasticity. A recent experiment evaluated plasticity in maize by studying variation in Genotype by Environment interactions (GxE) for a number of phenotypes in 858 inbred lines across 21 locations across North America (Gage et al. 2017). Results demonstrated that genes selected for high yield in temperate climates in North America correlated with low variance in GxE. This suggests a loss of plasticity accompanying selection for stable crop performance across environments, a major goal for breeders. In addition, GxE was mainly explained by regulatory regions (Gage et al. 2017), an observation in agreement with previous

findings indicating that most phenotypic variation in maize is due to gene regulation (Wallace et al. 2014).

Recent work on maize and *parviglumis* growing under environmental conditions mimicking those encountered at the time of maize domestication (comparatively lower CO2 atmospheric concentration and lower temperatures) gives better insights into this phenomenon. The results showed that teosintes grown in these conditions exhibit contemporary maize-like phenotypes (Piperno et al. 2015). In contrast, modern maize has lost this plastic response. Over 2000 candidate loci associated with phenotypic changes showed altered expression in teosintes but not in maize, implying that they are no longer environmentally responsive (Figure2; Lorant et al. 2017). Such loss of phenotypic plasticity may limit the ability of maize to cope with environmental variability in the face of current climate changes.

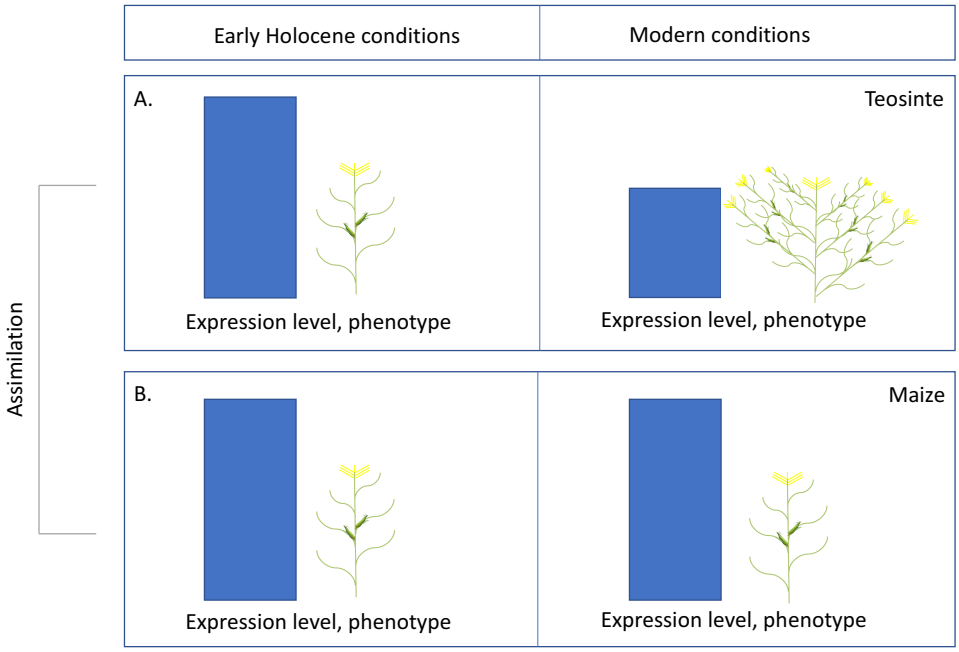


Figure 2: Schematic representation of differences in plastic responses between maize and teosinte in Early-Holocene (EH) conditions.

(A) *parviglumis* plants exhibit maize-like phenotypes in the EH conditions (vegetative architecture, inflorescence sexuality and seed maturation). Phenotypes of *parviglumis* in modern conditions are typical of today's plants. These changes in phenotypes are associated with altered expression levels of over 2000 candidate loci in teosinte, here we represent the schematic expression of one gene between the two environments in teosinte. **(B)** In contrast, these same traits and underlying gene expression remain unchanged in maize between EH and modern conditions.

II.5 How convergent is adaptation?

Convergent adaptation is the result of independent events of similar phenotypic changes to adapt to analogous environmental constraints (Wood, Burke, and Rieseberg 2005). In this review, we concentrated on genetic convergence in populations of the same, or closely related, species which are the results of convergent evolution at the molecular level. By molecular convergence, we include convergence at the same nucleotide positions, genes or orthologues. Several studies illustrate this, suggesting that genomes may respond in predictable ways to selection (Chan et al. 2010; Colosimo et al. 2005; Pearce et al. 2009; Roesti et al. 2014; Stern 2013). The selected alleles can originate from independent mutation events in different lineages, from shared ancestral variation or enter in the population by introgression (Stern 2013).

A classical way to study convergence is experimental evolution. During these experiments, replicates of the same genotype are grown for many generations in new environments. Such studies have often shown that convergent evolution is common (Riehle, Bennett, and Long 2001; Tenaillon et al. 2012).

Domestication is a striking example of phenotypic convergence with common traits usually referred to as the domestication syndrome. These phenotypes include, but are not limited to, larger fruits or gains, less branching, loss of shattering, and loss of seed dormancy (Gaut 2015). QTL

mapping can be performed to identify the genes controlling these phenotypes in different species. As an example, seeds on wild grasses shed naturally at maturity. During the domestication of such species this natural trait was rapidly selected against since it causes inefficient harvesting (Fuller et al. 2014). QTL mapping of sorghum, rice and maize reveals that the *Shattering1* genes are involved in the loss of the dispersal mechanism and were under convergent evolution during their domestication (Lin et al. 2012).

But genetic convergence can also be observed in much shorter evolutionary time, at the intraspecific level across populations. Here genome scans for extreme differentiation in allele frequency between multiple pairs of diverged populations along gradients, for instance, are typically employed. This method has been used to test for convergent adaptation in highland maize landraces and teosintes. Fustier et al (Fustier et al. 2017) found several instances (24/40) of convergence involving the same haplotype in two gradients of adaptation to high altitude in teosintes. In maize, the Mesoamerican and South American populations independently adapted from distinct lowland populations to high elevation conditions (Vigouroux et al. 2008). These populations exhibit several similar phenotypic characteristics not observed in lowland populations such as changes in inflorescence morphology and stem coloration. A study found that highland adaptation is likely due to a combination of introgression events, selection on standing genetic variation and independent *de novo* mutations (Takuno et al. 2015). These studies also showed that convergent evolution involving identical nucleotide changes is uncommon and most selected loci arise from standing genetic variation present in lowland populations. This is not surprising given the relative short time frame of highland adaptation in maize compared to teosinte subspecies.

Recently, a new method has been developed to infer modes of convergence (Lee and Coop 2017), using covariance of allele frequencies in windows around a selected site to explicitly

compare different models of origin for a selected variant. This novel method should give a better insight on the genetic mechanisms underlying convergence.

II.6 Mechanisms of genetic adaptation in maize and teosintes

Populations of teosinte have long evolved under natural selection. In contrast, maize populations have been under artificial human selection that moved phenotypes towards optimal traits tailored to agriculture during a shorter time frame of ~9,000 years (Piperno and Flannery 2001; Matsuoka et al. 2002; Fukunaga et al. 2005). These time scales have left distinct genetic signatures. In theory, traits fixed by domestication should involve genes with larger effect sizes, and standing variation should be a major contributor to domestication (Wallace, Larsson, and Buckler 2014). This is supported by crosses between maize and teosinte that led to the discovery of six main QTLs responsible for major phenotypic differences between them, notably vegetative architecture and inflorescence sexuality (Beadle 1972; Briggs et al. 2007). Among these QTLs, genes with major phenotypic effects have been discovered such as *tb1* and *tga1* (*teosinte glume architecture1*). In addition to these major genes, a collection of targets (2 to 4% of the genome according to Wright et al. 2005 and Hufford, Xu, et al. 2012) have likely contributed to the domesticated phenotype. In contrast, Genome Wide Association (GWA) studies on traits selected over much longer time scale such as drought tolerance or flowering time have highlighted only minor effect loci that rarely contribute to more than 5% of the phenotypic variation (Buckler et al. 2009; Cook et al. 2012; Wallace, Larsson, and Buckler 2014).

In addition to the time frame over which adaptation occurs, another important factor for evolution is the nature of variation for selection to act on. Maize and teosintes are genetically very diverse, with as much nucleotide diversity in coding regions between two maize lines as there are between humans and chimpanzees (Tian, Stevens, and Buckler 2009). This diversity is even higher

in intergenic regions (Tenaillon et al. 2001; Buckler, Gaut, and McMullen 2006). Some adaptive mutations are found in coding sequences. Examples include non-synonymous changes in the *tga1* gene responsible for the "naked kernel" maize phenotype, and in the *diacylglycerol acyltransferase* (*DGAT1-2*) gene resulting in elevated kernel oil content in maize lines (Wang et al. 2005; Zheng et al. 2008). But most observations support adaptation from regulatory non-coding sequences. Indeed, in comparison with *Arabidopsis*, where adaptive variants are enriched in coding sequences (Hancock et al. 2011), in maize and teosinte these are predominantly found in non-coding region: estimates in *Zea* show that non-coding variation may explain as much of the phenotypes as the coding regions (Chia et al. 2012; Rodgers-Melnick et al. 2016). Selection on regulatory sequences drive important expression changes; hence, genes displaying footprints of selection in maize are usually more expressed than in teosintes (Hufford, Xu, et al. 2012), and are associated with modified co-expression networks (Swanson-Wagner et al. 2012).

Adaptive variation also results from structural variants. In contrast to the *Arabidopsis* or rice genomes where Transposable Elements (TEs) account for 20-40% of sequence, the maize genome is composed of about 85% TEs (Schnable et al. 2009; Tenaillon, Hollister, and Gaut 2010). Genome size varies considerably within *Zea* resulting in over 30% differences among maize lines or landraces (Chia et al. 2012; Diez et al. 2014; Muñoz-Diez et al. 2012). Because of their deleterious effect, TEs are often negatively selected and silenced by DNA methylation (Hollister and Gaut 2009). But some may also impact gene expression and function in a beneficial manner by various mechanisms such as gene inactivation or differential expression caused by insertion in regulatory regions (Waters et al. 2017) or TE-mediated genomic rearrangements causing gene insertion, deletion or duplication (reviewed in Vitte et al. 2014). A handful of examples of their beneficial impact has been reported in *Zea*. A classic example in maize is at the *tb1* locus, where

a transposon inserted in the cis-regulatory region, doubling expression (Studer et al. 2011a). Teosinte, like most grasses, produces numerous branches tipped by a male inflorescence. In contrast, maize has only one main stalk terminated by a single tassel with repressed development of lateral branches. The increased expression level of *tb1* is the major contributor to this apical dominance (Studer et al. 2011a). Beyond TEs, Copy Number Variants (CNVs) are also common in the maize genome (Springer et al. 2009) and they contribute significantly to phenotypic variation (Chia et al. 2012; Maron et al. 2013).

Another important player in adaptation in *Zea* is gene flow. Indeed, teosinte populations are found in sympatry with maize and hybridization between them is common (Baltazar et al. 2005). Highland maize shows up to 20% *mexicana* introgression, which has likely facilitated their adaptation to high elevations (van Heerwaarden et al. 2011; Hufford et al. 2013). An ancient DNA study revealed that ancestral highland maize already showed evidence of introgression from *mexicana* (Da Fonseca et al. 2015). Introgressed regions found at high frequency in highland maize overlap with previously identified QTLs driving adaptive traits (Hufford et al. 2013; Lauter et al. 2004), emphasizing the importance of introgression during post-domestication adaptation. Similarly, recent results suggest that admixture between distinct genetic groups has facilitated adaptation to mid-latitudes in North America and Europe (Brandenburg et al. 2017).

II.7 What constraints adaptation?

Genetic adaptation can proceed through a single beneficial mutation that occurs after the onset of selection pressure, in which case the classical genetic footprint of a “hard” selective sweep is observed. Alternatively, it can proceed through a single mutation segregating in the population

before the onset of selection (standing genetic variation), or through recurrent beneficial mutations. In these latter cases, adaptation produces a "soft" sweep footprint (Hermisson and Pennings 2017).

Hard sweeps are characterized by local shifts in allele frequencies due to the hitchhiking of neutral sites around a selected *de novo* variant occurring on a specific haplotype. Such changes in allele frequencies can easily be detected by genome scans. In contrast, soft sweeps, which derive from multiple adaptive alleles sweeping in the population are substantially harder to detect at a genome-wide scale. When spatially structured species are subject to new selection pressures, independent adaptation can arise in different populations, hence increasing the total number of selective sweeps (Franssen, Kofler, and Schlötterer 2017). Population structure may also mislead analyses based on species-wide sampling, causing multiple independent hard sweeps in different populations to appear soft.

The relative contribution of hard and soft sweeps has been a long-standing debate and ultimately raises the important question of what limits adaptation. Experimental evolution in model organisms with short generation time such as *Escherichia Coli*, yeast and *Drosophila melanogaster* can provide insights into those questions (Bell and Collins 2008; Burke et al. 2010; Tenaillon et al. 2012; Burke, Liti, and Long 2014; Graves et al. 2017; Good et al. 2017). What emerges from these studies is that relevant parameters include mutations, drift, selection and the power to detect selection targets (Franssen, Kofler, and Schlötterer 2017; Schlötterer et al. 2014). We surveyed these parameters in eight divergent selection experiments undertaken in maize (Table 1) and detail below our interpretations. By applying continuous directional selection on a given quantitative trait, such experiments aim to quantify and understand the limits of selection. However, it should be noted none of the cited work (Table 1) has included multiple replicates.

One of the most puzzling observations across experiments is that the response to selection is generally steady over time. In the Golden Glow (GG) experiment, the response varies from 4.7% to 8.7% of the original phenotypic value per cycle of selection across 24 cycles (De Leon and Coors 2002). In the Krug Yellow Dent (KYD), it was estimated at 1.6% and 2.5% per cycle respectively, for high and low seed size direction (Lopez-Reynoso and Hallauer 1998). In the Iowa Stalk Synthetic (BSSS), the response was of 3.9% per cycle for higher grain yield (Lamkey 1992). In the Iowa Long Ear Synthetic (BSLE), an increase of 1.4% and a decrease of 1.9% per cycle for high and low ear length was observed (Lopez-Reynoso and Hallauer 1998). The results were more equivocal for Burn's White (BW), for which the response is much stronger and steadier towards high (between 0.1% and 0.3%) than low values (between 0% and 0.32%) for both protein and oil content. This pattern of shift between a strong and steady response to a plateau-like response for the low trait values is explained by physiological limits. Hence after 65 generations a lower limit for protein content is reached while the percentage of oil in the grain (close to 0% in the late generations) is no longer detectable (Dudley and Lambert 2010; Moose, Dudley, and Rocheford 2004). A similar situation has been reported for some of the late flowering families of MBS847 and F252 that are not able to produce seeds in the local climate conditions where they are selected while the early still display a significant response after 16 generations (Durand et al. 2015). Overall mutations do not appear limiting regardless of the design, whether it started from highly inbred material or a diverse set of intercrossed landraces (Table 1).

What differs from one experiment to another, however, is the genomic footprint of the response to selection. Such footprints have been investigated in all but the BW and BSLE design. In GG, in which the mutational target size – the number of sites affecting the trait – was restricted, the effective population size was the highest of all and the selection was intense, the signal is

consistent with genome-wide soft sweeps (De Leon and Coors 2002; Maita and Coors 1996). In KYD, characterized by a larger mutational target, stronger drift (smaller effective population size), but weaker selection, both hard and soft sweeps are observed (Odhiambo and Compton 1987). In BSSS, in which the mutational target size is the largest, the effective population size small and the selection intense, the signal is consistent with hard sweeps (Gerke et al. 2015). In F252 and MBS that display the most limiting standing variation, and at the same time the strongest drift and selection of all experiments, a rapid fixation of mutations explains the response to selection (Durand et al. 2010; Durand et al. 2015). Effective population size primarily determines the likelihood of soft sweeps. Hence, when θ (four times the product of effective population size and neutral mutation rate) is equal or above 1, and selection is strong enough, adaptation proceeds from multiple *de novo* mutations or standing variation (Messer and Petrov 2013). Below 1, soft sweeps' contribution diminishes with θ. In cited experiments (Table 1), selection is strong but $\theta << 1$. Nevertheless, hard and soft sweeps were associated respectively with the lowest (in F252 and MBS) and highest (in GG) effective population size, consistent with *Ne* being a key player. Comparisons among experiments contribute to point out the parameters of importance and their interactions that together shape the genomic patterns of the response to selection.

An additional layer of complexity that may substantially impact evolutionary trajectories is that of genetic correlations among traits. Such correlations may emerge from genes with pleiotropic effects, epistatic interactions among genes, and/or loci in tight linkage affecting various traits. While some studies have found that covariance between traits rarely affect adaptation (Agrawal and Stinchcombe 2009), several examples instead suggest that they may either constrain or facilitate adaptation as predicted by Lande (1979). For instance, in *Arabidopsis thaliana* a recent study indicates that polymorphisms with intermediate degrees of pleiotropy favored rapid

adaptation to micro-habitats *in natura*. In the case of domestication, tight linkage between genes conferring the so-called domestication syndrome has been invoked as a mechanism facilitating adaptation to the cultivated environment in allogamous species, preventing gene flow from wild relatives to break co-adapted suites of alleles (Thierry D'Ennequin et al. 1999). It turns out that rather than clustering, plant domestication genes identified so far are mainly transcription factors (reviewed in Martínez-Ainsworth and Tenaillon 2016) most of which likely display strong epistatic interactions. *tb1* in maize, for instance, interacts with another locus on a different chromosome to alter the sex of maize inflorescences. The introgression of the *tb1* teosinte allele alone changes only ~20% of the inflorescence sex but the introgression of both alleles converts 90% of maize's female flowers to male (Lukens and Doebley 1999). The maize *tb1* allele segregates at low frequency in teosinte populations but is rarely found associated with the domesticated allele of chromosome 3, as both are likely to evolve under negative selection in teosinte (Doebley, Stec, and Gustus 1995; Lukens and Doebley 1999). Their association in maize has however facilitated the acquisition of the domesticated phenotype.

II.8 Conclusion

Ongoing global warming has drastic effects on maize production, with an estimated impact of temperature and precipitation on yield of 3.8% worldwide between 1980 and 2008 (Lobell, Schlenker, and Costa-Roberts 2011). Predicted changes that include further increases in temperatures and decline in rainfall, as well as shifts of pests and diseases, represent a huge challenge. There is, therefore, a pressing need to better understand the dynamics and genomic bases of adaptation. Future climate projections predict that changes in temperature will impact the distribution and survival of *Zea* populations, and will have an even greater impact on wild species

(Hufford, Martínez-Meyer, et al. 2012; Ureta et al. 2012). However, most modeling studies focused on the climate tolerance of species, while the response to climate can depend on other factors such as the level of plasticity and local adaptation. This suggests that the response should be studied at the level of individual populations to better understand the basis of adaptation.

DS experiments	F252 (F252)	MBS847 (MBS)	Krug Yellow Dent (KYD)	Burn's White (BW)	Burn's White (BW)	Golden Glow (GG)	Iowa Stiff Stalk Synthetic (BSSS)	Iowa Long Ear Synthetic (BSLE)
References [a]	[1,2]	[1,2]	[3]	[4,5]	[4,5]	[6,7]	[10,11,12]	[14]
Directions (High/Low) [b]	H/L	H/L	H/L	H/L	H/L	H	H	H/L
Trait [c]	Flowering	Flowering	Seed size	Protein	Oil	Ears/plant	Grain yield	Ear length
Material type [d]	Inbred	Inbred	OP variety	OP variety	OP variety	OP variety	Synthetic population	Synthetic population
Mutational target [e]	>60 QTLs[8]	>60 QTLs[8]	>300 loci[9]	102 to 178 factors	14 to 69 factors	limited[6]	large[13]	25 QTLs[15]
Standing variation [f]	1.9%	0.19%	pervasive	pervasive	pervasive	pervasive	pervasive	pervasive
Census population size [g]	1000	1000	1200 to 1500	60 to 120	60 to 120	4250 (1-12) 14250 (13-30)	>1240	4000
N_e [h]	3.1 to 20.2	5.8 to 13.5	369	4 to 12	4 to 12	667	10 to 20	14
Selection coefficient (%)	1	1	8	20	20	0.5 to 5	5	7.5
Heritability [i]	0.14/0.13	0.13/0.16		0.21/0.07	0.23/0.23	0.88	0.4	0.05
Number of founders [j]	2 haplotypes	2 haplotypes	100 founders	24 ears (H) 12 ears (L)	24 ears (H) 12 ears (L)	~300 founders	16 founders	12 founders
Reproductive mode	Selfing	Selfing	Outcrossing	Outcrossing	Outcrossing	Outcrossing	Outcrossing	Outcrossing
Sampling [k]	All/ind	All/ind	All/bulk	All/bulk	All/bulk	All/bulk	All/bulk	All/bulk
Number of generations	16	16	30	114	114	30	17	27

Table 1. Description of eight long-term (>16 generations) Divergent Selection (DS) experiments in maize with groups of features primarily (but not exclusively) related to Mutations (3), Drift (1), Selection (2) and Power to detect selection targets (5) highlighted by groups.

[a] : References from which values were taken for each DS experiment are indicated in superscript.

[b] : Direction of selection towards higher and/or lower values than the initial material.

[c] : Protein and Oil designate protein and oil content of the grain, Ears/plan relates to prolificacy.

[d] : Inbred: Inbred line; OP variety: Open Pollinated population.

[e] : Number of factors in BW was estimated from the trait value, predicted gain and additive genetic variance.

[f] : Standing variation was estimated from 50k SNP array for F252 and MBS.

[g] : For GG, 4250 individuals were evaluated from cycles 1 to 12, and 14250 in the following cycles.

[h] : Effective population size (N_e) estimates given from the variance of offspring number (Crow, Kimura, and others 1970), range is given when N_e was estimated at each generation

[i] : Broad-sense heritability estimated from genetic variation between progenies of the same family. Average values across generations is reported here.

[j] : expressed either as number of haplotypes (a single founder=individual bears 2 haplotypes), number of founders, or number of ears (all individuals of a given ear share identical mother but different fathers). For GG, most selection cycles used 300 founders.

[k] : Seeds from all time points (All) are available, and were either collected separately on each selected individual (/ind) or in bulk (/bulk).

[1]: Durand et al. 2010

[2]: Durand et al. 2015

[3]: Odhiambo and Compton 1987

[4]: Moose, Dudley, and Rocheford 2004

[5]: Dudley and Lambert 2010

[6]: De Leon and Coors 2002

[7]: Maita and Coors 1996

[8]: Buckler et al. 2009

[9]: Liu et al. 2017

[10]: Lamkey 1992

[11]: Hallauer, Carena, and Filho 2010

[12]: Gerke et al. 2015

[13]: Yang et al. 2017

[14]: Lopez-Reynoso 1997

[15]: Ross, Hallauer, and Lee 2006

II.9 Acknowledgements

Maud Tenaillon acknowledges the Kavli Institute for Theoretical Physics at UCSB, supported in part by the National Science Foundation under Grant No. NSF PHY-1125915. The authors wish to acknowledge funding support from the US National Science Foundation (IOS-0922703 and IOS-1238014) as well as Michelle Stitzer, Mirko Ledda and Domenica Manicacci for comments.

II.10 References

Agrawal, A. F., and J. R. Stinchcombe. 2009. "How Much Do Genetic Covariances Alter the Rate of Adaptation?" *Proceedings of the Royal Society B: Biological Sciences* 276 (1659):1183–91. https://doi.org/10.1098/rspb.2008.1671.

Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. "Connecting Genomic Patterns of Local Adaptation and Niche Suitability in Teosintes." *Molecular Ecology* 26 (16):4226–40. https://doi.org/10.1111/mec.14203.

Anderson, J. T., and M. A. Geber. 2010. "Demographic Source-Sink Dynamics Restrict Local Adaptation in Elliott's Blueberry (Vaccinium Elliottii)." *Evolution* 64 (2):370–84. https://doi.org/10.1111/j.1558-5646.2009.00825.x.

Anderson, J. T., C. R. Lee, C. A. Rushworth, R. I. Colautti, and T. Mitchell-Olds. 2013. "Genetic Trade-Offs and Conditional Neutrality Contribute to Local Adaptation." In *Molecular Ecology*, 22:699–708. https://doi.org/10.1111/j.1365-294X.2012.05522.x.

Auld, J. R., A. A. Agrawal, and R. A. Relyea. 2010. "Re-Evaluating the Costs and Limits of Adaptive Phenotypic Plasticity." *Proceedings of the Royal Society B: Biological Sciences* 277 (1681):503–11. https://doi.org/10.1098/rspb.2009.1355.

Baltazar, M., J. De Jesús Sánchez-Gonzalez, L. De La Cruz-Larios, and J. B. Schoper. 2005. "Pollination between Maize and Teosinte: An Important Determinant of Gene Flow in Mexico." *Theoretical and Applied Genetics* 110 (3):519–26. https://doi.org/10.1007/s00122-004-1859-6.

Beadle, G. W. 1972. “Mystery of Maize.” *Field Museum Natural History Bulletin*.

Bebber, D. P., M. A. T. Ramotowski, and S. J. Gurr. 2013. “Crop Pests and Pathogens Move Polewards in a Warming World.” *Nature Climate Change* 3 (11):985–88. https://doi.org/10.1038/nclimate1990.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, and J. Ross-Ibarra. 2016. “Recent Demography Drives Changes in Linked Selection across the Maize Genome.” *Nature Plants* 2 (7):16084. https://doi.org/10.1038/nplants.2016.84.

Beldade, P., A. Mateus, and R. A. Keller. 2011. “Evolution and Molecular Mechanisms of Adaptive Developmental Plasticity.” *Mol Ecol* 20 (7):1347–63. https://doi.org/10.1111/j.1365-294X.2011.05016.x.

Bell, G., and S. Collins. 2008. “Adaptation, Extinction and Global Change.” *Evolutionary Applications* 1 (1):3–16. https://doi.org/10.1111/j.1752-4571.2007.00011.x.

Brandenburg, J. T., T. Mary-Huard, G. Rigaill, S. J. Hearne, H. Corti, J. Joets, C.Vitte, A. Charcosset, S. D. Nicolas, and M. I. Tenaillon. 2017. “Independent Introductions and Admixtures Have Contributed to Adaptation of European Maize and Its American Counterparts.” *PLoS Genetics* 13 (3). https://doi.org/10.1371/journal.pgen.1006666.

Briggs, W. H, M. D. McMullen, B. S. Gaut, and J. Doebley. 2007. “Linkage Mapping of Domestication Loci in a Large Maize Teosinte Backcross Resource.” *Genetics* 177 (3):1915–28. https://doi.org/10.1534/genetics.107.076497.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, et al. 2009. “The Genetic Architecture of Maize Flowering Time.” *Science* 325 (5941):714–18. https://doi.org/10.1126/science.1174276.

Buckler, E. S., B. S. Gaut, and M. D. McMullen. 2006. “Molecular and Functional Diversity of Maize.” *Current Opinion in Plant Biology*. https://doi.org/10.1016/j.pbi.2006.01.013.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, and A. D. Long. 2010. “Genome-Wide Analysis of a Long-Term Evolution Experiment with Drosophila.” *Nature* 467 (7315):587–90. https://doi.org/10.1038/nature09352.

Burke, M. K., G. Liti, and A. D. Long. 2014. “Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of Saccharomyces Cerevisiae.” *Molecular Biology and Evolution* 31 (12):3228–39. https://doi.org/10.1093/molbev/msu256.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, et al. 2010. “Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitxl Enhancer.” *Science* 327 (5963):302–5. https://doi.org/10.1126/science.1182213.

Chia, J. M., C. Song, P. J. Bradbury, D. Costich, N. De Leon, J. Doebley, R. J. Elshire, et al. 2012. "Maize HapMap2 Identifies Extant Variation from a Genome in Flux." *Nature Genetics* 44 (7):803–7. https://doi.org/10.1038/ng.2313.

Chuine, I. 2010. "Why Does Phenology Drives Species Distribution?" *Philosophical Transactions of the Royal Society of London B* 365:3149–60. https://doi.org/doi: 10.1098/rstb.2010.0142.

Clark, J. S., C. Fastie, G. Hurtt, S. T. Jackson, C. Johnson, A. King, M. Lewis, et al. 1998. "Reid's Paradox of Plant Rapid Migration Dispersal Theory and Interpretation of Paleoecological Records." *BioScience* 48 (1):13–24. https://doi.org/10.2307/1313224.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, H. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. "Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles." *Science* 307 (5717):1928–33. https://doi.org/10.1126/science.1107239.

Cook, J. P., M. D. McMullen, J. B. Holland, F. Tian, P. Bradbury, J. Ross-Ibarra, E. S. Buckler, and S. A. Flint-Garcia. 2012. "Genetic Architecture of Maize Kernel Composition in the Nested Association Mapping and Inbred Association Panels." *PLANT PHYSIOLOGY* 158 (2):824–34. https://doi.org/10.1104/pp.111.185033.

Crow, J. F., M. Kimura, and others. 1970. "An Introduction to Population Genetics Theory." *An Introduction to Population Genetics Theory.* New York, Evanston and London: Harper & Row, Publishers.

Cvijovic, I., B. H Good, and M. M. Desai. 2017. "The Effect of Strong Purifying Selection on Genetic Diversity." *Doi.org*, 211557. https://doi.org/10.1101/211557.

Davis, M. B., and R. G. Shaw. 2001. "Range Shifts and Adaptive Responses to Quaternary Climate Change." *Science* 292 (April):673–79. https://doi.org/10.1126/science.292.5517.673.

DeWitt, T. J., A. Sih, and D. S. Wilson. 1998. "Costs and Limits of Phenotypic Plasticity." *Trends in Ecology and Evolution*. https://doi.org/10.1016/S0169-5347(97)01274-3.

Diez, C. M., E. Meca, M. I. Tenaillon, and B. S. Gaut. 2014. "Three Groups of Transposable Elements with Contrasting Copy Number Dynamics and Host Responses in the Maize (Zea Mays Ssp. Mays) Genome." *PLoS Genetics* 10 (4). https://doi.org/10.1371/journal.pgen.1004298.

Diggle, P. K., and J. S. Miller. 2013. "Developmental Plasticity, Genetic Assimilation, and the Evolutionary Diversification of Sexual Expression in Solanum." *Am J Bot* 100 (6):1050–60. https://doi.org/10.3732/ajb.1200647.

Doebley, J., A. Stec, and C. Gustus. 1995. “Teosinte branched1 and the Origin of Maize: Evidence for Epistasis and the Evolution of Dominance.” *Genetics* 141 (1):333–46.

Dudley, J. W., and R. J. Lambert. 2010. “100 Generations of Selection for Oil and Protein in Corn.” In *Plant Breeding Reviews*, 79–110. https://doi.org/10.1002/9780470650240.ch5.

Durand, E., M. I. Tenaillon, X. Raffoux, S. Thépot, M. Falque, P. Jamin, A. Bourgais, A. Ressayre, and C. Dillmann. 2015. “Dearth of Polymorphism Associated with a Sustained Response to Selection for Flowering Time in Maize.” *BMC Evolutionary Biology* 15 (1):103. https://doi.org/10.1186/s12862-015-0382-5.

Durand, E., M. I. Tenaillon, C. Ridel, D. Coubriche, P. Jamin, S. Jouanne, A. Ressayre, A. Charcosset, and C. Dillmann. 2010. “Standing Variation and New Mutations Both Contribute to a Fast Response to Selection for Flowering Time in Maize Inbreds.” *BMC Evolutionary Biology* 10:2. https://doi.org/10.1186/1471-2148-10-2.

Eyre-Walker, a, R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. “Investigation of the Bottleneck Leading to the Domestication of Maize.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (April):4441–46. https://doi.org/10.1073/pnas.95.8.4441.

Fang, Z., T. Pyhäjärvi, A. L. Weber, R. K. Dawe, J. C. Glaubitz, J. de Jesus Sánchez González, C. Ross-Ibarra, J. Doebley, P. L. Morrell, and J. Ross-Ibarra. 2012. “Megabase-Scale Inversion Polymorphism in the Wild Ancestor of Maize.” *Genetics* 191 (3):883–94. https://doi.org/10.1534/genetics.112.138578.

Fonseca, R. R. Da, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. A. Samaniego, et al. 2015. “The Origin and Evolution of Maize in the Southwestern United States.” *Nature Plants* 1. https://doi.org/10.1038/nplants.2014.3.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. 2011. “A Map of Local Adaptation in Arabidopsis Thaliana.” *Science* 334 (6052):86–89. https://doi.org/10.1126/science.1209271.

Franssen, S. U., R. Kofler, and C. Schlötterer. 2017. “Uncovering the Genetic Signature of Quantitative Trait Evolution with Replicated Time Series Data.” *Heredity* 118 (1):42–51. https://doi.org/10.1038/hdy.2016.98.

Fukunaga, K., J. Hill, Y. Vigouroux, Y. Matsuoka, J. Sanchez, K. J. Liu, E. S. Buckler, and J. Doebley. 2005. “Genetic Diversity and Population Structure of Teosinte.” *Genetics* 169 (4):2241–54. https://doi.org/genetics.104.031393.10.1534/genetics.104.031393.

Fuller, D. Q., T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, and M. D. Purugganan. 2014. “Convergent Evolution and Parallelism in Plant Domestication Revealed by an Expanding Archaeological Record.” *Proceedings of the National Academy*

of Sciences 111 (17):6147–52. https://doi.org/10.1073/pnas.1308937110.

Fustier, M. A., J. T. Brandenburg, S. Boitard, J. Lapeyronnie, L. E. Eguiarte, Y. Vigouroux, D. Manicacci, and M. I. Tenaillon. 2017. "Signatures of Local Adaptation in Lowland and Highland Teosintes from Whole-Genome Sequencing of Pooled Samples." *Molecular Ecology* 26 (10):2738–56. https://doi.org/10.1111/mec.14082.

Gage, J. L., D. Jarquin, C. Romay, A. Lorenz, E. S. Buckler, S. Kaeppler, N. Alkhalifah, et al. 2017. "The Effect of Artificial Selection on Phenotypic Plasticity in Maize." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01450-2.

Gaut, B. S. 2015. "Evolution Is an Experiment: Assessing Parallelism in Crop Domestication and Experimental Evolution." *Molecular Biology and Evolution* 32 (7):1661–71. https://doi.org/10.1093/molbev/msv105.

Gerke, J. P., J. W. Edwards, K. E. Guill, J. Ross-Ibarra, and M. D. McMullen. 2015. "The Genomic Impacts of Drift and Selection for Hybrid Performance in Maize." *Genetics* 201 (3):1201–11. https://doi.org/10.1534/genetics.115.182410.

Gilbert, S. F., and D. Epel. 2009. "Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution." *Sinauer Associates*.

Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. "The Dynamics of Molecular Evolution over 60,000 Generations." *Nature* 551 (7678):45–50. https://doi.org/10.1038/nature24287.

Graves, J. L., K. L. Hertweck, M. A. Phillips, M. V. Han, L. G. Cabral, T. T. Barter, L. F. Greer, et al. 2017. "Genomics of Parallel Experimental Evolution in Drosophila." *Molecular Biology and Evolution* 34 (4):831–42. https://doi.org/10.1093/molbev/msw282.

Hallauer, A. R., M. J. Carena, and J.B. M. Filho. 2010. "Quantitative Genetics in Maize Breeding." In *Handbook of Plant Breeding*, 1–680. https://doi.org/10.1017/CBO9781107415324.004.

Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz, F. G. Sperone, C. Toomajian, F. Roux, and J. Bergelson. 2011. "Adaptation to Climate Across the Arabidopsis Thaliana Genome." *Science* 334 (6052):83–86. https://doi.org/10.1126/science.1209244.

Heerwaarden, J. van, J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. de Jesus Sanchez Gonzalez, and J. Ross-Ibarra. 2011. "Genetic Signals of Origin, Spread, and Introgression in a Large Sample of Maize Landraces." *Proceedings of the National Academy of Sciences* 108 (3):1088–92. https://doi.org/10.1073/pnas.1013011108.

Hermisson, J., and P. S. Pennings. 2017. "Soft Sweeps and beyond: Understanding the Patterns

and Probabilities of Selection Footprints under Rapid Adaptation." *Methods in Ecology and Evolution* 8 (6):700–716. https://doi.org/10.1111/2041-210X.12808.

Hill, J. K., C. D. Thomas, and B. Huntley. 1999. "Climate and Habitat Availability Determine 20th Century Changes in a Butterfly's Range Margin." *Proceedings of the Royal Society B: Biological Sciences* 266 (1425):1197–1206. https://doi.org/10.1098/rspb.1999.0763.

Hollister, J. D., and B. S. Gaut. 2009. "Epigenetic Silencing of Transposable Elements: A Trade-off between Reduced Transposition and Deleterious Effects on Neighboring Gene Expression." *Genome Research* 19 (8):1419–28. https://doi.org/10.1101/gr.091678.109.

Hudson, R. R. 2002. "Generating Samples under a Wright-Fisher Neutral Model of Genetic Variation." *Bioinformatics* 18 (2):337–38. https://doi.org/10.1093/bioinformatics/18.2.337.

Hufford, M. B., P. Bilinski, T. Pyhäjärvi, and Jeffrey Ross-Ibarra. 2012. "Teosinte as a Model System for Population and Ecological Genomics." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.08.004.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. "The Genomic Signature of Crop-Wild Introgression in Maize." *PLoS Genetics* 9 (5). https://doi.org/10.1371/journal.pgen.1003477.

Hufford, M. B., E. Martínez-Meyer, B. S. Gaut, L. E. Eguiarte, and M. I. Tenaillon. 2012. "Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight." *PLoS ONE* 7 (11). https://doi.org/10.1371/journal.pone.0047659.

Hufford, M. B., E. Martinez-Meyer, B. S. Gaut, L. E. Eguiarte, and M. I. Tenaillon. 2012. "Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight." *PLoS One* 7 (11):e47659. https://doi.org/10.1371/journal.pone.0047659.

Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J. M. Chia, R. A. Cartwright, R. J. Elshire, et al. 2012. "Comparative Population Genomics of Maize Domestication and Improvement." *Nat Genet* 44 (7):808–11. https://doi.org/10.1038/ng.2309.

Hung, H-Y., L. M. Shannon, F. Tian, P. J. Bradbury, C. Chen, S. A. Flint-Garcia, M. D. McMullen, et al. 2012. "ZmCCT and the Genetic Basis of Day-Length Adaptation Underlying the Postdomestication Spread of Maize." *Proceedings of the National Academy of Sciences* 109 (28):E1913–21. https://doi.org/10.1073/pnas.1203189109.

Jaenicke-Despres, V. 2003. "Early Allelic Selection in Maize as Revealed by Ancient DNA." *Science* 302 (5648):1206–8. https://doi.org/10.1126/science.1089056.

Kawecki, T. J., and D. Ebert. 2004. "Conceptual Issues in Local Adaptation." *Ecology Letters*

7:1225–41. https://doi.org/10.1111/j.1461-0248.2004.00684.x.

Kleunen, M. van, and M. Fischer. 2005. “Constraints on the Evolution of Adaptive Phenotypic Plasticity in Plants.” *The New Phytologist* 166 (1):49–60. https://doi.org/10.1111/j.1469-8137.2004.01296.x.

Kuzawa, C. W., and J. M. Bragg. 2012. “Plasticity in Human Life History Strategy.” *Current Anthropology* 53 (S6). JSTOR:S369--S382.

Kwak, M., O. Toro, D. G. Debouck, and P. Gepts. 2012. “Multiple Origins of the Determinate Growth Habit in Domesticated Common Bean (Phaseolus Vulgaris).” *Annals of Botany* 110 (8):1573–80. https://doi.org/10.1093/aob/mcs207.

Lamkey, K. R. 1992. “Fifty Years of Recurrent Selection in the Iowa Stiff Stalk Synthetic Maize Population.” *Maydica* 37 (1):19.

Lande, R. 1979. “Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry.” *Evolution* 33 (1):402. https://doi.org/10.2307/2407630.

Lauter, N., C. Gustus, A. Westerbergh, and J. Doebley. 2004. “The Inheritance and Evolution of Leaf Pigmentation and Pubescence in Teosinte.” *Genetics* 167 (4):1949–59. https://doi.org/10.1534/genetics.104.026997.

Lee, K. M, and G. Coop. 2017. “Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data.” *Genetics*, genetics.300417.2017. https://doi.org/10.1534/genetics.117.300417.

Leon, N. De, and J. G. Coors. 2002. “Twenty-Four Cycles of Mass Selection for Prolificacy in the Golden Glow Maize Population.” *Crop Science* 42 (2):325–33. https://doi.org/10.2135/cropsci2002.3250.

Lin, Z., X. Li, L. M. Shannon, C. T. Yeh, M. L. Wang, G. Bai, Z. Peng, et al. 2012. “Parallel Domestication of the Shattering1 Genes in Cereals.” *Nature Genetics* 44 (6):720–24. https://doi.org/10.1038/ng.2281.

Linhart, Y. B., and M. C. Grant. 1996. “EVOLUTIONARY SIGNIFICANCE OF LOCAL GENETIC DIFFERENTIATION IN PLANTS.” *Annual Review of Ecology and Systematics* 27 (1):237–77. https://doi.org/10.1146/annurev.ecolsys.27.1.237.

Liu, J., J. Huang, H. Guo, L. Lan, H. Wang, Y. Xu, X. Yang, et al. 2017. “The Conserved and Unique Genetic Architecture of Kernel Size and Weight in Maize and Rice.” *Plant Physiology* 175 (2). Am Soc Plant Biol:774–85.

Lobell, D. B, W. Schlenker, and J. Costa-Roberts. 2011. “Climate Trends and Global Crop Production since 1980.” *Science (New York, N.Y.)* 333 (6042):616–20. https://doi.org/10.1126/science.1204531.

Lopez-Reynoso, J. J. De. 1997. “Divergent Mass Selection for Ear Length in the Iowa Long Ear Synthetic Maize Cultivar after Twenty-Seven Cycles of Selection.” Digital Repository@ Iowa State University, http://lib. dr. iastate. edu/.

Lopez-Reynoso, J. J. De, and A. R. Hallauer. 1998. “Twenty-Seven Cycles of Divergent Mass Selection for Ear Length in Maize.” *Crop Science* 38 (4):1099–1107. https://doi.org/10.2135/cropsci1998.0011183X003800040035x.

Lorant, A., S. Pedersen, I. Holst, M. B. Hufford, K. Winter, D. Piperno, and J. Ross-Ibarra. 2017. “The Potential Role of Genetic Assimilation during Maize Domestication.” *PLoS ONE* 12 (9). https://doi.org/10.1371/journal.pone.0184202.

Lukens, L. N., and J. Doebley. 1999. “Epistatic and Environmental Interactions for Quantitative Trait Loci Involved in Maize Evolution.” *Genetical Research* 74 (3):291–302. https://doi.org/10.1017/S0016672399004073.

Lundy, M., I. Montgomery, and J. Russ. 2010. “Climate Change-Linked Range Expansion of Nathusius’ Pipistrelle Bat, Pipistrellus Nathusii (Keyserling & Blasius, 1839).” *Journal of Biogeography* 37 (12):2232–42. https://doi.org/10.1111/j.1365-2699.2010.02384.x.

Maita, R., and J. G. Coors. 1996. “Twenty Cycles of Biparental Mass Selection for Prolificacy in the Open-Pollinated Maize Population Golden Glow.” *Crop Science* 36 (6). Crop Science Society of America:1527–32.

Marais, D. L. Des, K. M. Hernandez, and T. E. Juenger. 2013. “Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment.” *Annual Review of Ecology, Evolution, and Systematics* 44. Annual Reviews:5–29.

Maron, L. G, C. T Guimarães, M. Kirst, P. S. Albert, J. A. Birchler, P. J. Bradbury, E. S. Buckler, et al. 2013. “Aluminum Tolerance in Maize Is Associated with Higher MATE1 Gene Copy Number.” *Proceedings of the National Academy of Sciences of the United States of America* 110:5241–46. https://doi.org/10.1073/pnas.1220766110.

Martínez-Ainsworth, N. E., and M. I. Tenaillon. 2016. “Superheroes and Masterminds of Plant Domestication.” *Comptes Rendus Biologies* 339 (7–8):268–73. https://doi.org/10.1016/j.crvi.2016.05.005.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. S. Buckler, and J. Doebley. 2002. “A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping.” *Proc Natl Acad Sci U S A* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

Mercer, K., Á. Martínez-Vásquez, and H. R. Perales. 2008. “Asymmetrical Local Adaptation of Maize Landraces along an Altitudinal Gradient.” *Evolutionary Applications* 1 (3):489–500.

https://doi.org/10.1111/j.1752-4571.2008.00038.x.

Messer, P. W., and D. A. Petrov. 2013. “Population Genomics of Rapid Adaptation by Soft Selective Sweeps.” *Trends in Ecology and Evolution* 28 (11):659–69. https://doi.org/10.1016/j.tree.2013.08.003.

Moeller, D. A., and P. Tiffin. 2008. “Geographic Variation in Adaptation at the Molecular Level: A Case Study of Plant Immunity Genes.” *Evolution* 62 (12):3069–81. https://doi.org/10.1111/j.1558-5646.2008.00511.x.

Monleon, V. J., and H. E. Lintz. 2015. “Evidence of Tree Species’ Range Shifts in a Complex Landscape.” *PLoS ONE* 10 (1). https://doi.org/10.1371/journal.pone.0118069.

Moose, S. P., J. W. Dudley, and T. R. Rocheford. 2004. “Maize Selection Passes the Century Mark: A Unique Resource for 21st Century Genomics.” *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2004.05.005.

Muñoz-Diez, C., C. Vitte, J. Ross-Ibarra, B. S. Gaut, and M. I. Tenaillon. 2012. “Using Nextgen Sequencing to Investigate Genome Size Variation and Transposable Element Content.” *Topics in Current Genetics*. https://doi.org/10.1007/978-3-642-31842-9-3.

Nicotra, A. B., O. K. Atkin, S. P. Bonser, A. M. Davidson, E. J. Finnegan, U. Mathesius, P. Poot, et al. 2010. “Plant Phenotypic Plasticity in a Changing Climate.” *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2010.09.008.

Odhiambo, M. O., and W. A. Compton. 1987. “Twenty Cycles of Divergent Mass Selection for Seed Size in Corn.” *Crop Science* 27 (6). Crop Science Society of America:1113–16.

Pauli, H., M. Gottfried, S. Dullinger, O. Abdaladze, M. Akhalkatsi, J. L. B. Alonso, G. Coldea, et al. 2012. “Recent Plant Diversity Changes on Europe’s Mountain Summits.” *Science* 336 (6079):353–55. https://doi.org/10.1126/science.1219033.

Pearce, R. J., H. Pota, M. S. B. Evehe, E. H. Bâ, G. Mombo-Ngoma, A. L. Malisa, R. Ord, et al. 2009. “Multiple Origins and Regional Dispersal of Resistant Dhps in African Plasmodium Falciparum Malaria.” *PLoS Medicine* 6 (4). https://doi.org/10.1371/journal.pmed.1000055.

Piperno, D. R., and K. V. Flannery. 2001. “The Earliest Archaeological Maize (Zea Mays L.) from Highland Mexico: New Accelerator Mass Spectrometry Dates and Their Implications.” *Proc Natl Acad Sci U S A* 98 (4):2101–3. https://doi.org/10.1073/pnas.98.4.2101.

Piperno, D. R., I. Holst, K. Winter, and O. McMillan. 2015. “Teosinte before Domestication: Experimental Study of Growth and Phenotypic Variability in Late Pleistocene and Early Holocene Environments.” *Quaternary International* 363. Elsevier:65–77.

Poets, A. M., Z. Fang, M. T. Clegg, and P. L. Morrell. 2015. “Barley Landraces Are

Characterized by Geographically Heterogeneous Genomic Origins." *Genome Biology* 16 (1). https://doi.org/10.1186/s13059-015-0712-3.

Pyhäjärvi, T., M. B. Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biol Evol* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Riehle, M. M., A. F. Bennett, and A. D. Long. 2001. "Genetic Architecture of Thermal Adaptation in Escherichia Coli." *Proceedings of the National Academy of Sciences* 98 (2):525–30. https://doi.org/10.1073/pnas.98.2.525.

Rodgers-Melnick, E., D. L. Vera, H. W. Bass, and E. S. Buckler. 2016. "Open Chromatin Reveals the Functional Maize Genome." *Proceedings of the National Academy of Sciences* 113 (22):E3177–84. https://doi.org/10.1073/pnas.1525244113.

Roesti, M., S. Gavrilets, A. P. Hendry, W. Salzburger, and D. Berner. 2014. "The Genomic Signature of Parallel Adaptation from Shared Genetic Variation." *Molecular Ecology* 23 (16):3944–56. https://doi.org/10.1111/mec.12720.

Ross-Ibarra, J., M. Tenaillon, and B. S. Gaut. 2009. "Historical Divergence and Gene Flow in the Genus Zea." *Genetics* 181 (4):1399–1413. https://doi.org/10.1534/genetics.108.097238.

Ross, A. J., A. R. Hallauer, and M. Lee. 2006. "Genetic Analysis of Traits Correlated with Maize Ear Length." *Maydica* 51 (2):301–13.

Savolainen, O., M. Lascoux, and J. Merilä. 2013. "Ecological Genomics of Local Adaptation." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3522.

Schlötterer, C., R. Kofler, E. Versace, R. Tobler, and S. U. Franssen. 2014. "Combining Experimental Evolution with next-Generation Sequencing: A Powerful Tool to Study Adaptation from Standing Genetic Variation." *Heredity*, no. April:1–10. https://doi.org/10.1038/hdy.2014.86.

Schluter, D. 2009. "Evidence for Ecological Speciation and Its Alternative." *Science*. https://doi.org/10.1126/science.1160006.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, et al. 2009. "The B73 Maize Genome: Complexity, Diversity, and Dynamics." *Science* 326 (5956):1112–15. https://doi.org/10.1126/science.1178534.

Springer, N. M., K. Ying, Y. Fu, T. Ji, C. T. Yeh, Y. Jia, W. Wu, et al. 2009. "Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/absence Variation (PAV) in Genome Content." *PLoS Genetics* 5 (11). https://doi.org/10.1371/journal.pgen.1000734.

Standen, E. M, T. Y. Du, and H. C. E. Larsson. 2014. "Developmental Plasticity and the Origin

of Tetrapods." *Nature* 513 (7516):54–58. https://doi.org/10.1038/nature13708.

Stern, D. L. 2013. "The Genetic Causes of Convergent Evolution." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3483.

Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. "Identification of a Functional Transposon Insertion in the Maize Domestication Gene tb1." *Nat Genet* 43 (11):1160–63. https://doi.org/10.1038/ng.942.

Swanson-Wagner, R., R. Briskine, R. Schaefer, M. B. Hufford, J. Ross-Ibarra, C. L. Myers, P. Tiffin, and N. M. Springer. 2012. "Reshaping of the Maize Transcriptome by Domestication." *Proc Natl Acad Sci U S A* 109 (29):11878–83. https://doi.org/10.1073/pnas.1201961109.

Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski, J. Holland, M. Kruse-Peeples, et al. 2017. "Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America." *Science* 357 (6350):512–15. https://doi.org/10.1126/science.aam9425.

Takuno, S., P. Ralph, K. Swart, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. 2015. "Independent Molecular Basis of Convergent Highland Adaptation in Maize." *Genetics* 200 (4):1297–1312. https://doi.org/10.1534/genetics.115.178327.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, and B. S. Gaut. 2001. "Patterns of DNA Sequence Polymorphism along Chromosome 1 of Maize (Zea Mays Ssp. Mays L.)." *Proceedings of the National Academy of Sciences of the United States of America* 98 (16):9161–66. https://doi.org/10.1073/pnas.151244298.

Tenaillon, M. I., J. D. Hollister, and B. S. Gaut. 2010. "A Triptych of the Evolution of Plant Transposable Elements." *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2010.05.003.

Tenaillon, M. I., J. U'Ren, O. Tenaillon, and B. S. Gaut. 2004. "Selection versus Demography: A Multilocus Investigation of the Domestication Process in Maize." *Molecular Biology and Evolution* 21 (7):1214–25. https://doi.org/10.1093/molbev/msh102.

Tenaillon, O., A. Rodriguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. 2012. "The Molecular Diversity of Adaptive Convergence." *Science* 335 (6067):457–61. https://doi.org/10.1126/science.1212986.

Thierry D'Ennequin, M. Le, B. Toupance, T. Robert, B. Godelle, and P. H. Gouyon. 1999. "Plant Domestication: A Model for Studying the Selection of Linkage." *Journal of Evolutionary Biology* 12 (6):1138–47. https://doi.org/10.1046/j.1420-9101.1999.00115.x.

Thomas, C. D., and J. J. Lennon. 1999. "Birds Extend Their Ranges Northwards." *Nature* 399

(May):213. https://doi.org/10.1038/20335.

Thompson, J., A. Charpentier, G. Bouguet, F. Charmasson, S. Roset, B. Buatois, P. Vernet, and P. Gouyon. 2013. "Evolution of a Genetic Polymorphism with Climate Change in a Mediterranean Landscape." *Proceedings of the National Academy of Sciences of the United States of America* 110 (8):2893–97. https://doi.org/10.1073/pnas.1215833110.

Tian, F., N. M. Stevens, and E. S. Buckler. 2009. "Tracking Footprints of Maize Domestication and Evidence for a Massive Selective Sweep on Chromosome 10." *Proceedings of the National Academy of Sciences* 106 (Supplement_1):9979–86. https://doi.org/10.1073/pnas.0901122106.

Tiffin, P., and J. Ross-Ibarra. 2014. "Advances and Limits of Using Population Genetics to Understand Local Adaptation." *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2014.10.004.

Ureta, C., E. Martínez-Meyer, H. R. Perales, and E. R. Álvarez-Buylla. 2012. "Projecting the Effects of Climate Change on the Distribution of Maize Races and Their Wild Relatives in Mexico." *Global Change Biology* 18 (3):1073–82. https://doi.org/10.1111/j.1365-2486.2011.02607.x.

Valverde, B. E. 2007. "Status and Management of Grass-Weed Herbicide Resistance in Latin America." *Weed Technology* 21 (2):310–23. https://doi.org/10.1614/WT-06-097.1.

Vigouroux, Y., J. C. Glaubitz, Y. Matsuoka, M. M. Goodman, J. Sánchez G., and J. Doebley. 2008. "Population Structure and Genetic Diversity of New World Maize Races Assessed by DNA Microsatellites." *American Journal of Botany* 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097.

Vitte, C., M. A. Fustier, K. Alix, and M. I. Tenaillon. 2014. "The Bright Side of Transposons in Crop Evolution." *Briefings in Functional Genomics and Proteomics* 13 (4):276–95. https://doi.org/10.1093/bfgp/elu002.

Wallace, J. G., S. J. Larsson, and E. S. Buckler. 2014. "Entering the Second Century of Maize Quantitative Genetics." *Heredity*. https://doi.org/10.1038/hdy.2013.6.

Wallace, J. G., P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E. S. Buckler. 2014. "Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize." *PLoS Genetics* 10 (12). https://doi.org/10.1371/journal.pgen.1004845.

Wang, H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, M. Faller, K. Bomblies, L. Lukens, and J. F. Doebley. 2005. "The Origin of the Naked Grains of Maize." *Nature* 436 (7051):714–19. https://doi.org/10.1038/nature03863.

Waters, A. J., I. Makarevitch, J. Noshay, L. T. Burghardt, C. N. Hirsch, C. D. Hirsch, and N. M.

Springer. 2017. “Natural Variation for Gene Expression Responses to Abiotic Stress in Maize.” *Plant Journal* 89 (4):706–17. https://doi.org/10.1111/tpj.13414.

Wennersten, L., and A. Forsman. 2012. “Population-Level Consequences of Polymorphism, Plasticity and Randomized Phenotype Switching: A Review of Predictions.” *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2012.00231.x.

Williams, J. W., B. N. Shuman, T. Webb, P. J. Bartlein, and P. L. Leduc. 2004. “Late-Quaternary Vegetation Dynamics in North America: Scaling from Taxa to Biomes.” *Ecological Monographs*. https://doi.org/10.1890/02-4045.

Wood, T. E., J. M. Burke, and L. H. Rieseberg. 2005. “Parallel Genotypic Adaptation: When Evolution Repeats Itself.” In *Genetica*, 123:157–70. https://doi.org/10.1007/s10709-003-2738-9.

Wright, S. I, I. V. Bi, S. G Schroeder, M. Yamasaki, J. F Doebley, Michael D McMullen, and Brandon S Gaut. 2005. “The Effects of Artificial Selection on the Maize Genome.” *Science (New York, N.Y.)* 308 (5726):1310–14. https://doi.org/10.1126/science.1107891.

Yang, J., S. Mezmouk, A. Baumgarten, E. S. Buckler, K. E. Guill, M. D. McMullen, R. H. Mumm, and J. Ross-Ibarra. 2017. “Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize.” *PLoS Genetics* 13 (9). https://doi.org/10.1371/journal.pgen.1007019.

Yang, Q., Z. Li, W. Li, L. Ku, C. Wang, J. Ye, K. Li, et al. 2013. “CACTA-like Transposable Element in ZmCCT Attenuated Photoperiod Sensitivity and Accelerated the Postdomestication Spread of Maize.” *Proceedings of the National Academy of Sciences* 110 (42):16969–74. https://doi.org/10.1073/pnas.1310949110.

Zheng, P., W. B. Allen, K. Roesler, M. E. Williams, S. Zhang, J. Li, K. Glassman, et al. 2008. “A Phenylalanine in DGAT Is a Key Determinant of Oil Content and Composition in Maize.” *Nature Genetics* 40 (3):367–72. https://doi.org/10.1038/ng.85.

III. THE INTERPLAY OF DEMOGRAPHY AND SELECTION DURING MAIZE DOMESTICATION AND EXPANSION

Li Wang[1,2], Timothy M. Beissinger[3,4,5], Anne Lorant[3], Claudia Ross-Ibarra[3], Jeffrey Ross-Ibarra[3,6*], Matthew B. Hufford[1*]

[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University

[2]Genome Informatics Facility, Iowa State University

[3]Department of Plant Sciences, University of California

[4]USDA-ARS Plant Genetics Research Unit

[5]Divisions of Plant and Biological Sciences, University of Missouri

[6]Genome Center and Center for Population Biology, University of California

Keywords: Demography, Domestication, Genetic load, Maize

Published: November 13, 2017 in Genome Biology https://doi.org/10.1186/s13059-017-1346-4

III.1 Abstract

Background

The history of maize has been characterized by major demographic events, including population size changes associated with domestication and range expansion, and gene flow with wild relatives. The interplay between demographic history and selection has shaped diversity across maize populations and genomes.

Results

We investigate these processes using high-depth resequencing data from 31 maize landraces spanning the pre-Columbian distribution of maize, and four wild teosinte individuals (*Zea mays* ssp. *parviglumis*). Genome-wide demographic analyses reveal that maize experienced pronounced declines in effective population size due to both a protracted domestication bottleneck and serial founder effects during post-domestication spread, while *parviglumis* in the Balsas River Valley experienced population growth. The domestication bottleneck and subsequent spread led to an increase in deleterious alleles in the domesticate compared to the wild progenitor. This cost is particularly pronounced in Andean maize, which has experienced a more dramatic founder event compared to other maize populations. Additionally, we detect introgression from the wild teosinte *Zea mays* ssp. *mexicana* into maize in the highlands of Mexico, Guatemala, and the southwestern USA, which reduces the prevalence of deleterious alleles likely due to the higher long-term effective population size of teosinte.

Conclusions

These findings underscore the strong interaction between historical demography and the efficiency of selection and illustrate how domesticated species are particularly useful for

understanding these processes. The landscape of deleterious alleles and therefore evolutionary potential is clearly influenced by recent demography, a factor that could bear importantly on many species that have experienced recent demographic shifts.

III.2 Introduction

Genomes are shaped over the course of their evolutionary history through a complex interaction of demography and selection. Neutral processes that comprise a species' demographic history, such as stochastic changes in population size and migration events, influence both the pool of diversity upon which selection can act and its efficiency. Selection and genetic drift then jointly determine the fate of this diversity.

After the development of agriculture, both crops and humans have experienced profound demographic shifts that left clear signatures in genome-wide patterns of diversity (Li et al. 2008; Beissinger et al. 2016). Early agriculturalists sampled a subset of the diversity present in crop wild relatives, resulting in an initial demographic bottleneck for many domesticates (Doebley, Gaut, and Smith 2006). Subsequent to domestication, humans and their crops experienced a process of global expansion facilitated by the rise of agriculture (Gignoux, Henn, and Mountain 2011). In many cases expansion was accompanied by gene flow with close relatives, a demographic process that further altered patterns of diversity (Hufford et al. 2013; Prüfer et al. 2014).

Recent interest in the effects of demography on functional variation has led to a growing body of theory that is increasingly supported by empirical examples. To date, the relationship between demography and selection has been most thoroughly explored in the context of deleterious alleles. While theory suggests mutation load (i.e., the reduction in mean fitness caused by the presence of deleterious alleles) may be insensitive to demography over long periods (Do et al.

2015; Simons et al. 2014), empirical results are consistent with load being shaped by demography over shorter timescales (Harris and Nielsen 2016; Fu et al. 2014; Zhang et al. 2016; Marsden et al. 2016). For example, evidence in both plant and animal species has revealed increased mutation load in populations that have undergone recent, sudden declines in effective population size (N_e) (Fu et al. 2014; Zhang et al. 2016; Marsden et al. 2016; Liu et al. 2017). Similarly, in geographically expanding populations, repeated sub-sampling of diversity (i.e., serial founder effects) can occur during migration away from a center of origin (Austerlitz et al. 1997; Slatkin and Excoffier 2012), a phenomenon shown to have decreased genetic diversity and increased counts of deleterious alleles in human populations more distant from Africa (Henn et al. 2015; Ramachandran et al. 2005). Finally, gene flow may also affect genome-wide patterns of deleterious variants, particularly when occurring between populations with starkly contrasting N_e. For instance, during the Out-of-Africa migration, modern humans inter-mated with the Neanderthal species, a close relative with substantially lower N_e and higher mutation load (Harris and Nielsen 2016). The higher mutation load in Neanderthals presented a cost of gene flow, and subsequent purifying selection appears to have limited the amount of Neanderthal introgression near genes in the modern human genome (Harris and Nielsen 2016; Juric, Aeschbacher, and Coop 2016).

The domesticated plant maize (*Zea mays* ssp. *mays*) has a history of profound demographic shifts accompanied by selection for agronomic performance and adaptation to novel environments, making it an ideal system in which to study the interaction between demography and selection. Maize was domesticated in a narrow region of southwest Mexico from the wild plant teosinte (*Zea mays* ssp. *parviglumis*; hereafter, *parviglumis* (Matsuoka et al. 2002; Piperno et al. 2009; van Heerwaarden et al. 2011) and experienced an associated genetic bottleneck that removed a substantial proportion of the diversity found in its progenitor (Hufford, Xu, et al. 2012; Wright et

al. 2005). Archaeological evidence suggests that after initial domestication, maize spread across the Americas, reaching the southwestern USA by approximately 4500 years before the present (BP) (Merrill et al. 2009) and coastal South America as early as 6700 years BP (Grobman et al. 2012). Gene flow into maize from multiple teosinte species has been documented in geographical regions outside of its center of origin (Hufford et al. 2013; Ross-Ibarra, Tenaillon, and Gaut 2009). To date, genetic studies of demography and selection in maize have primarily focused on initial domestication (Tenaillon et al. 2004), only broadly considering the effects of subsequent change in population size on diversity (Beissinger et al. 2016) and largely disregarding the spatial effects of geographic expansion and gene flow (but see (van Etten and Hijmans 2010)). Furthermore, the effect of maize demography on the prevalence of deleterious alleles has yet to receive in-depth attention.

Here, we investigate the genome-wide effects of demographic change in maize during domestication and subsequent expansion using high-depth resequencing data from a panel of maize landraces. We present evidence for a protracted domestication bottleneck, further loss of diversity during crop expansion, and gene flow between maize and its wild relatives outside of its center of origin. We then explore how this demographic history has shaped genome-wide patterns of deleterious alleles.

III.3 Results

Maize population size change during domestication and expansion

We resequenced 31 maize individuals, each from one open-pollinated landrace, representing six geographical regions that span the pre-Columbian range of maize cultivation

(southwestern US highlands, 6 individuals; Central Mexican Plateau, 6 individuals; Mexican lowlands, 5 individuals; Guatemalan highlands, 3 individuals; South American lowlands, 6 individuals; Andes, 5 individuals). In addition, we resequenced four wild *parviglumis* individuals from a single population located in the Balsas River Valley in Mexico (Fig. 1a). The median sequencing depth was 29X, with a range of 24–53X, resulting in a data set consisting of 49,508,640 single nucleotide polymorphisms (SNPs). Landrace accessions were selected to broadly reflect the diversity of maize in the Americas and to be representative of defined ecogeographic regions based on consultation with experts on landrace germplasm (Major Goodman, personal communication) and on descriptions in the Races of Maize handbooks (Races of maize https://www.ars.usda.gov/midwest-area/ames/plant-introduction-research/docs/races-of-maize/).

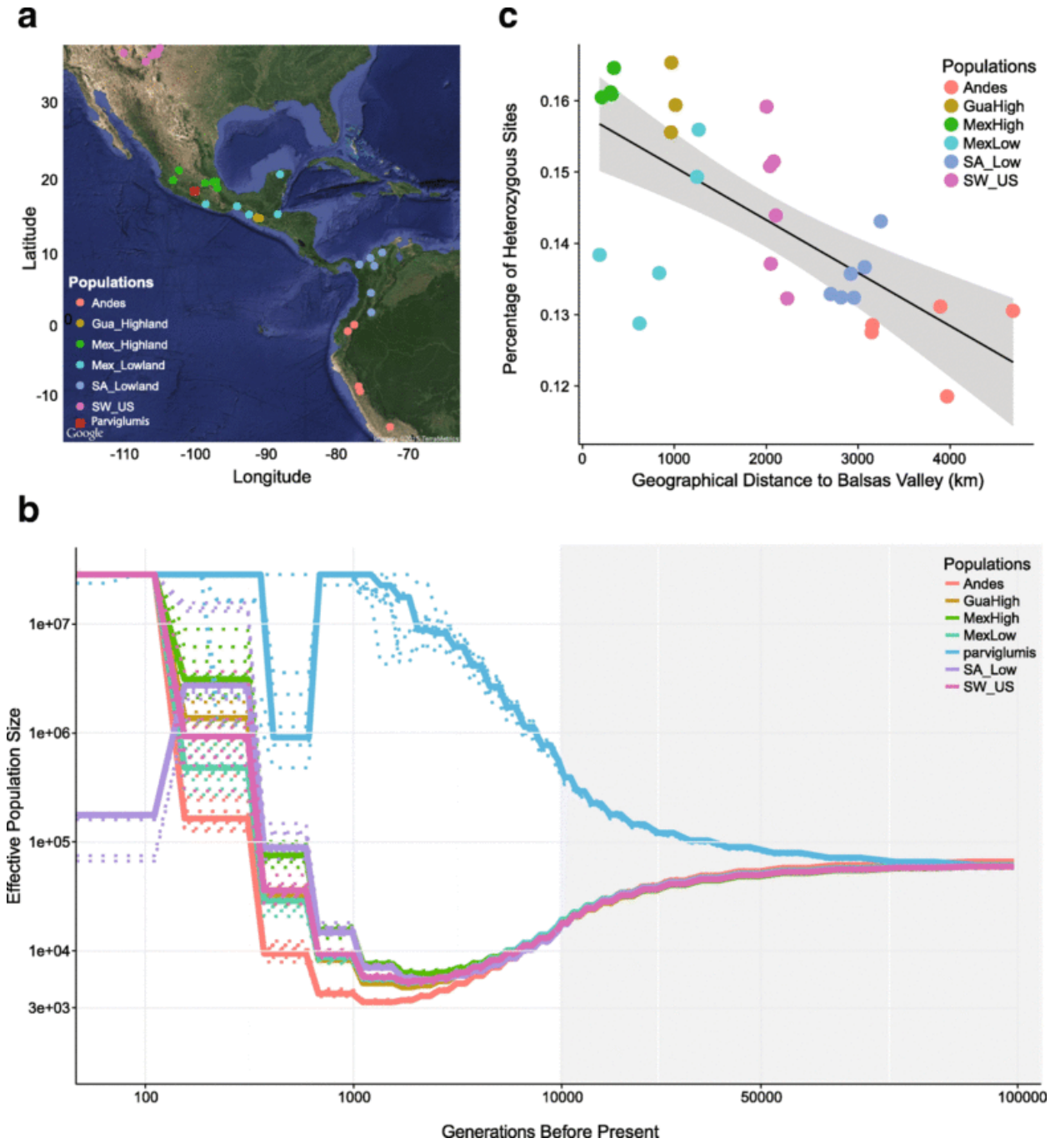


Fig. 1 Maize domestication and expansion.

a Sampling locations. **b** Estimates of effective population size over time (mutation rate =$3*10^{-8}$, generation time = 1 year). *Dashed lines* represent bootstrapping results. The *x* axis is *log*10 scaled when time is less than 10,000 generations BP and linear when greater than 10,000 generations BP as indicated by the gray

background. **c** The percentage of polymorphic sites versus distance from the maize domestication center. Abbreviations for populations: *GuaHigh* Guatemalan highlands, *MexHigh* Mexican highlands, *MexLow* Mexican lowlands, *SA_Low* South American lowlands, *SW_US* southwestern US highlands.

We first estimated historical changes in effective population size (N_e) of maize and *parviglumis* using the multiple sequentially Markovian coalescent (MSMC) (Schiffels and Durbin 2014). Consistent with archaeological evidence (Piperno et al. 2009), we find that the demographic histories of the various maize populations begin to diverge from one another approximately 10,000 years BP (Fig. 1b). Surprisingly, our single population of *parviglumis* diverges from maize much earlier, around 75,000 years BP. All maize populations show a gradual decline in diversity concomitant with divergence from *parviglumis*, but the slope becomes more pronounced around the time of domestication. This period of declining N_e continues until the recent past (≈ 1100−2400 years BP) and is followed by extremely rapid population growth, suggesting recovery from domestication post-dated expansion of maize across the Americas. In contrast to our results in maize, *parviglumis* shows an increase in N_e which also lasts until the recent past (≈ 1200−1800 years BP). To determine if linked selection associated with domestication could bias estimates of N_e in maize (see (Schrider, Shanku, and Kern 2016)), we masked previously identified domestication candidates (Hufford, Xu, et al. 2012) and observed nearly identical results (Additional file 1: Figure S1A).

One explanation for the prolonged population size reduction in maize following the onset of domestication would be repeated colonization bottlenecks during the spread of maize across the Americas. Genome-wide levels of heterozygosity across our maize samples are consistent with this idea, showing a strong negative correlation (R^2=0.3636, p=0.0004; Fig. 1c) with distance from the center of maize domestication in the Balsas River Basin. To confirm this trend, we performed

a similar analysis with a much larger sample of published genotyping data (n=3520; Additional file 1: Figure S1B) (Hearne et al. 2017) and observed similar results.

While the gradual decrease in genetic diversity seen with distance from the Balsas indicates serial founder effects, our analyses also point to a more extreme founder event in the Andean highlands of South America. Andean landraces show a deeper bottleneck in our MSMC analysis (Fig. 1b), have the lowest overall diversity (Additional file 1: Figure S2), and show both a distinct reduction of low frequency alleles and a greater proportion of derived homozygous alleles compared to other populations (Additional file 1: Figure S2). To shed light on the timing of this extreme founder event, we assessed evidence for recent inbreeding. Inbreeding coefficients in Andean samples were quite low and not statistically different from other populations (all F<0.002 and p>0.05 based on a Wilcoxon test). Likewise, no significant difference could be found across populations in the number of runs of homozygosity (ROHs) longer than 1 cM (p>0.05 in all cases, Wilcoxon test). Using simple conversions between generations and the genetic length of an inherited region in the genome (Thompson 2013), these results provide further evidence for limited recent (< 50 generations) inbreeding in the Andes. However, when ROHs were limited to those shorter than $0.05c\ M$ and longer than $0.005c\ M$ (inbreeding from approximately 1000–10,000 generations in the past), Andean samples demonstrated significantly greater cumulative ROHs compared to all (p<0.05, Wilcoxon test) but the South American lowland population (p=0.165, Wilcoxon test; Additional file 1: Figure S3). Together, these lines of evidence are consistent with an unusually strong founder event during colonization of the Andes.

Introgression from wild maize in highland populations

Adaptive introgression from the wild teosinte taxon *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) has previously been observed in maize in the highlands of Mexico (Hufford et al. 2013). Our broad sampling allowed us to investigate whether introgressed *mexicana* haplotypes have spread to highland maize populations outside of Mexico, potentially playing a role in adaptation in other regions. In order to test this hypothesis, we calculated Patterson's *D* statistic (Durand et al. 2011) across all maize populations. All individuals from both the Mexican and Guatemalan highlands exhibited highly significant evidence for shared ancestry with *mexicana* (Additional file 1: Figure S4). Maize from the southwestern USA also showed more limited evidence of introgression, consistent with findings from ancient DNA suggesting this region was originally colonized by admixed maize from the highlands of Mexico (Da Fonseca et al. 2015). In contrast, the distribution of z-scores for South American populations overlapped zero, providing no evidence for substantial spread of *mexicana* haplotypes to this region.

We localized introgression to chromosomal regions through genome-wide calculation of the $fd^{\wedge}$ statistic (Martin, Davey, and Jiggins 2015). Megabase-scale regions of introgression were identified in both Mexican and Guatemalan highland populations that correspond to those reported by (Hufford et al. 2013) on chromosomes 4 and 6 (Fig. 2; Additional file 1: Figure S5). On chromosome 3 (at around 75−90 Mb), a large, previously unidentified region of introgression can be found in the Mexican and southwestern US highlands (Fig. 2; Additional file 1: Figure S5). This region overlaps a putative chromosomal inversion associated with flowering time in maize landraces (Romero Navarro et al. 2017) and in the maize nested association mapping population (Buckler et al. 2009) and may be an example of *mexicana* contribution to modern maize lines.

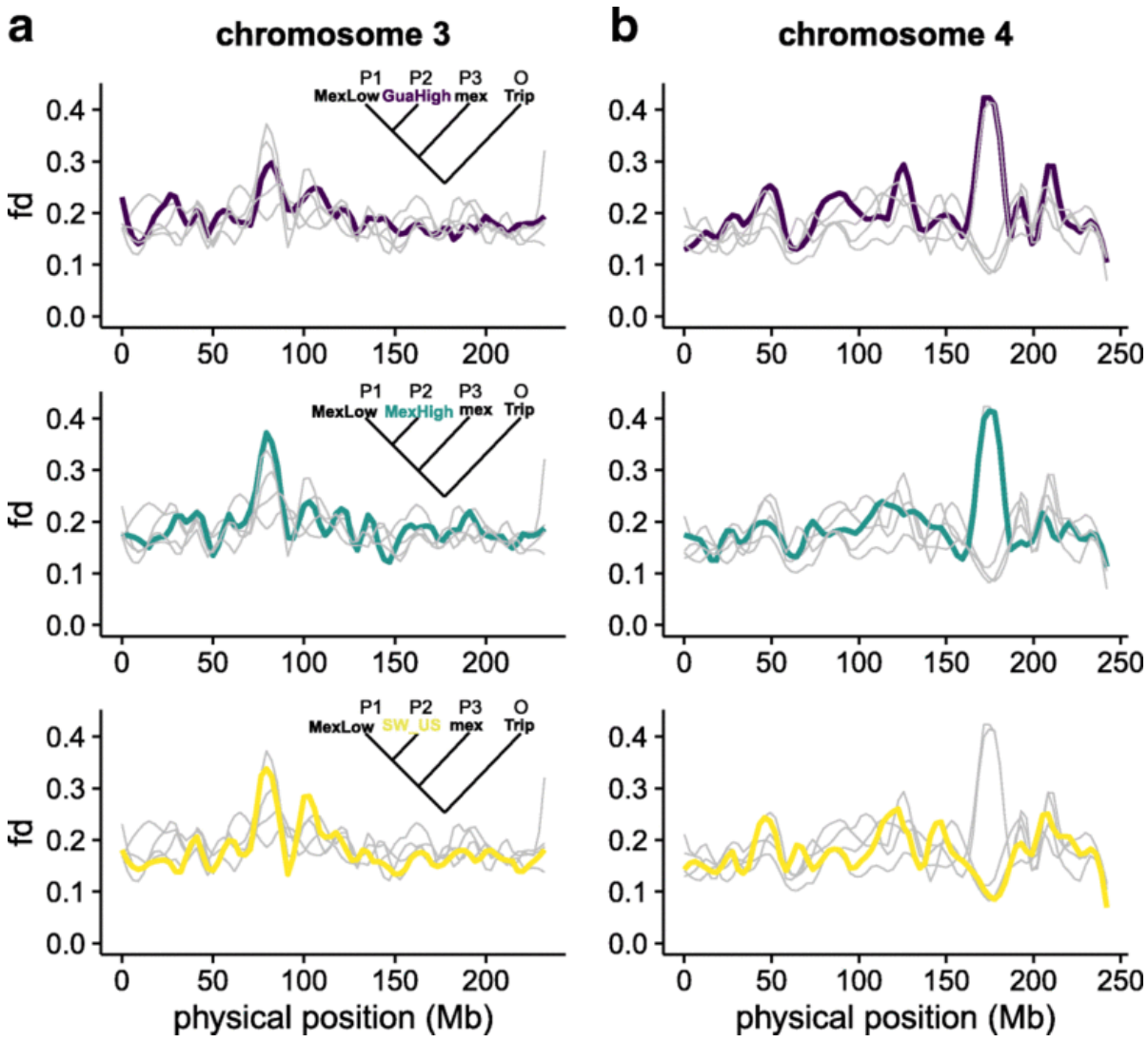


Fig. 2 Introgression from *mexicana* into maize landraces. Loess regression of *fd*^ is plotted for all five populations on **a** chromosome 3 and **b** chromosome 4. Each plot highlights a single population, with other populations shown in *gray*. The Mexican lowlands population is used as a reference and thus not plotted. No significant introgression was detected in the South American lowlands or the Andes, and loess regressions for these populations are only shown as *gray lines*. The statistic *fd*^ was calculated based on the tree in which P2 is varied across populations. *mex mexicana*, *Trip Tripsacum*

The influence of demography on accumulation of deleterious alleles

Population-specific changes in historical N_e should influence the efficiency of purifying selection and alter genome-wide patterns of deleterious variants (Fu et al. 2014). Introgression from a species with substantially different N_e may also influence the abundance and distribution of deleterious alleles in the genome (Harris and Nielsen 2016; Juric, Aeschbacher, and Coop 2016). Below we evaluate the effects of major demographic events during the pre-Columbian history of maize on patterns of deleterious alleles.

Domestication and deleterious alleles

We first compared counts of deleterious alleles in Mexican lowland maize individuals to four *parviglumis* individuals from a single population in the Balsas River Valley. Maize from the Mexican lowlands has not experienced substantial introgression from wild relatives and is near the center of maize origin (van Heerwaarden et al. 2011), and thus best reflects the effects of domestication alone. After identifying putatively deleterious mutations using Genomic Evolutionary Rate Profiling (GERP) (Cooper et al. 2005), we calculated the number of derived deleterious alleles per genome under both an additive and a recessive model across four levels of mutation severity (see Methods for details). Maize showed significantly more deleterious alleles than teosinte under both additive (<1 0% more; p=0.0079, Wilcoxon test; Additional file 1: Figure S6) and recessive (< 20−30% more; p=0.0079; Fig. 3) models across all categories (Additional file 1: Figure S7). Additionally, maize contained more than twice as many fixed deleterious alleles than teosinte (57,881 versus 26,947) and 10% fewer segregating deleterious alleles (429,837 versus 478,594), effects expected under a domestication bottleneck (Fig. 3c; (Simons et al. 2014)). GERP load (GERP score × frequency of deleterious alleles), a more direct proxy of mutation load quantified at the population level, revealed a similar trend (additive model: maize median =23.635,

teosinte median =22.791, p=0.008, Wilcoxon test; recessive model: maize median =14.922, teosinte median =12.231, p=0.008). Maize, like other domesticates (Marsden et al. 2016; Liu et al. 2017; Renaut and Rieseberg 2015; Günther and Schmid 2010), thus appears to have a higher mutation load compared to its wild progenitor *parviglumis*.

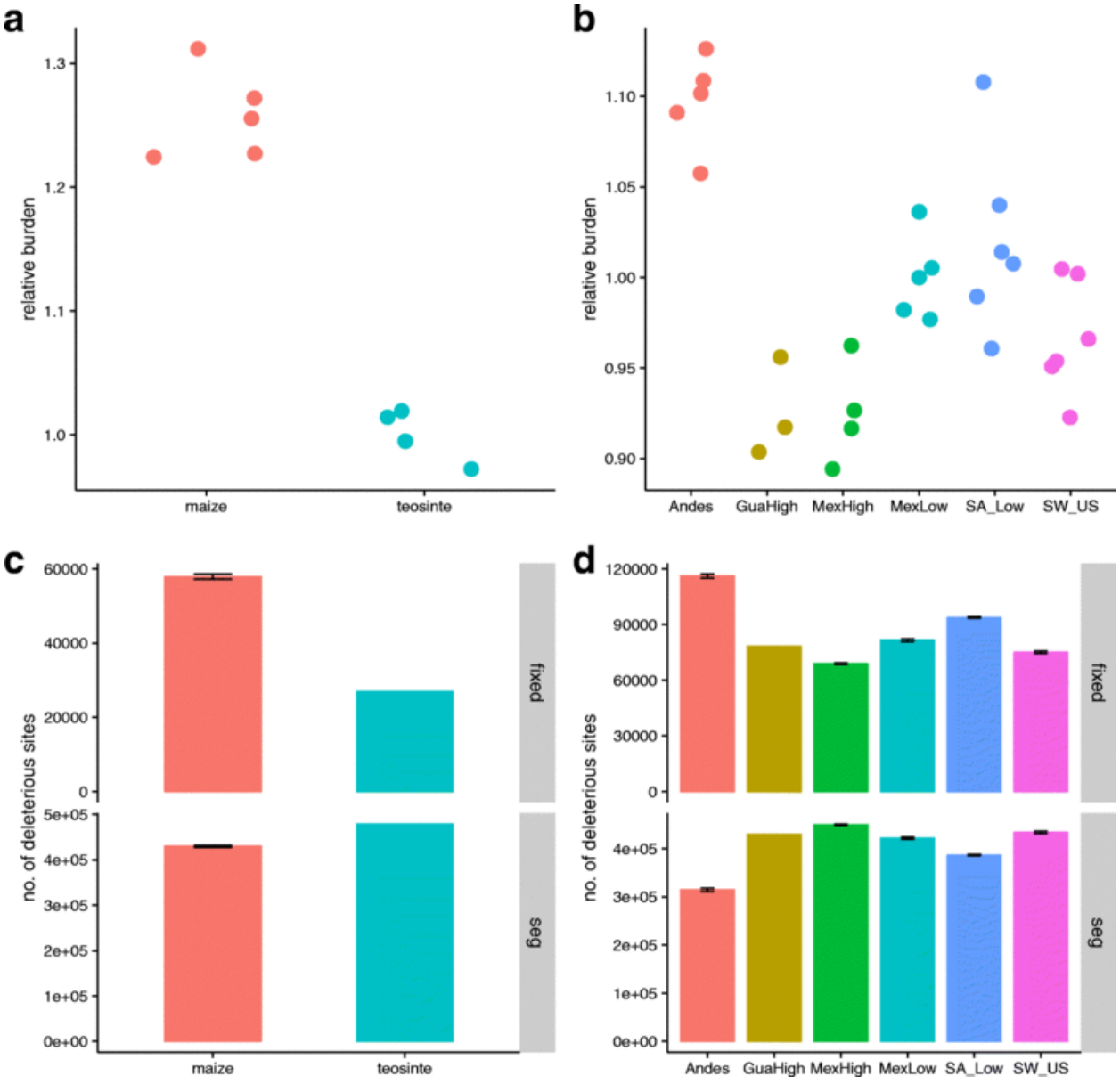


Fig. 3 Burden of deleterious mutations during maize domestication and expansion.

Comparison of counts of deleterious alleles at the individual level **a** between *parviglumis* and maize (mean value in *parviglumis* population was used as the standard to calculate the relative burden) and **b** among

maize populations (mean value in Mexican lowland population was utilized as the standard to calculate the relative burden) under a recessive model. Comparison of fixed versus segregating (seg) deleterious alleles at the population level **c** between *parviglumis* and maize and **d** among maize populations. A jackknife sub-sampling approach ($n = 4$) was utilized for maize in **c** and for individual maize populations ($n = 3$) in **d**

While the elevated mutation load we observe in maize relative to *parviglumis* may be driven primarily by the domestication bottleneck, positive selection on causal variants underlying domestication phenotypes may also fix nearby deleterious variants through genetic hitchhiking, which would result in a higher number of deleterious alleles in regions linked to domestication loci (Kono et al. 2016; Renaut and Rieseberg 2015). To test this hypothesis, we first confirmed that 420 previously identified domestication candidates (Hufford, Xu, et al. 2012) showed evidence of selection in our data (Additional file 1: Figure S8), and then assessed the distribution of deleterious alleles in and near (5 kb upstream and downstream) these genes by calculating the number of deleterious alleles per base pair under both recessive and additive models. No significant difference was found in the prevalence of deleterious alleles near domestication and random sets of genes (Additional file 1: Figure S9), suggesting the increased mutation load we observe in maize has been driven primarily by the genome-wide effects of the domestication bottleneck rather than linkage associated with selection on specific genes.

The effect of the Andean founder event on deleterious alleles

The extreme founder event observed in the Andes could potentially alter genome-wide patterns of deleterious variants beyond the effects of domestication alone. Under a recessive model, maize from the Andes contains significantly more deleterious alleles than any other population (Fig. 3b; Additional file 1: Figure S7; all p values <0.02, Wilcoxon test), and this difference becomes more extreme when considering more severe (i.e., higher GERP score) mutations (Additional file 1: Figure S7). In contrast, we observe no significant difference under

an additive model (Additional file 1: Figure S6; Additional file 1: Figure S7). The Andean founder event therefore appears to have resulted in higher mutation load than seen in other maize populations. This result is further supported by a higher proportion of fixed deleterious alleles within the Andes and fewer segregating deleterious alleles (Additional file 1: Figure S10; Fig. 3d), a result comparable to the differences observed between maize and *parviglumis*.

Introgression decreases the prevalence of deleterious alleles

Highly variable rates of *mexicana* introgression were detected across our landrace populations (Fig. 2; Additional file 1: Figure S4; Additional file 1: Figure S5). To explore the potential effects of introgression on the genomic distribution of deleterious alleles, we fit a linear model in which the number of deleterious sites is predicted by introgression (represented by $f_{d}\hat{}$) and gene density (exonic base pairs per centimorgan) in 10-kb non-overlapping windows in the Mexican highland population where we found the strongest evidence for *mexicana* introgression. Gene density was included as a predictor in the regression to control for the positive correlation observed between gene density and both introgression (p=3.48e−08) and deleterious alleles (p≈0). When identifying deleterious alleles under both additive and recessive models, we found a strong negative correlation with introgression (i.e., fewer deleterious alleles in introgressed regions; p≈0 under both models). These findings likely reflect the larger ancestral N_e and more efficient purifying selection in *mexicana*.

III.4 Discussion

Demographic studies in domesticated species have focused largely on identifying progenitor population(s) and quantifying the effect of the domestication bottleneck on genetic

diversity (Hufford, Xu, et al. 2012; Lam et al. 2010; Zhu et al. 2007). It is likely, however, that the demographic history of domesticates is generally more complex than a simple bottleneck followed by recovery (Meyer et al. 2016; Zhou et al. 2017). Many crops and domesticated animals have expanded from defined centers of origin to global distributions, experiencing population size changes and gene flow from closely related taxa throughout their histories (Gaut, Díez, and Morrell 2015). With this in mind, we have characterized maize demography from domestication through initial expansion in order to provide a more complete assessment of the influence of demography on deleterious variants.

Historical changes in maize population size

Early models of maize demography suggested the ratio of the domestication bottleneck size and duration was between ≈ 2.5:1 and ≈ 5:1, but little statistical support was found for specific estimates of these individual parameters (Wright et al. 2005; Tenaillon et al. 2004; Eyre-Walker et al. 1998). Most recently, Beissinger et al. (Beissinger et al. 2016) fit a model assuming a bottleneck followed by instantaneous exponential recovery. While our results concur with the most recent model in finding a similar bottleneck size (≈ 10% compared to ≈ 5% in Beissinger et al.) and that the modern N_e of maize is quite large, the flexibility of MSMC also allowed us to estimate the duration of the bottleneck. We find that the domestication bottleneck may have lasted much longer than previously believed, spanning ≈ 9000 generations and only beginning to recover in the recent past (Fig. 1b). Analysis of bottlenecks during African rice and grape domestication have also suggested a duration of several thousand generations (Meyer et al. 2016; Zhou et al. 2017), indicating that demographic bottlenecks during crop evolution may have generally occurred over substantial periods of time. Previous work has suggested population structure can generate

spurious signals of population size change in methods like MSMC (Mazet et al. 2016; Nielsen and Beaumont 2009), such that individuals sampled from a single deme of a highly structured population can falsely demonstrate signatures of a population bottleneck similar to what we observe in maize (Mazet et al. 2016). Given that our maize landraces are sampled from broad ecogeographic regions, however, this effect should be minimal. Moreover, a similar analysis in an Americas-wide sample of maize landraces demonstrated qualitatively similar results (Beissinger et al. 2016).

In addition to a strong bottleneck during domestication, our finding that levels of diversity decline in populations increasingly distant from the center of maize domestication are suggestive of serial founder effects during the spread of maize across the Americas (Fig. 1c; Additional file 1: Figure S1). Serial founder effects are the result of multiple sampling events during which small founder populations are repeatedly drawn from ancestral pools, leading to a stepwise increase in genetic drift and a concomitant decrease in genetic diversity. During maize range expansion, serial founder effects would have occurred if seed carried to each successive colonized region was limited to a sample of whole ears that contained a fraction of the diversity found in the source population (van Etten and Hijmans 2010). Movement of entire ears involves a collective transfer of seeds that are either full or maternal half-siblings and could lead to more substantial founder effects than would be seen if dispersal were truly random. Such "kin-structured" migration, which is common in nature, has theoretically been demonstrated to increase inbreeding due to a reduction in the number of effective colonists (Whitlock and McCauley 1990). Consistent with serial founder effects, other researchers have found a correlation between geographic and genetic distance in maize landraces (van Heerwaarden et al. 2011; Vigouroux et al. 2008), though this was previously attributed to limited migration across the species range leading to isolation by distance (IBD).

Neutral expectations of allele frequencies across populations under serial founder effects differ substantially from those predicted under equilibrium conditions (Slatkin and Excoffier 2012). For example, Slatkin and Excoffier (Slatkin and Excoffier 2012) have demonstrated that allele frequency clines previously attributed to adaptation could be generated entirely by neutral processes under expansion. Many of the world's crops have experienced such histories of expansion, and studies attempting to identify loci underlying crop adaptation during post-domestication spread to new environments may most accurately compare empirical data to neutral expectations under a serial founder effects demography (Slatkin and Excoffier 2012).

While a history of serial founder effects partially explains the variation in diversity across maize landraces, there are deviations from this model. For example, our combined results showing increased ROHs (Additional file 1: Figure S3), lower nucleotide diversity (Additional file 1: Figure S2), and smaller effective population size (Fig. 1) in the Andes all suggest a pronounced, ancient founder event and are in agreement with previous work modeling demography in this region (Takuno et al. 2015). The founder event in the Andes may reflect initially limited cultivation due to the poor performance of maize in this region relative to established root and tuber staples (Pearsall 2008); maize cultivation may have only become widespread after an initial lag period necessary for adaptation. Additionally, we observe somewhat higher than expected nucleotide diversity in maize landraces from the highlands of Mexico and Guatemala (Fig. 1c), which may be linked to the introgression we have detected from *mexicana*.

In striking contrast to the bottleneck we observe in maize, the effective population size in *parviglumis* increases steadily from the time of initial maize domestication until the recent past. Multiple population genetic studies have reported negative genome-wide values of Tajima's D in *parviglumis* from the Balsas River Valley (Beissinger et al. 2016; Ross-Ibarra, Tenaillon, and Gaut

2009; Wright et al. 2005) findings characteristic of an expanding population. Likewise, analyses of pollen content in sediment cores from Mexico suggest herbaceous vegetation and grasslands have expanded over the last 10,000 years due to changing environmental conditions during the Holocene and human management of vegetation with fire (Piperno et al. 2007; Correa-Metrio et al. 2012). While our *parviglumis* samples are drawn from a single population in the Balsas, these data collectively suggest *parviglumis* from this region has experienced expansion over the last several millennia.

Consistent with archaeological evidence of the timing of initial maize domestication (Piperno et al. 2009), we find that maize demographies begin to diverge ≈ 10,000 generations BP, a time that appears to coincide with a steeper decline in maize N_e as well. In contrast, we estimate the timing of the split between maize and our single population of *parviglumis* to be ≈ 75,000 generations BP, potentially reflecting population structure in *parviglumis*. Beissinger et al. (Beissinger et al. 2016), using samples from additional populations, also find an estimate of maize-*parviglumis* divergence older than the probable onset of domestication, suggesting that currently available sequences of *parviglumis* may not sample well from the populations directly ancestral to domesticated maize.

The prevalence of gene flow during maize diffusion

Increasingly, range-wide analyses of crops and their wild relatives have identified evidence of gene flow during post-domestication expansion from newly sympatric populations of their progenitor taxa and closely related species (Bredeson et al. 2016; Miao, Wang, and Li 2017; Poets et al. 2015). Consistent with previous results from genotyping data (Hufford et al. 2013; van Heerwaarden et al. 2011; Doebley, Goodman, and Stuber 1987), we find strong support for

introgression from *mexicana* to maize in the highlands of Mexico. While *mexicana* is not currently found in the highlands of Guatemala, we also find strong evidence for *mexicana* introgression in maize from this region, suggesting either *mexicana* was at one time more broadly distributed, or, perhaps more likely, that highland maize from Mexico was introduced to the Guatemalan highlands. Support is also found for *mexicana* introgression in the southwestern USA at specific chromosomal regions such as a putative inversion polymorphism on chromosome 3 (Fig. 2). These results confirm previous findings suggesting maize from the highlands of Mexico originally colonized the southwestern USA (Da Fonseca et al. 2015). The more limited signal of *mexicana* introgression here may be due to subsequent gene flow from lowland maize as suggested by (Da Fonseca et al. 2015). Very little evidence is found for *mexicana* haplotypes extending into South America, as highland-adapted haplotypes would likely have been maladaptive and removed by selection as maize traversed the lowland regions of Central America (Takuno et al. 2015).

Impacts of demography on accumulation of deleterious variants

Previous work in maize has characterized genome-wide trends in deleterious alleles across modern inbred maize lines, revealing that inbreeding during the formation of modern lines has likely purged many recessive deleterious variants (Yang et al. 2017) and that complementation of deleterious alleles likely underlies the heterosis observed in hybrid breeding programs (Yang et al. 2017; Gerke et al. 2015). Additionally, (Beissinger et al. 2016) revealed that purifying selection has removed a greater extent of pairwise diversity (θ_π) near genes in *parviglumis* than in maize due to the higher historical N_e in *parviglumis*, but that this trend is reversed when considering younger alleles due to the recent dramatic expansion in maize population size. To date, however, few links have been made between the historical demography of maize domestication and

expansion and the prevalence of deleterious alleles. Our analysis reveals, for the first time, that demography has played a pivotal role in determining both the geographic and genomic landscapes of deleterious alleles in maize.

Population size and deleterious variants

Previous studies have suggested a "cost of domestication" in which a higher burden of deleterious alleles is found in domesticates compared to their wild progenitors (Kono et al. 2016; Lu et al. 2006; Marsden et al. 2016; Renaut and Rieseberg 2015; Schubert et al. 2014). Consistent with these results, we detect an excess of deleterious alleles in maize relative to *parviglumis* (Fig. 3; Additional file 1: Figure S6; Additional file 1: Figure S7), which could be caused by two potential factors. First, reduced population size during the domestication bottleneck could result in deleterious alleles drifting to higher allele frequency. Second, hitchhiking caused by strong positive selection on domestication genes could cause linked deleterious alleles to rise in frequency (Lu et al. 2006; Marsden et al. 2016). While we find support for the former in maize, we see little evidence of the latter. Recent studies have reported contrasting results regarding the effect of selective sweeps in patterning the distribution of deleterious alleles. For example, putative selective sweeps in cassava showed a paucity of deleterious alleles, a result that was attributed to purifying selection (Ramu et al. 2017). Sweep regions in grape exhibited an overall decrease in the number of deleterious alleles but an increase in the ratio of deleterious mutations to synonymous variants, a pattern suggesting deleterious alleles may have hitchhiked along with the targets of positive directional selection (Zhou et al. 2017). Finally, selective sweeps in Asian rice contained a roughly equivalent ratio of deleterious mutations to synonymous mutations when compared to neutral regions (Liu et al. 2017). Clearly, further exploration is warranted to clarify

the effect of selection on the distribution of deleterious mutations. In addition to the cost of domestication, we find a cost of geographic expansion that is likely driven by serial founder effects. The increase in deleterious alleles during expansion is most pronounced in the Andes and may be symptomatic of the extreme founder event we propose above.

Differences in the number of deleterious alleles between maize and *parviglumis* and non-Andean and Andean maize are more dramatic under a recessive model than an additive model. This trend may indicate that the bulk of deleterious alleles in maize are at least partially recessive, such that heterozygous sites contribute less to a reduction in individual fitness. Previous work in human populations has shown that the majority of deleterious mutations are recessive or partially recessive (McQuillan et al. 2012), and data from knock-out mutations in yeast have revealed that large-effect mutations tend to be more recessive (Agrawal and Whitlock 2011). Likewise, both theory and empirical evaluation across a number of organisms suggest that mildly deleterious mutations are likely to be partially recessive (Manna, Martin, and Lenormand 2011). In maize, Yang et al. (Yang et al. 2017) have found that most deleterious alleles are at least partially recessive and note a negative correlation between the severity of a deleterious variant and its dominance. Our results thus match nicely both with previous empirical data and theoretical expectations showing that recent population bottlenecks should only show strong differences in load under a recessive model (Simons et al. 2014).

Introgression and deleterious variants

Very few studies have investigated the effects of introgression from a taxon with substantially different N_e on the genomic landscape of deleterious variants. The best example is found in the human literature, where confirmation has been found that introgression from

Neanderthals with low ancestral N_e increased the overall mutation load in modern humans (Harris and Nielsen 2016; Juric, Aeschbacher, and Coop 2016). We report here the opposite pattern in maize, as introgression appears to have reduced the proportion of deleterious variants. Nonetheless, the underlying interpretation is similar: the taxon donating alleles *mexicana* has had a larger ancestral N_e than maize (Ross-Ibarra, Tenaillon, and Gaut 2009), and introgressed haplotypes have thus experienced more efficient long-term purging of deleterious alleles.

III.5 Conclusion

We have demonstrated that demography during the domestication and expansion of maize across the Americas has profoundly influenced putative functional variation across populations and within individual genomes. More generally, we have learned that population size changes and gene flow from close relatives with contrasting effective population size will influence the distribution of deleterious alleles in species undergoing rapid shifts in demography. The significance of our results extends far beyond maize. For example, invasive species that have recently experienced founder events followed by expansion, endangered species subject to precipitous declines in N_e, species with a history of post-glacial expansion, and new species expanding their range will all likely show clear genetic signals of the interplay between demography and selection. This interaction bears importantly on the adaptive potential of both individual populations and species. By fully characterizing this relationship, we can better understand how the current evolutionary trajectory of a species has been influenced by its history.

III.6 Methods

Samples, whole genome resequencing, and read mapping

A total of 31 maize landrace accessions were obtained from the US Department of Agriculture (USDA)'s National Plant Germplasm System and through collaborators (Additional file 2: Table S1). Samples were chosen from four highland populations (Andes, Mexican highlands, Guatemalan highlands, and southwestern US highlands) and two lowland populations (Mexican and South American lowlands) (Fig. 1a). In addition, four open-pollinated *parviglumis* samples were selected from a single population in the Balsas River Valley in Mexico. DNA was extracted from leaves using a standard cetyltrimethyl ammonium bromide (CTAB) protocol (Doyle and Doyle 1987). Library preparation and Illumina HiSeq 2000 sequencing (100-bp paired-end) were conducted by BGI (Shenzhen, China) following their established protocols. the Burrows-Wheeler Aligner (BWA) v.0.7.5.a (Li and Durbin 2010) was used to map reads to the maize B73 reference genome v3 (GenBank BioProject PRJNA72137) (Schnable et al. 2009) with default settings. The duplicate molecules in the realigned bam files were removed with MarkDuplicates in Picardtools v.1.106 (http://broadinstitute.github.io/picard), and indels were realigned with the Genome Analysis Toolkit (GATK) v.3.3-0 (Depristo et al. 2011). Sites with mapping quality less than 30 and base quality less than 20 were removed, and only uniquely mapped reads were included in downstream analyses.

Demography of maize domestication and diffusion

The MSMC method (Schiffels and Durbin 2014), which models ancestral relationships under recombination and mutation and has been used in several plant species (Meyer et al. 2016; Zhou et al. 2017), was utilized to infer effective population size changes in both *parviglumis* and

maize. SNPs were called via HaplotypeCaller and filtered via VariantFiltration in GATK (Depristo et al. 2011) across all samples. SNPs with the following metrics were excluded from the analysis: QD <2.0; FS >60.0; MQ <40.0; MQRankSum <−12.5; ReadPosRankSum <−8.0. Vcftools v.0.1.12 (Danecek et al. 2011) was used to further filter SNPs to include only bi-allelic sites. Following these data filtering steps, our data set consisted of 49 million SNPs. SNPs were phased using BEAGLE v.4.0 (Browning and Browning 2007) with SNP data from the maize HapMap2 panel (Chia et al. 2012) used as a reference. Only sites with depth between half and twice of the mean depth were included in analyses. In addition, the software SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml) was used to mask genomic regions in which reads were not uniquely mapped. The mappability mask file for MSMC was generated by stepping in 1-bp increments across the maize genome to generate 100-bp single-end reads, which were then mapped back to the maize B73 reference genome (Schnable et al. 2009). Sites with the majority of overlapping 100-mers mapped uniquely without mismatch were determined to be "SNPable" sites and used for the MSMC analyses. For effective population size inference in MSMC, we used 5×4+25×2+5×4 as the pattern parameter, and the value *m* was set as half of the heterozygosity in *parviglumis* and maize populations, respectively.

In order to explore the trend of genetic diversity away from the domestication center, the correlation between the percentage of polymorphic sites and the geographic distance to the Balsas River Valley (latitude 18.099138, longitude –100.243303) was examined via linear regression. Geographical distance in kilometers was calculated based on great circle distance using the haversine transformation (Ramachandran et al. 2005). The correlation between percentage of heterozygous sites and distance away from domestication center was also explored in the SeeDs

data set. SNPs with more than 50% missing samples and samples with more than 50% missing genotypes were removed from the SeeDs data set.

Population structure, genetic diversity, and inbreeding coefficients

We first evaluated population structure using principal component analysis (PCA) with ngsCovar (Fumagalli et al. 2013) in ngsTools (Fumagalli et al. 2014) based on the matrix of posterior probabilities of SNP genotypes produced in Analysis of Next Generation Sequencing Data (ANGSD) v.0.614 (Korneliussen, Albrechtsen, and Nielsen 2014), and then utilized NGSadmix v.32 (Skotte, Korneliussen, and Albrechtsen 2013) to investigate the admixture proportion of each accession. The NGSadmix analysis was conducted based on genotype likelihoods for all individuals, which were generated with ANGSD (options -GL 2 -doGlf 2 -SNP_pval 1e−6), and K from 2 to 10 was set to run the analysis for sites present in a minimum of 77% of all individuals (24 in 31). A clear outlier in the Mexican highland population was detected, RIMMA0677, a sample from relatively low altitude, which was suspected to contain a divergent haplotype. A neighbor-joining tree of SNPs within an inversion polymorphism on chromosome 4 that includes a diagnostic highland haplotype (Hufford et al. 2013) was constructed with the R package phangorn (Schliep 2011). The sample RIMMA0677 was not clustered with other highland samples, but embedded within lowland haplotypes (Additional file 1: Figure S11), so it was removed from further analyses.

The genetic diversity measures Watterson's θ and θ_π were calculated in ANGSD (Korneliussen, Albrechtsen, and Nielsen 2014) for each population. The neutrality test statistic Tajima's D was calculated with an empirical Bayes approach (Korneliussen et al. 2013) implemented in ANGSD by first estimating a global site frequency spectrum (SFS) then

calculating posterior sample allele frequencies using the global SFS as a prior. The three statistics were summarized across the genome using 10-kb non-overlapping sliding windows.

Inbreeding coefficients for each individual were estimated with ngsF (Vieira et al. 2013) with initial values of F_{IS} set to be uniform at 0.01 with an epsilon value of $1e{-}5$.

In addition, SNPs were polarized using the *Tripsacum dactyloides* genome to assess the frequency of derived homozygous sites in each maize landrace population. *T. dactyloides* short reads were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (SRR447804–SRR447807), mapped to the B73 reference genome v3 (Schnable et al. 2009) with BWA (Li and Durbin 2010), and incorporated into SNP calling as described above.

Runs of homozygosity

SNPs were down-sampled to contain one SNP in a 2-kb window to identify segments representing homozygosity by descent (i.e., autozygosity) rather than by chance. PLINK v.1.07 (Purcell et al. 2007) was applied to identify segments of ROHs in a window containing 20 SNPs, among which the number of the maximum missing SNPs was set to 2 and the number of the maximum heterozygous sites was set to 1. The shortest length of final ROHs was set to be 300 kb. Physical distances were converted into genetic distances based on a recent genetic map (Ogut et al. 2015).

Detection of introgression

To assess per-genome evidence of population admixture between maize landraces and teosinte, we calculated the *D* statistic using ANGSD (Korneliussen, Albrechtsen, and Nielsen

2014). The statistic was calculated using trees of the form (((X, low),*mexicana*),*T. dactyloides*). One accession from the Mexican lowland population was randomly sampled as the "low" taxon, and each sample from all other populations except the Mexican lowland was set as "X". The *mexicana* accession TIL25 from the maize HapMap2 project (Chia et al. 2012) was treated as the third column species. The *D* statistic was calculated in a 1-kb block, and then jackknife bootstrapping was conducted to estimate significance.

In addition, the f_d^ statistic (Martin, Davey, and Jiggins 2015) was calculated based on a similar tree form (((*P*1,*P*2),*P*3),*O*), but using allele frequencies across multiple individuals for each position on the tree. We fixed P_1 as the Mexican lowland population, P_3 as two lines of *mexicana* (TIL08 and TIL25), and *T. dactyloides* as the outgroup. P_2 was set to each of the four highland populations and the South American lowland population.

The f_d^ statistic was calculated in 10-kb non-overlapping windows across the genome with the python script egglib_sliding_windows.py (https://github.com/johnomics/Martin_Davey_Jiggins_evaluating_introgression_statistics), which makes use of the EggLib library (De Mita and Siol 2012). The input file was generated by first identifying genotypes using ANGSD (-doMajorMinor 1 -doMaf 1 -GL 2 -doGeno 4 -doPost 1 -postCutoff 0.95 -SNP_pval 1e−6) followed by format adjustments with a custom script (see "Availability of data and materials"). Outliers were detected by setting the 95% quantile of the f_d^ distribution in the South American lowland population as the cutoff.

Estimating burden of deleterious mutations

We estimated the individual burden of deleterious alleles based on GERP scores (Davydov et al. 2010) for each site in the maize genome, which reflects the strength of purifying selection

based on constraint in a whole genome alignment of 13 plant species (Rodgers-Melnick et al. 2015). The alignment and estimated GERP scores are available at iplant (https://doi.org/10.7946/P2WS60). Scores above 0 may be interpreted as historically subject to purifying selection, and mutations at such sites are likely deleterious. We identified *Sorghum bicolor* alleles in the multiple species alignment as ancestral and defined the non-*Sorghum* allele as the derived allele. Only biallelic sites were included for our evaluation. Inclusion of the maize B73 reference genome when calculating GERP scores (Rodgers-Melnick et al. 2015) introduces a bias toward underestimation of the burden of deleterious alleles in maize versus teosinte populations. Therefore, we corrected the GERP scores of sites where the B73 allele is derived following (Simons et al. 2014). Briefly, we divided SNPs where the B73 allele is ancestral into bins of 1% derived allele frequency based on maize HapMap3 (Bukowski et al. 2015) and used this frequency distribution to estimate the posterior probability of GERP scores for SNPs where the B73 allele is derived.

The sum of GERP scores multiplied by deleterious allele frequency for each SNP site was used as a proxy of individual burden of deleterious alleles under an additive model ($HET*0.5+HOM*1$). This burden was calculated under a recessive model as the sum of GERP scores multiplied by one for each deleterious homozygous site ($HOM*1$). For a better understanding of the variation of individual burden among sites under varied selection strength, we partitioned the deleterious SNPs into four categories ($-2<$ GERP ≤ 0, nearly neutral; $0<$ GERP ≤ 2, slightly deleterious; $2<$ GERP ≤ 4, moderately deleterious; GERP >4, strongly deleterious) and recapitulated the preceding statistics.

III.7 Acknowledgements and funding

We thank Dr. Andrew Severin and Dr. Arun Seetharam for bioinformatic support and two anonymous reviewers for their helpful suggestions.

This study was supported by the US Department of Agriculture (USDA #2009-65300-05668), the USDA Agricultural Research Service, the National Science Foundation (NSF IOS #1546719), USDA Hatch project (CA-D-PLS-2066-H), and startup funds from Iowa State University.

III.8 References

Agrawal, A. F., and M. C. Whitlock. 2011. "Inferences about the Distribution of Dominance Drawn from Yeast Gene Knockout Data." *Genetics* 187 (2):553–66. https://doi.org/10.1534/genetics.110.124560.

Austerlitz, F., B. Jung-Muller, B. Godelle, and P-H Gouyon. 1997. "Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization." *Theoretical Population Biology* 51 (2):148–64. https://doi.org/10.1006/tpbi.1997.1302.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, and J. Ross-Ibarra. 2016. "Recent Demography Drives Changes in Linked Selection across the Maize Genome." *Nature Plants* 2 (7):16084. https://doi.org/10.1038/nplants.2016.84.

Bredeson, J. V., J. B. Lyons, S. E. Prochnik, G. A. Wu, C. M. Ha, E. Edsinger-Gonzales, J. Grimwood, et al. 2016. "Sequencing Wild and Cultivated Cassava and Related Species Reveals Extensive Interspecific Hybridization and Genetic Diversity." *Nature Biotechnology* 34 (5):562–70. https://doi.org/10.1038/nbt.3535.

Browning, S. R., and B. L. Browning. 2007. "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering." *The American Journal of Human Genetics* 81 (5):1084–97. https://doi.org/10.1086/521987.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, et

al. 2009. “The Genetic Architecture of Maize Flowering Time.” *Science* 325 (5941):714–18. https://doi.org/10.1126/science.1174276.

Bukowski, R., X. Guo, Y. Lu, C. Zou, B. He, Z. Rong, B. Wang, et al. 2015. “Construction of the Third Generation Zea Mays Haplotype Map.” *bioRxiv*, 26963. https://doi.org/10.1101/026963.

Chia, J. M., C. Song, P. J. Bradbury, D. Costich, N. De Leon, J. Doebley, R. J. Elshire, et al. 2012. “Maize HapMap2 Identifies Extant Variation from a Genome in Flux.” *Nature Genetics* 44 (7):803–7. https://doi.org/10.1038/ng.2313.

Cooper, G. M., E. A. Stone, G. Asimenos, E. D. Green, S. Batzoglou, and A. Sidow. 2005. “Distribution and Intensity of Constraint in Mammalian Genomic Sequence.” *Genome Research* 15 (7):901–13. https://doi.org/10.1101/gr.3577405.

Correa-Metrio, A., S. Lozano-García, S. Xelhuantzi-López, S. Sosa-Nájera, and S. E. Metcalfe. 2012. “Vegetation in Western Central Mexico during the Last 50 000 Years: Modern Analogs and Climate in the Zacapu Basin.” *Journal of Quaternary Science* 27 (5):509–18. https://doi.org/10.1002/jqs.2540.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15):2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, and S. Batzoglou. 2010. “Identifying a High Fraction of the Human Genome to Be under Selective Constraint Using GERP++.” *PLoS Computational Biology* 6 (12). https://doi.org/10.1371/journal.pcbi.1001025.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data.” *Nature Genetics* 43 (5):491–501. https://doi.org/10.1038/ng.806.

Do, R., D. Balick, H. Li, I. Adzhubei, S. Sunyaev, and D. Reich. 2015. “No Evidence That Selection Has Been Less Effective at Removing Deleterious Mutations in Europeans than in Africans.” *Nature Genetics* 47 (2):126–31. https://doi.org/10.1038/ng.3186.

Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. “The Molecular Genetics of Crop Domestication.” *Cell*. https://doi.org/10.1016/j.cell.2006.12.006.

Doebley, J., M. M. Goodman, and C. W. Stuber. 1987. “Patterns of Isozyme Variation between Maize and Mexican Annual Teosinte.” *Economic Botany* 41 (2):234–46. https://doi.org/10.1007/BF02858971.

Doyle, J. J., and J. L. Doyle. 1987. “A Rapid DNA Isolation Procedure for Small Quantities of

Fresh Leaf Tissue." *Phytochemical Bulletin* 19:11–15. https://doi.org/10.2307/4119796.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. "Testing for Ancient Admixture between Closely Related Populations." *Molecular Biology and Evolution* 28 (8):2239–52. https://doi.org/10.1093/molbev/msr048.

Etten, J. van, and R. J. Hijmans. 2010. "A Geospatial Modelling Approach Integrating Archaeobotany and Genetics to Trace the Origin and Dispersal of Domesticated Plants." *PLoS ONE* 5 (8). https://doi.org/10.1371/journal.pone.0012060.

Eyre-Walker, a, R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. "Investigation of the Bottleneck Leading to the Domestication of Maize." *Proceedings of the National Academy of Sciences of the United States of America* 95 (April):4441–46. https://doi.org/10.1073/pnas.95.8.4441.

Fonseca, R. R. Da, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. Alfredo Samaniego, et al. 2015. "The Origin and Evolution of Maize in the Southwestern United States." *Nature Plants* 1. https://doi.org/10.1038/nplants.2014.3.

Fu, W., R. M. Gittelman, M. J. Bamshad, and J. M. Akey. 2014. "Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations." *American Journal of Human Genetics* 95 (4):421–36. https://doi.org/10.1016/j.ajhg.2014.09.006.

Fumagalli, M., F. G. Vieira, T. S. Korneliussen, T. Linderoth, E. Huerta-Sánchez, A. Albrechtsen, and R. Nielsen. 2013. "Quantifying Population Genetic Differentiation from next-Generation Sequencing Data." *Genetics* 195 (3):979–92. https://doi.org/10.1534/genetics.113.154740.

Fumagalli, M., F. G. Vieira, T. Linderoth, and R. Nielsen. 2014. "NgsTools: Methods for Population Genetics Analyses from next-Generation Sequencing Data." *Bioinformatics* 30 (10):1486–87. https://doi.org/10.1093/bioinformatics/btu041.

Gaut, B. S., C. M. Díez, and P. L. Morrell. 2015. "Genomics and the Contrasting Dynamics of Annual and Perennial Domestication." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2015.10.002.

Gerke, J. P., J. W. Edwards, K. E. Guill, J. Ross-Ibarra, and M. D. McMullen. 2015. "The Genomic Impacts of Drift and Selection for Hybrid Performance in Maize." *Genetics* 201 (3):1201–11. https://doi.org/10.1534/genetics.115.182410.

Gignoux, C. R., B. M. Henn, and J. L. Mountain. 2011. "Rapid, Global Demographic Expansions after the Origins of Agriculture." *Proceedings of the National Academy of Sciences* 108 (15):6044–49. https://doi.org/10.1073/pnas.0914274108.

Grobman, A., D. Bonavia, T. D. Dillehay, D. R. Piperno, J. Iriarte, and I. Holst. 2012.

"Preceramic Maize from Paredones and Huaca Prieta, Peru." *Proceedings of the National Academy of Sciences* 109 (5):1755–59. https://doi.org/10.1073/pnas.1120270109.

Günther, T., and K. J. Schmid. 2010. "Deleterious Amino Acid Polymorphisms in Arabidopsis Thaliana and Rice." *Theoretical and Applied Genetics* 121 (1):157–68. https://doi.org/10.1007/s00122-010-1299-4.

Harris, K., and R. Nielsen. 2016. "The Genetic Cost of Neanderthal Introgression." *Genetics* 203 (2):881–91. https://doi.org/10.1534/genetics.116.186890.

Hearne, S., E. S. Buckler, C. Chen, S. Mitchell, K. Swarts, H. Li, and A. Romero. 2017. "Imputed GbS Derived SNPs for Maize Landrace Accessions Represented in the SeeD-Maize GWAS Panel: Imputation Using FILLIN." CIMMYT Research Data & Software Repository Network. https://doi.org/11529/10036.

Heerwaarden, J. van, J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. de Jesus Sanchez Gonzalez, and J. Ross-Ibarra. 2011. "Genetic Signals of Origin, Spread, and Introgression in a Large Sample of Maize Landraces." *Proceedings of the National Academy of Sciences* 108 (3):1088–92. https://doi.org/10.1073/pnas.1013011108.

Henn, B. M., L. R. Botigué, C. D. Bustamante, A. G. Clark, and S. Gravel. 2015. "Estimating the Mutation Load in Human Genomes." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3931.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. "The Genomic Signature of Crop-Wild Introgression in Maize." *PLoS Genetics* 9 (5). https://doi.org/10.1371/journal.pgen.1003477.

Hufford, M. B, X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J-M. Chia, R. A Cartwright, R. J Elshire, et al. 2012. "Comparative Population Genomics of Maize Domestication and Improvement." *Nat Genet* 44 (7):808–11. https://doi.org/10.1038/ng.2309.

Juric, I., S. Aeschbacher, and G. Coop. 2016. "The Strength of Selection against Neanderthal Introgression." *PLoS Genetics* 12 (11). https://doi.org/10.1371/journal.pgen.1006340.

Kono, T. J.Y., F. Fu, M. Mohammadi, P. J. Hoffman, C. Liu, R. M. Stupar, K. P. Smith, P. Tiffin, J. C. Fay, and P. L. Morrell. 2016. "The Role of Deleterious Substitutions in Crop Genomes." *Molecular Biology and Evolution* 33 (9):2307–17. https://doi.org/10.1093/molbev/msw102.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15 (1). https://doi.org/10.1186/s12859-014-0356-4.

Korneliussen, T. S., I. Moltke, A. Albrechtsen, and R. Nielsen. 2013. "Calculation of Tajima's D

and Other Neutrality Test Statistics from Low Depth next-Generation Sequencing Data." *BMC Bioinformatics* 14 (1). https://doi.org/10.1186/1471-2105-14-289.

Lam, H. M., X. Xu, X. Liu, W. Chen, G. Yang, F. Ling Wong, M. Wah Li, et al. 2010. "Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection." *Nature Genetics* 42 (12):1053–59. https://doi.org/10.1038/ng.715.

Li, H., and R. Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 26 (5):589–95. https://doi.org/10.1093/bioinformatics/btp698.

Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, et al. 2008. "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation." *Science* 319 (5866):1100–1104. https://doi.org/10.1126/science.1153717.

Liu, Q., Y. Zhou, P. L. Morrell, B. S. Gaut, and S. Ge. 2017. "Deleterious Variants in Asian Rice and the Potential Cost of Domestication." *Molecular Biology and Evolution* 34 (4):908–24. https://doi.org/10.1093/molbev/msw296.

Lu, J., T. Tang, H. Tang, J. Huang, S. Shi, and C. I. Wu. 2006. "The Accumulation of Deleterious Mutations in Rice Genomes: A Hypothesis on the Cost of Domestication." *Trends in Genetics*. https://doi.org/10.1016/j.tig.2006.01.004.

Manna, F., G. Martin, and T. Lenormand. 2011. "Fitness Landscapes: An Alternative Theory for the Dominance of Mutation." *Genetics* 189 (3):923–37. https://doi.org/10.1534/genetics.111.132944.

Marsden, C. D., D. O-D. Vecchyo, D. P. O'Brien, J. F. Taylor, O. Ramirez, C. Vilà, T. Marques-Bonet, R. D. Schnabel, R. K. Wayne, and K. E. Lohmueller. 2016. "Bottlenecks and Selective Sweeps during Domestication Have Increased Deleterious Genetic Variation in Dogs." *Proceedings of the National Academy of Sciences* 113 (1):152–57. https://doi.org/10.1073/pnas.1512501113.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. "Evaluating the Use of ABBA-BABA Statistics to Locate Introgressed Loci." *Molecular Biology and Evolution* 32 (1):244–57. https://doi.org/10.1093/molbev/msu269.

Matsuoka, Y., Y. Vigouroux, M. M Goodman, J. Sanchez G, E. S. Buckler, and J. Doebley. 2002. "A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping." *Proc Natl Acad Sci U S A* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi. 2016. "On the Importance of Being Structured: Instantaneous Coalescence Rates and Human Evolution-Lessons for Ancestral Population Size Inference?" *Heredity* 116 (4):362–71. https://doi.org/10.1038/hdy.2015.104.

McQuillan, R., N. Eklund, N. Pirastu, M. Kuningas, B. P. McEvoy, T. Esko, T. Corre, et al. 2012. “Evidence of Inbreeding Depression on Human Height.” *PLoS Genetics* 8 (7). https://doi.org/10.1371/journal.pgen.1002655.

Merrill, W. L., R. J. Hard, J. B. Mabry, G. J. Fritz, K. R. Adams, J. R. Roney, and A. C. MacWilliams. 2009. “The Diffusion of Maize to the Southwestern United States and Its Impact.” *Proceedings of the National Academy of Sciences* 106 (50):21019–26. https://doi.org/10.1073/pnas.0906075106.

Meyer, R. S., J. Young Choi, M. Sanches, A. Plessis, J. M. Flowers, J. A. K. Dorph, et al. 2016. “Domestication History and Geographical Adaptation Inferred from a SNP Map of African Rice.” *Nature Genetics* 48 (9):1083–88. https://doi.org/10.1038/ng.3633.

Miao, B., Z. Wang, and Y. Li. 2017. “Genomic Analysis Reveals Hypoxia Adaptation in the Tibetan Mastiff by Introgression of the Gray Wolf from the Tibetan Plateau.” *Molecular Biology and Evolution* 34 (3):734–43. https://doi.org/10.1093/molbev/msw274.

Mita, S. De, and M. Siol. 2012. “EggLib: Processing, Analysis and Simulation Tools for Population Genetics and Genomics.” *BMC Genetics* 13. https://doi.org/10.1186/1471-2156-13-27.

Nielsen, R., and M. A. Beaumont. 2009. “Statistical Inferences in Phylogeography.” *Molecular Ecology*. https://doi.org/10.1111/j.1365-294X.2008.04059.x.

Ogut, F., Y. Bian, P. J. Bradbury, and J. B. Holland. 2015. “Joint-Multiple Family Linkage Analysis Predicts within-Family Variation Better than Single-Family Analysis of the Maize Nested Association Mapping Population.” *Heredity* 114 (6):552–63. https://doi.org/10.1038/hdy.2014.123.

Pearsall, D. M. 2008. “Plant Domestication and the Shift to Agriculture in the Andes.” *The Handbook Of South American Archaeology*, 105–20. https://doi.org/10.1007/978-0-387-74907-5_7.

Piperno, D. R., A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau. 2009. “Starch Grain and Phytolith Evidence for Early Ninth Millennium B.P. Maize from the Central Balsas River Valley, Mexico.” *Proceedings of the National Academy of Sciences* 106 (13):5019–24. https://doi.org/10.1073/pnas.0812525106.

Piperno, D. R, J E. Moreno, J. Iriarte, I. Holst, M. Lachniet, J. G. Jones, A. J. Ranere, and R. Castanzo. 2007. “Late Pleistocene and Holocene Environmental History of the Iguala Valley, Central Balsas Watershed of Mexico.” *Proceedings of the National Academy of Sciences* 104 (29). National Acad Sciences:11874–81.

Poets, A. M., Z. Fang, M. T. Clegg, and P. L. Morrell. 2015. “Barley Landraces Are Characterized by Geographically Heterogeneous Genomic Origins.” *Genome Biology* 16

(1). https://doi.org/10.1186/s13059-015-0712-3.

Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, et al. 2014. "The Complete Genome Sequence of a Neanderthal from the Altai Mountains." *Nature* 505 (7481):43–49. https://doi.org/10.1038/nature12886.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A.R. Ferreira, D. Bender, J. Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *The American Journal of Human Genetics* 81 (3):559–75. https://doi.org/10.1086/519795.

Races of maize. "Races of Maize : USDA ARS." Accessed 15 Apr 2015. Accessed December 25, 2017. https://www.ars.usda.gov/midwest-area/ames/plant-introduction-research/docs/races-of-maize/.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. 2005. "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa." *Proceedings of the National Academy of Sciences* 102 (44):15942–47. https://doi.org/10.1073/pnas.0507611102.

Ramu, P., W. Esuma, R. Kawuki, I. Y. Rabbi, C. Egesi, J. V. Bredeson, R. S. Bart, J. Verma, E. S. Buckler, and F. Lu. 2017. "Cassava Haplotype Map Highlights Fixation of Deleterious Mutations during Clonal Propagation." *Nature Genetics* 49 (6):959–63. https://doi.org/10.1038/ng.3845.

Renaut, S., and L. H. Rieseberg. 2015. "The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops." *Molecular Biology and Evolution* 32 (9):2273–83. https://doi.org/10.1093/molbev/msv106.

Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler. 2015. "Recombination in Diverse Maize Is Stable, Predictable, and Associated with Genetic Load." *Proceedings of the National Academy of Sciences*, 201413864. https://doi.org/10.1073/pnas.1413864112.

Romero Navarro, J. A., M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, et al. 2017. "A Study of Allelic Diversity Underlying Flowering-Time Adaptation in Maize Landraces." *Nature Genetics* 49 (3):476–80. https://doi.org/10.1038/ng.3784.

Ross-Ibarra, J., M. Tenaillon, and B. S. Gaut. 2009. "Historical Divergence and Gene Flow in the Genus Zea." *Genetics* 181 (4):1399–1413. https://doi.org/10.1534/genetics.108.097238.

Schiffels, S., and R. Durbin. 2014. "Inferring Human Population Size and Separation History from Multiple Genome Sequences." *Nature Genetics* 46 (8):919–25.

https://doi.org/10.1038/ng.3015.

Schliep, K. P.. 2011. “Phangorn: Phylogenetic Analysis in R.” *Bioinformatics (Oxford, England)* 27 (4):592–93. https://doi.org/10.1093/bioinformatics/btq706.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, et al. 2009. “The B73 Maize Genome: Complexity, Diversity, and Dynamics.” *Science* 326 (5956):1112–15. https://doi.org/10.1126/science.1178534.

Schrider, D. R., A. G. Shanku, and A. D. Kern. 2016. “Effects of Linked Selective Sweeps on Demographic Inference and Model Selection.” *Genetics* 204 (3):1207–23. https://doi.org/10.1534/genetics.116.190223.

Schubert, M., H. Jónsson, D. Chang, C. D. Sarkissian, L. Ermini, A. Ginolhac, A. Albrechtsen, et al. 2014. “Prehistoric Genomes Reveal the Genetic Foundation and Cost of Horse Domestication.” *Proceedings of the National Academy of Sciences* 111 (52):E5661–69. https://doi.org/10.1073/pnas.1416991111.

Simons, Y. B., M. C. Turchin, J. K. Pritchard, and Guy Sella. 2014. “The Deleterious Mutation Load Is Insensitive to Recent Population History.” *Nature Genetics* 46 (3):220–24. https://doi.org/10.1038/ng.2896.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. “Estimating Individual Admixture Proportions from next Generation Sequencing Data.” *Genetics* 195 (3):693–702. https://doi.org/10.1534/genetics.113.154138.

Slatkin, M., and L. Excoffier. 2012. “Serial Founder Effects during Range Expansion: A Spatial Analog of Genetic Drift.” *Genetics* 191 (1):171–81. https://doi.org/10.1534/genetics.112.139022.

Takuno, S., P. Ralph, K. Swart, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. 2015. “Independent Molecular Basis of Convergent Highland Adaptation in Maize.” *Genetics* 200 (4):1297–1312. https://doi.org/10.1534/genetics.115.178327.

Tenaillon, M. I., J U’Ren, O. Tenaillon, and B. S. Gaut. 2004. “Selection versus Demography: A Multilocus Investigation of the Domestication Process in Maize.” *Molecular Biology and Evolution* 21 (7):1214–25. https://doi.org/10.1093/molbev/msh102.

Thompson, E. A. 2013. “Identity by Descent: Variation in Meiosis, across Genomes, and in Populations.” *Genetics*. https://doi.org/10.1534/genetics.112.148825.

Vieira, F. G., M. Fumagalli, A. Albrechtsen, and R. Nielsen. 2013. “Estimating Inbreeding Coefficients from NGS Data: Impact on Genotype Calling and Allele Frequency Estimation.” *Genome Research* 23 (11):1852–61. https://doi.org/10.1101/gr.157388.113.

Vigouroux, Y., J. C. Glaubitz, Y. Matsuoka, M. M. Goodman, J. Sánchez G., and J. Doebley.

2008. “Population Structure and Genetic Diversity of New World Maize Races Assessed by DNA Microsatellites.” *American Journal of Botany* 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097.

Whitlock, MC, and D. E McCauley. 1990. “Some Population Genetic Consequences of Colony Formation and Extinction: Genetic Correlations within Founding Groups.” *Evolution* 44 (7):1717–24. https://doi.org/10.2307/2409501.

Wright, S. I, I. V. Bi, S. G Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. “The Effects of Artificial Selection on the Maize Genome.” *Science (New York, N.Y.)* 308 (5726):1310–14. https://doi.org/10.1126/science.1107891.

Yang, J., S. Mezmouk, A. Baumgarten, E. S. Buckler, K. E. Guill, M. D. McMullen, R. H. Mumm, and J. Ross-Ibarra. 2017. “Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize.” *PLoS Genetics* 13 (9). https://doi.org/10.1371/journal.pgen.1007019.

Zhang, M., L. Zhou, R. Bawa, H. Suren, and J. A. Holliday. 2016. “Recombination Rate Variation, Hitchhiking, and Demographic History Shape Deleterious Load in Poplar.” *Molecular Biology and Evolution* 33 (11):2899–2910. https://doi.org/10.1093/molbev/msw169.

Zhou, Y., M. Massonnet, J. S Sanjak, D. Cantu, and B. S. Gaut. 2017. “Evolutionary Genomics of Grape (Vitis Vinifera Ssp. Vinifera) Domestication.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (44):11715–20. https://doi.org/10.1073/pnas.1709257114.

Zhu, Q., X. Zheng, J. Luo, B. S. Gaut, and S. Ge. 2007. “Multilocus Analysis of Nucleotide Variation of Oryza Sativa and Its Wild Relatives: Severe Bottleneck during Domestication of Rice.” *Molecular Biology and Evolution* 24 (3):875–88. https://doi.org/10.1093/molbev/msm005.

III.9 Supplementary Material

Additional file 1

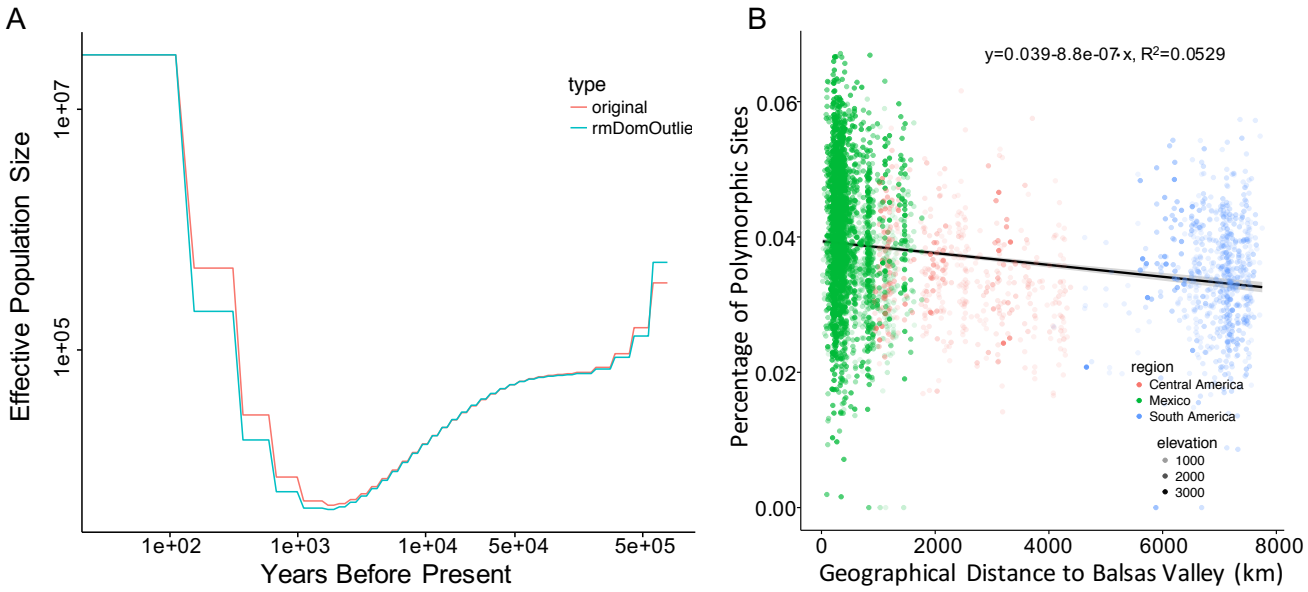


Figure S1. Demography of maize populations. A. MSMC results before and after masking candidate regions under selection during domestication. B. Percentage of heterozygous sites versus distance from the Balsas Valley in 3520 samples from the SeeDs data set.

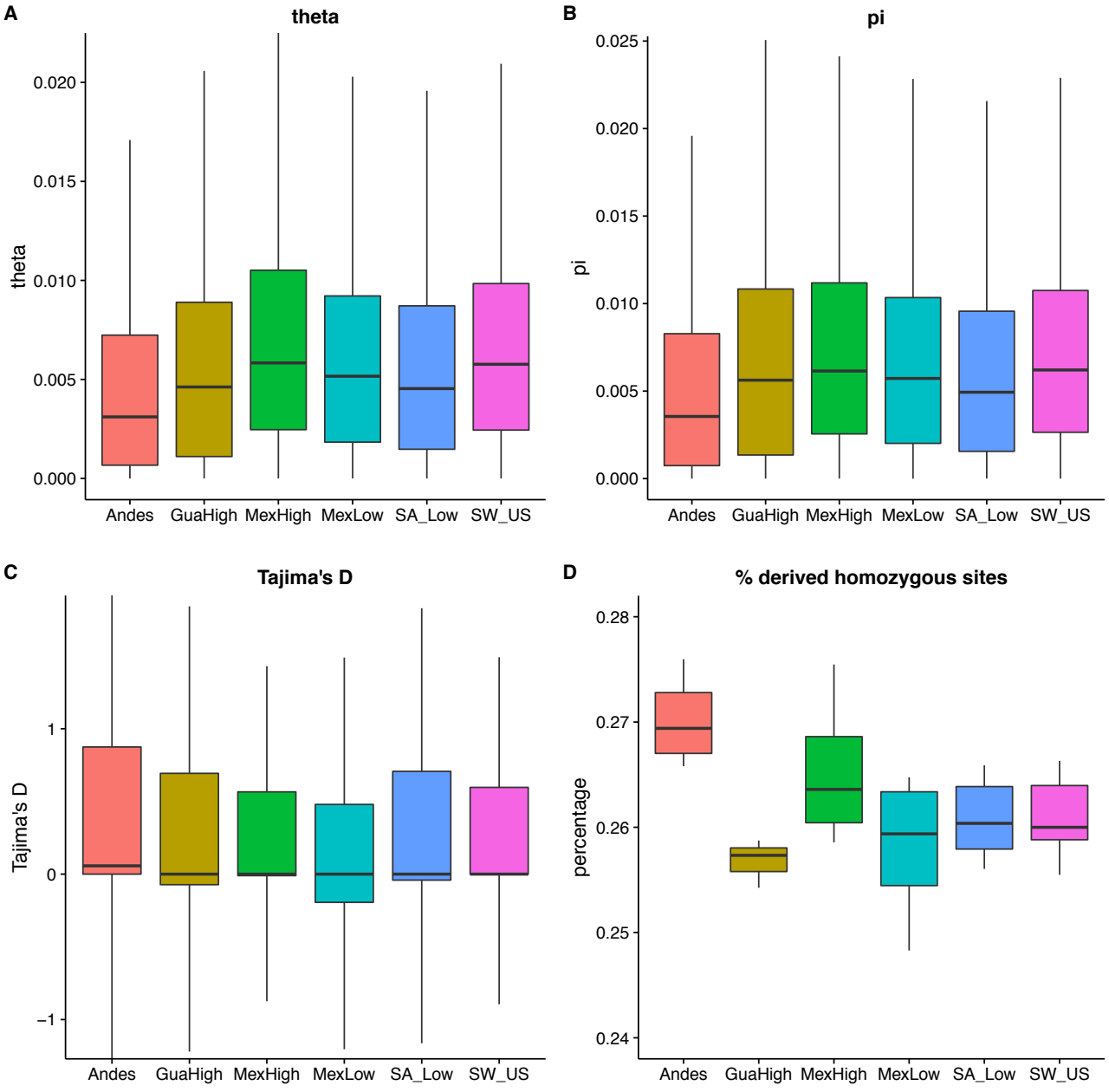


Figure S2. Watterson's theta (A), θ_π (B) and Tajima's D (C) are based on values in 10-kb non-overlapping windows across the genome. Percentage of derived homozygous sites was calculated for each individual and reported per population.

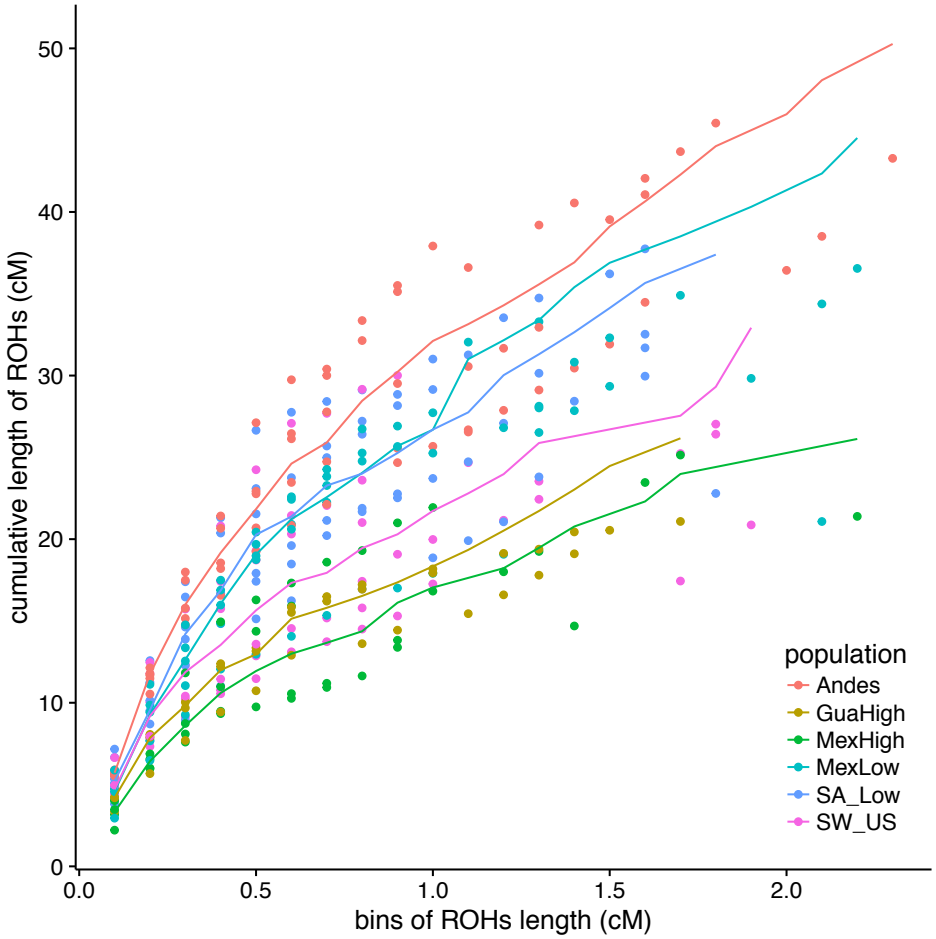


Figure S3. Cumulative length of ROHs in cM across populations. Lines indicate median values in each population. ROH: runs of homozygosity.

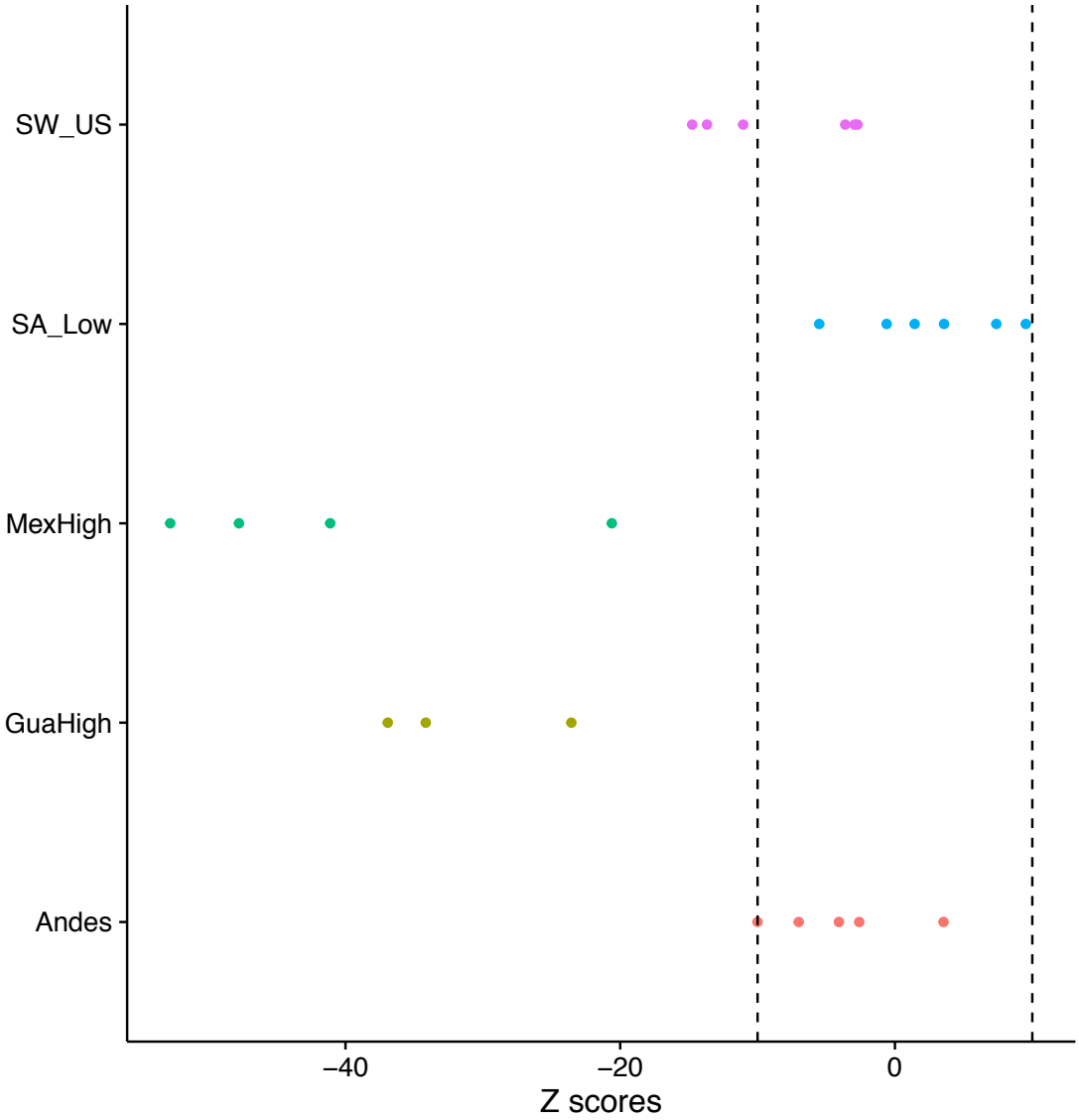


Figure S4. Calculation of D statistic across populations. Evidence of introgression from *mexicana* into Mexican highland, Guatemalan highland and Southwestern US highland maize populations. The dashed lines correspond to Z scores equal to -10 and 10.

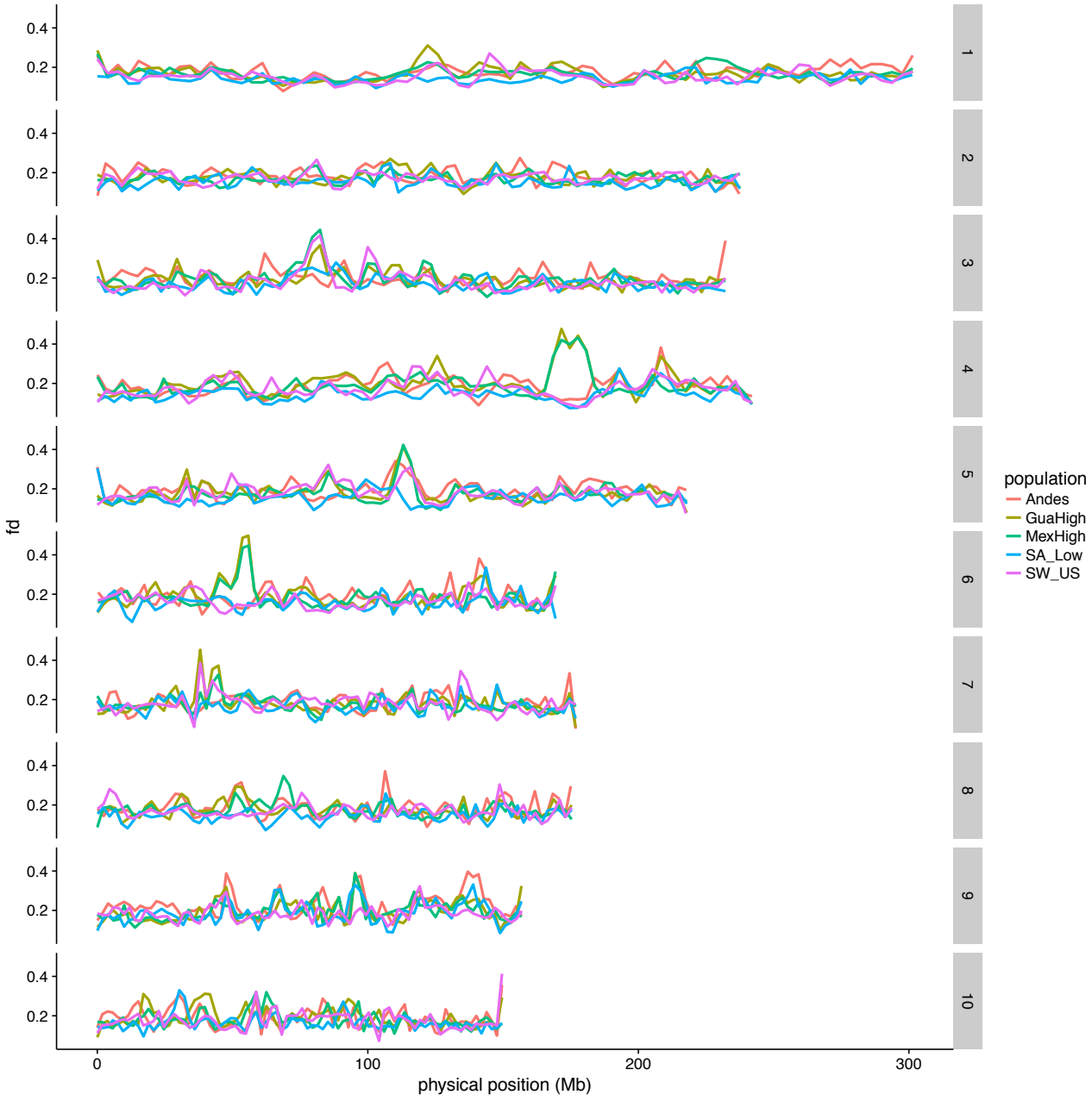


Figure S5. fd statistic results. Loess regression of fd in 10-kb non-overlapping windows across all chromosomes.

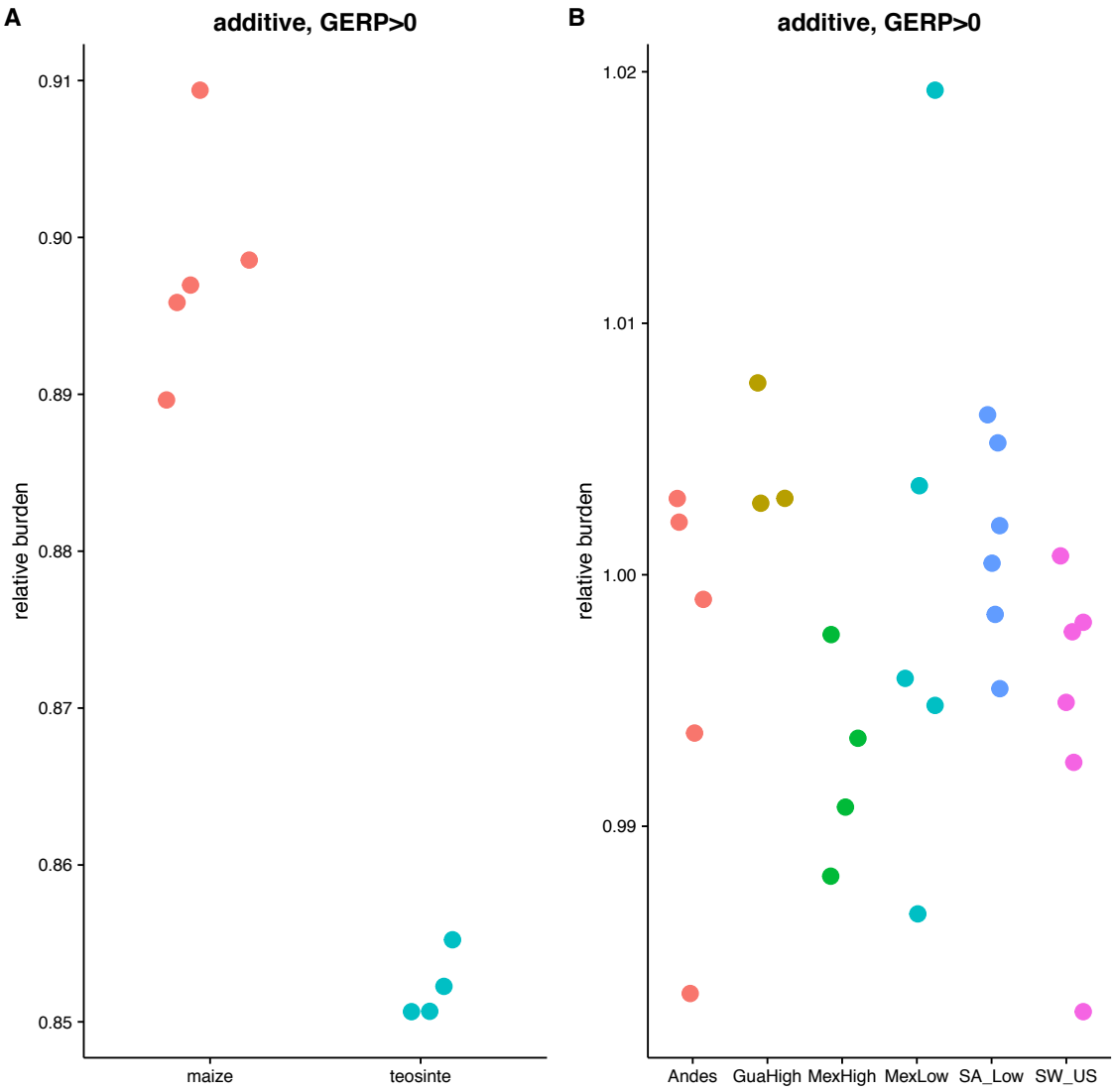


Figure S6. Relative burden of deleterious alleles under additive model between maize and teosinte (A) and among maize populations (B).

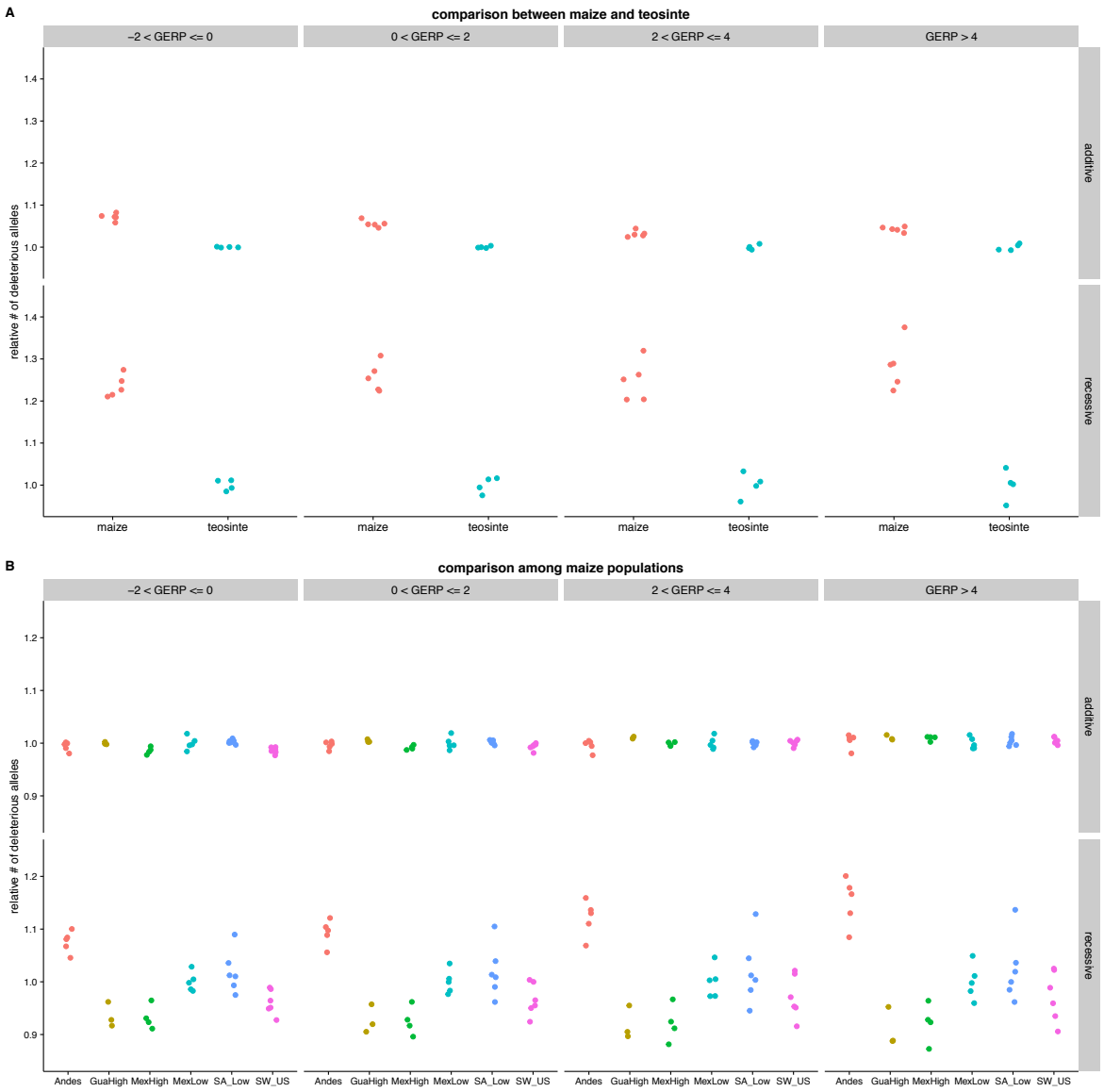


Figure S7. Relative burden of deleterious alleles under both additive and recessive models with different GERP partitions between maize and teosinte (A) and among maize populations (B).

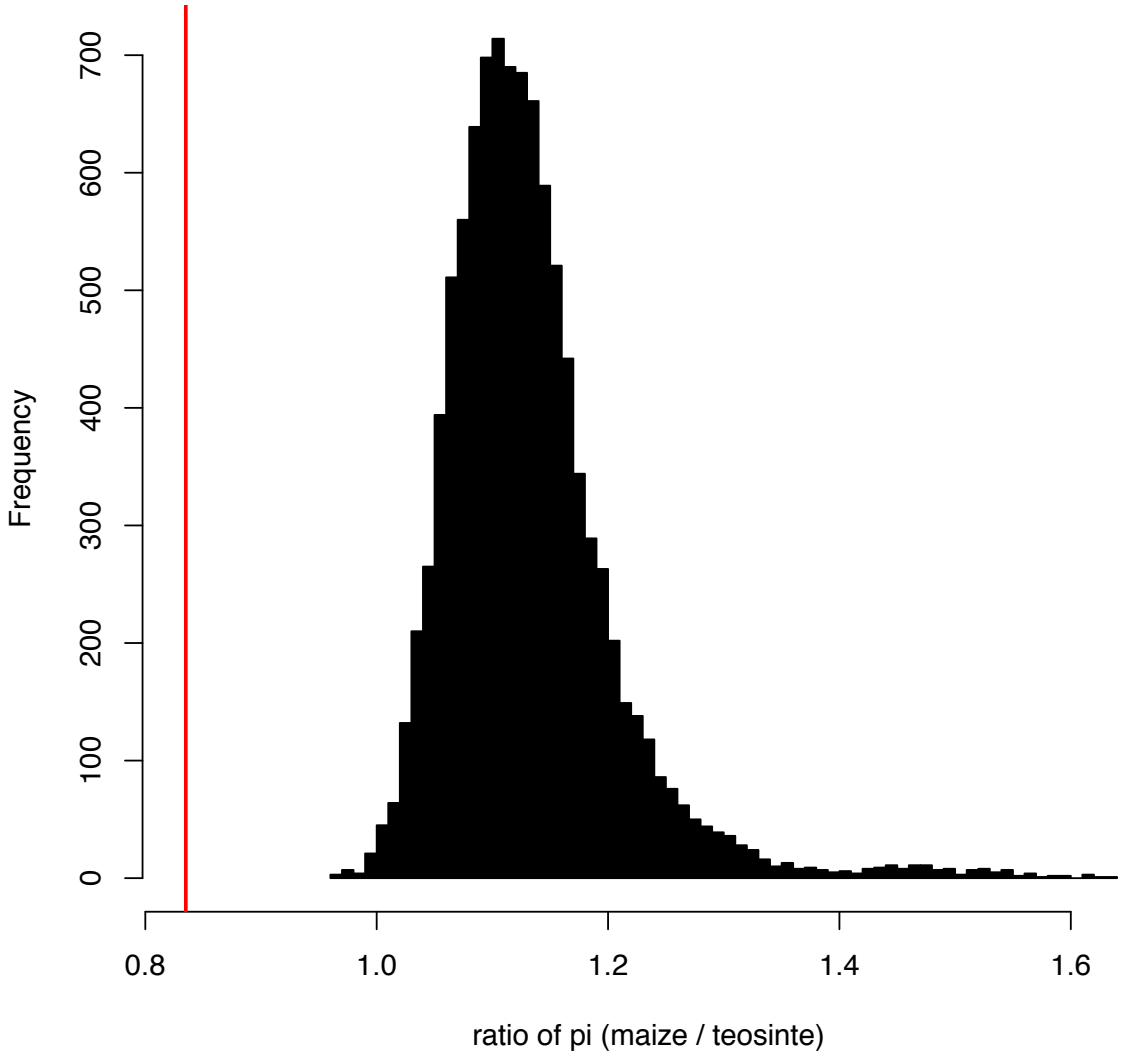


Figure S8. Domestication candidate genes exhibited lower θ_π ratio between maize and teosinte, a signal of selection in these genes. Distribution of ratio of θ_π between maize and teosinte in 420 domestication candidate genes (mean value was indicated with red line) against 10,000 replicates of genome-wide sampling of 420 random genes.

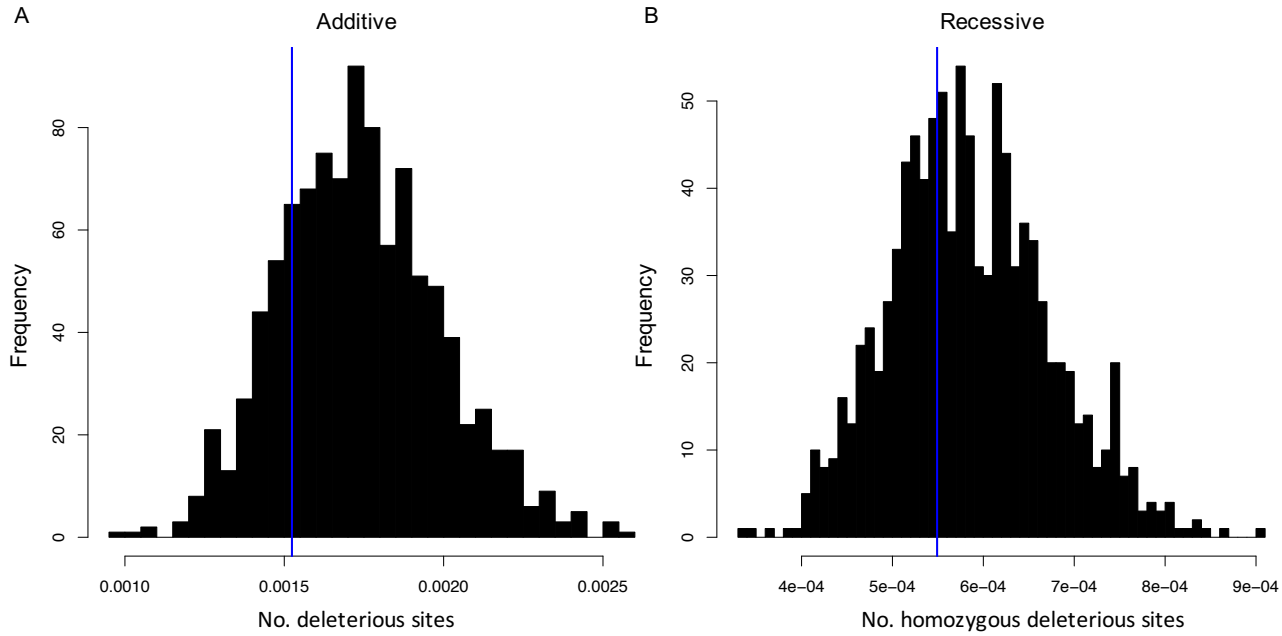


Figure S9. Distribution of number of deleterious sites per bp in 420 domestication candidate genes (indicated with blue line) compared to genome-wide random samples under an (A) additive model and (B) recessive model.

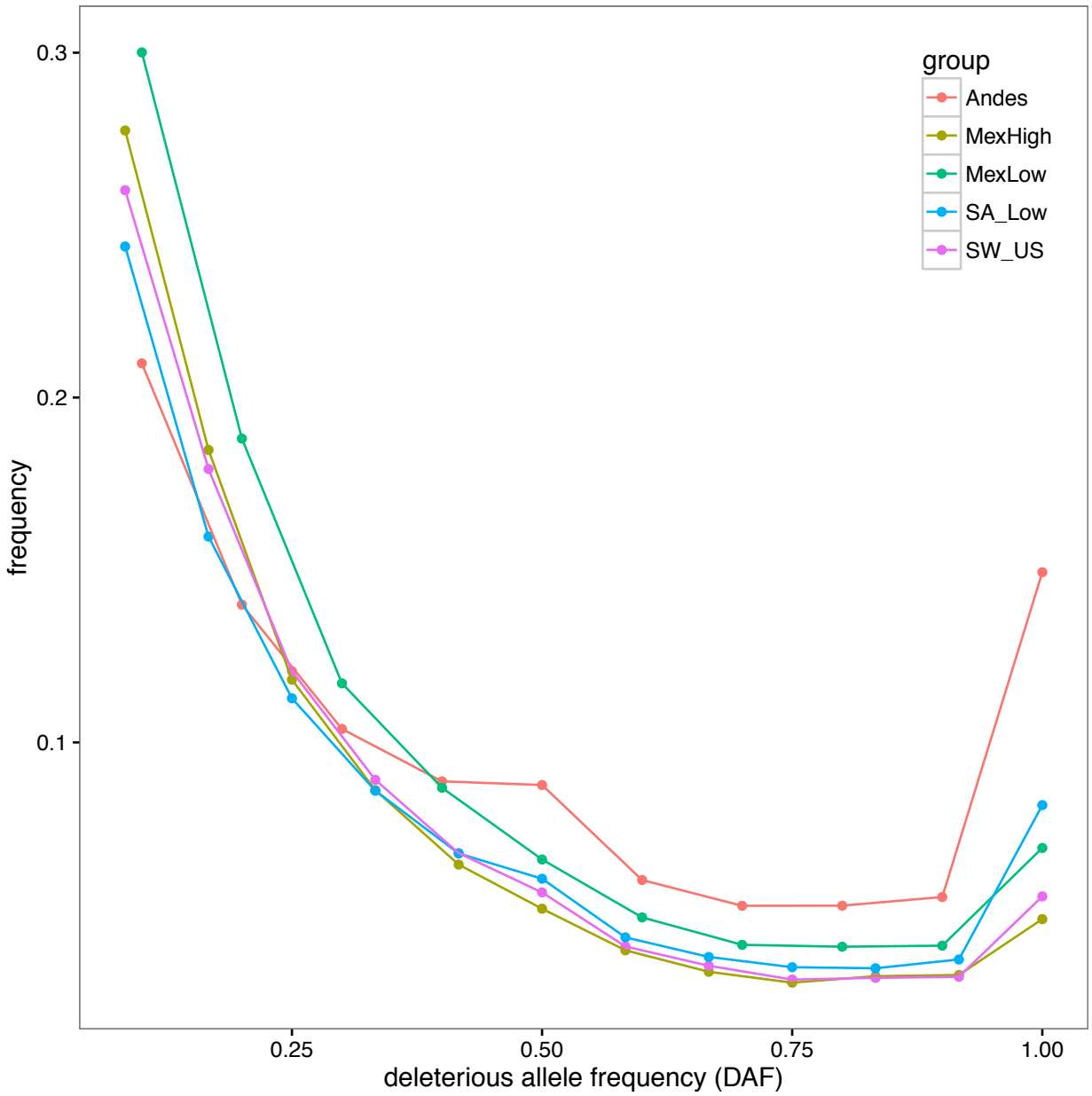


Figure S10. Site frequency spectrum of deleterious SNPs in five populations. GuaHigh is not included since the small sampling limited power for the SFS.

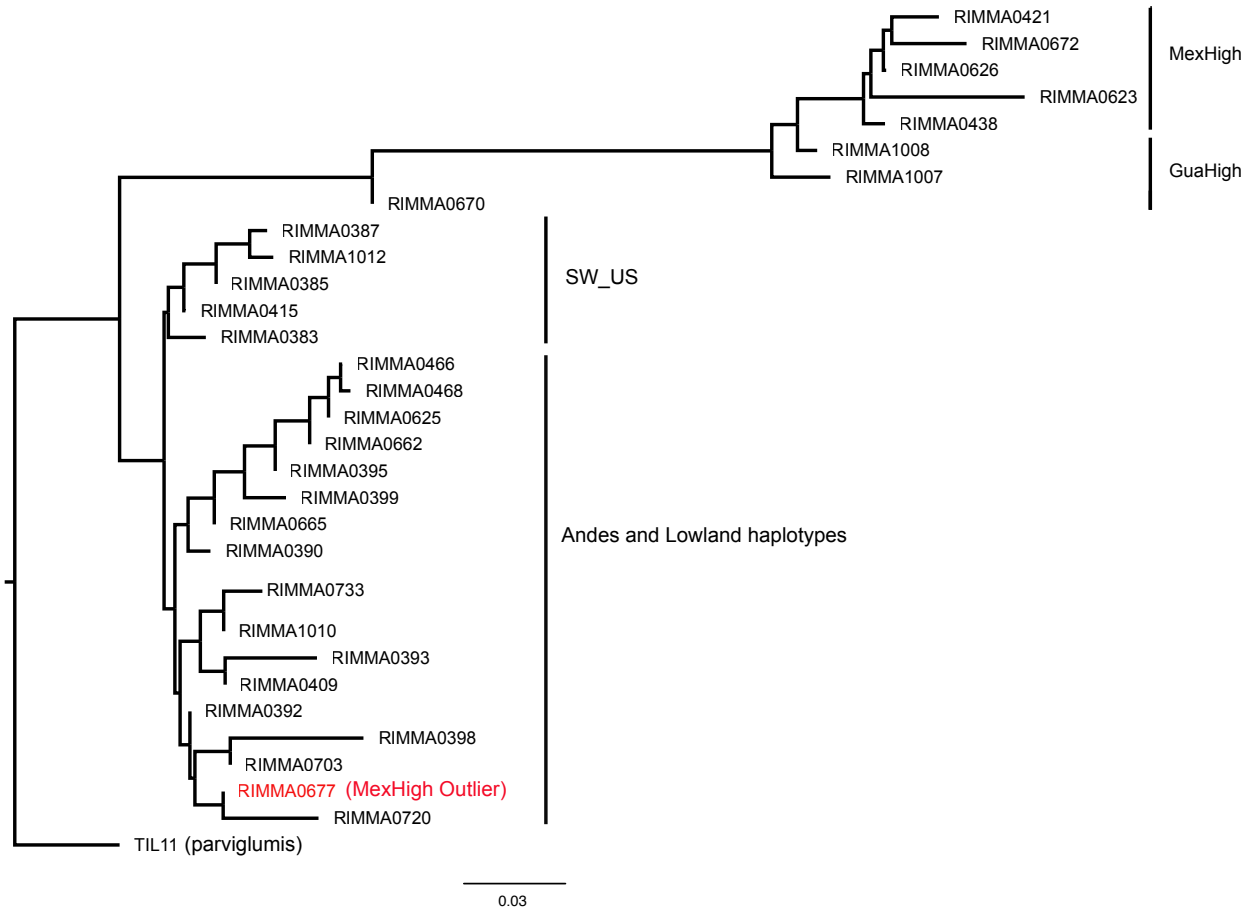


Figure S11. Neighbor Joining tree of SNPs from an inversion on chromosome 4 with a diagnostic haplotype for highland Mexican material.

Additional file 2

RI_Group	Latitude	Longitude	RI_Accession	Elevation	Locality
Andes	-14.317	-72.917	RIMMA0466	3600	Apurimac, Peru
Andes	-9.383	-77.167	RIMMA0468	3150	Ancash, Peru
Andes	-8.700	-77.383	RIMMA0625	2820	Ancash, Peru
Andes	0.000	-78.000	RIMMA0662	2195	Ecuador
Andes	-0.917	-78.917	RIMMA0665	2931	Ecuador
GuaHigh	14.967	-91.767	RIMMA0670	2378	San Marcos, Guatemala
GuaHigh	15.033	-91.783	RIMMA1007	3049	San Marcos, Guatemala
GuaHigh	14.917	-91.333	RIMMA1008	2774	Totonicapan, Guatemala
MexHigh	19.850	-97.983	RIMMA0421	2250	Puebla, Mexico
MexHigh	19.000	-97.383	RIMMA0438	2600	Puebla, Mexico
MexHigh	20.033	-103.683	RIMMA0623	2520	Jalisco, Mexico
MexHigh	19.883	-97.583	RIMMA0626	2260	Puebla, Mexico
MexHigh	19.683	-99.133	RIMMA0672	2256	Mexico, Mexico
MexHigh	21.367	-102.850	RIMMA0677	1951	Zacatecas, Mexico
MexLow	15.433	-92.900	RIMMA0409	107	Chiapas, Mexico
MexLow	20.833	-88.517	RIMMA0703	30	Yucatan, Mexico
MexLow	15.467	-88.850	RIMMA0720	39	Guatemala
MexLow	16.567	-94.617	RIMMA0733	107	Oaxaca, Mexico
MexLow	16.850	-99.067	RIMMA1010	201	La Concordia, Guerrero
SA_Low	4.517	-75.633	RIMMA0390	353	Caldas, Colombia
SA_Low	1.750	-75.583	RIMMA0392	555	Caqueta, Colombia
SA_Low	8.317	-75.150	RIMMA0393	100	Cordoba, Colombia
SA_Low	8.500	-77.267	RIMMA0395	30	Choco, Colombia
SA_Low	9.433	-75.700	RIMMA0398	27	Magdalena, Colombia
SA_Low	10.183	-74.050	RIMMA0399	250	Magdalena, Colombia
SW_US	34.900	-107.583	RIMMA0383	2073	Acoma Pueblo, NM, USA
SW_US	36.050	-106.283	RIMMA0384	2134	San Lorenzo Pueblo, NM, USA
SW_US	36.450	-105.550	RIMMA0385	2134	Taos Pueblo, NM, USA
SW_US	35.617	-106.733	RIMMA0387	1829	Jemez Pueblo, NM, USA
SW_US	35.900	-110.667	RIMMA0415	1941	Hotevilla, Arizona, USA
SW_US	35.762	-105.933	RIMMA1012	2073	Tesuque Pueblo, NM

IV. THE POTENTIAL ROLE OF GENETIC ASSIMILATION DURING MAIZE DOMESTICATION

Anne Lorant[1], Sarah Pedersen[2], Irene Holst[3], Matthew B. Hufford[2], Klaus Winter[3], Dolores Piperno[3,4], Jeffrey Ross-Ibarra[1,5*]

[1]Dept. of Plant Sciences, University of California, Davis, CA, USA

[2]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

[3]Smithsonian Tropical Research Institute, Panama, Republic of Panama

[4]Department of Anthropology, Smithsonian National Museum of Natural History, Washington, DC, USA

[5]Genome Center and Center for Population Biology, University of California, Davis, CA, USA

Keywords: Maize , Domestication, Gene expression, Paleogenetics, carbon dioxide

Published: September 8, 2017 in PlosOne https://doi.org/10.1371/journal.pone.0184202

IV.1 Abstract

Domestication research has largely focused on identification of morphological and genetic differences between extant populations of crops and their wild relatives. Little attention has been paid to the potential effects of environment despite substantial known changes in climate from the time of domestication to modern day. In recent research, the exposure of teosinte (i.e., wild maize) to environments similar to the time of domestication, resulted in a plastic induction of domesticated phenotypes in teosinte. These results suggest that early agriculturalists may have selected for genetic mechanisms that cemented domestication phenotypes initially induced by a plastic response of teosinte to environment, a process known as genetic assimilation. To better understand this phenomenon and the potential role of environment in maize domestication, we examined differential gene expression in maize (*Zea mays* ssp. *mays*) and teosinte (*Zea mays* ssp. *parviglumis*) between past and present conditions. We identified a gene set of over 2000 loci showing a change in expression across environmental conditions in teosinte and invariance in maize. In fact, overall we observed both greater plasticity in gene expression and more substantial changes in co-expressionnal networks in teosinte across environments when compared to maize. While these results suggest genetic assimilation played at least some role in domestication, genes showing expression patterns consistent with assimilation are not significantly enriched for previously identified domestication candidates, indicating assimilation did not have a genome-wide effect.

IV.2 Introduction

The development of agricultural societies 12,000–9,000 years ago (ka) was one of the most transformative events in human and ecological history and was made possible by plant and animal domestication (Larson et al. 2014; Piperno and Flannery 2001). During domestication, crops evolved a suite of phenotypic traits, collectively known as the domestication syndrome, that distinguish them from their wild relatives (Gepts 2012). Modifications due to domestication frequently include, for example, gigantism in the harvested plant part, reduced branching, and loss of shattering (Gepts 2012). Scientists have sought for centuries to understand the evolution of crops during domestication, making inferences based on imperfect genetic and archaeological data. Population genetic analysis of changes associated with domestication are limited by the still sparse availability of ancient DNA, and the archaeobotanical record is often chronologically coarse and geographically uneven (e.g, (Larson et al. 2014; Piperno and Flannery 2001)). As a result of these limitations, our current understanding of the morphological and molecular differences between domesticates and their wild ancestors is based almost exclusively on living representatives of those taxa. Most of what is known about maize domestication, for example, has been drawn from comparisons between extant cultivated and wild plants. Today, profound morphological differences in vegetative architecture and inflorescence sexuality distinguish domesticated maize (*Zea mays* ssp. *mays*) and its wild ancestor teosinte (*Zea mays* ssp. *parviglumis* Iltis and Doebley; hereafter *parviglumis*). Modern teosinte has long lateral branches tipped by tassels (male inflorescences) and secondary branches bearing ears (female inflorescences) with a few small seeds covered by hard fruit cases that mature sequentially over a period of a few months. Maize, in contrast, has a single main stem terminating in a tassel and few dramatically shortened lateral branches terminated by ears instead of tassels. Maize seeds are not covered by fruit cases and its

cobs mature at about the same time. These differences, the most dramatic documented for any major crop/ancestor pair, led to a century-long debate about maize ancestry (Beadle 1972; Matsuoka et al. 2002; Doebley 2004).

Because of its importance economically and as a genetic model organism, the genetics underlying the process of maize domestication has received considerable attention. Early crossing work by Beadle (Beadle 1972) suggested as few as five genes could be responsible for the major vegetative architecture and inflorescence sexuality differences between maize and teosinte. More recently, work mapping quantitative trait loci (QTL) found generally consistent results, identifying six major QTL (Briggs et al. 2007). The vegetative architecture and inflorescence sexuality differences noted above, for example, are to a large degree controlled by the major QTL teosinte branched1 (*tb1*) through a change in gene expression occurring early in plant development (Doebley, Stec, and Gustus 1995; Hubbard et al. 2002; Studer et al. 2011a). Evidence of positive selection during domestication has been found at many more loci than those identified as QTL, however (Bomblies and Doebley 2006; Studer et al. 2011b; Vollbrecht et al. 2005; Wang et al. 2005; Wills et al. 2013), as genome-wide scans find that as much as 5% of the genome may have played a functional role in domestication (Hufford, Xu, et al. 2012; Wright et al. 2005). While there are examples such as *tga1* in which selection acted on an amino acid substitution changing the protein sequence of a gene (Wang et al. 2005), considerable evidence suggests that much of the evolution during domestication was regulatory in nature. Not only do genes showing evidence of selection show directional changes in expression (Hufford, Xu, et al. 2012), but many of the transcription and co-expression networks of maize have been substantially modified during domestication (Swanson-Wagner et al. 2012), due in part to change in *cis* regulatory elements (Lemmon et al. 2014).

In spite of this large body of work, domestication research has primarily focused on comparisons of extant crops and wild relatives and has largely ignored the effects of changing environmental conditions during the timeframe of crop evolution. Agricultural beginnings occurred during a period of profound global environmental change as the Pleistocene was ending and transitioning to the Holocene interglacial period (Larson et al. 2014; Piperno 2011). It is well documented that atmospheric CO_2 and temperature were considerably lower than at present during both the Late Pleistocene (c. 14–11ka) and earliest Holocene (c. 11–9ka) (Ahn et al. 2004; Piperno et al. 2007; Hodell et al. 2008; Bush et al. 2009; Restrepo et al. 2012). Recent experimental work by Piperno and coauthors (Piperno et al. 2015) demonstrated remarkable phenotypic changes in teosinte exposed to temperatures and atmospheric CO_2 similar to those experienced during the Late Pleistocene and early Holocene. These changes included maize-like vegetative architecture, inflorescence sexuality, and seed maturation, together with decreased plant height, biomass, and seed yield (Piperno et al. 2015). This work points to the possibility that early cultivators may have worked with phenotypes considerably different from those of modern teosinte. Furthermore, because some of the observed changes under experimental environments appear to have been a result of phenotypic (developmental) plasticity, the results suggest a possible role for plasticity in maize domestication (Piperno et al. 2015).

Developmental or phenotypic plasticity refers to the inherent capacity of organisms to rapidly produce novel phenotypes through one of several developmental pathways in direct response to changing environment (e.g., (Beldade, Mateus, and Keller 2011; Gilbert and Epel 2009; Hendry 2016; Moczek et al. 2011; West-Eberhard 2003)). Plasticity is now established as a mainstream concept in evolution and ecology and is increasingly considered to be fundamental for understanding the genesis of phenotypes (Ledón-Rettig, Pfennig, and Crespi 2010; Magalhaes et

al. 2009; Pfennig and McGee 2010; Schneider et al. 2014). Both early and recent research has also shown that genetic modifications can cement plastic phenotypes, making them stable and heritable (Schlichting and Wund 2014; Waddington 1953). One such mechanism is genetic assimilation (GA), a process that was first investigated during the early period of the Modern Synthesis (Pigliucci 2006; Waddington 1953). Genetic assimilation involves a loss of plasticity and fixed expression across environments through reconfiguration of pre-existing genetic variation after a number of generations of growth in inducing conditions. Recent studies have demonstrated GA likely occurring in a variety of organisms, from tetrapods to *Solanum* spp. to early *Homo*, though its frequency and importance are still debated (Diggle and Miller 2013; Kuzawa and Bragg 2012; Standen, Du, and Larsson 2014).

Here we extend results from Piperno *et al*. (Piperno et al. 2015) on teosinte responses to environmental changes, investigating the potential role of plasticity in a transcriptome-wide analysis of differential gene expression in both teosinte and maize in modern and early Holocene climate conditions. We hypothesized that expression-level changes may have constituted an initial plastic response to changing environment at the time of domestication that was later canalized through the process of GA. We find a large number of loci that show environmentally-mediated differential expression in teosinte but not maize, including some with functions consistent with phenotypic differences observed between different experimental environments and between maize and teosinte. While population genetic evidence and enrichment analyses suggest these loci are not enriched for genes showing signals of selection during domestication, a number of loci nonetheless coincide with previously identified selective sweeps, potentially suggesting a role for GA during maize domestication. Finally, we also find a large number of genes differentially

expressed in teosinte that are not identified as domestication candidates but that may nevertheless shed important light on plant responses during domestication.

IV.3 Material and methods

Growth chamber experiment

Seeds were provided by the USDA North Central Regional Plant Introduction Station located in Ames, Iowa. We sampled three individuals of four natural populations of *parviglumis* representative of the current geographic and elevational range of the subspecies (Hufford, Martínez-Meyer, et al. 2012) as well as two individuals of four maize inbred lines (S1 Table).

We undertook two grow-outs in 2013 and 2014 with teosinte and maize, respectively, during their typical growing periods from July to December in two naturally-lit glass environmental chambers housed at the Gamboa field station at the Smithsonian Tropical Research Institute in Panama. One chamber was adjusted to Early Holocene (EH) temperature (ca. 23°C) and CO_2 (ca. 260–265 ppmv) levels determined for the low elevation Neotropics including Mesoamerica for ca. 10,000–9000 ka from paleoecological research and ice core data (Ahn et al. 2004; Piperno et al. 2007; Hodell et al. 2008; Bush et al. 2009; Restrepo et al. 2012). The other chamber served as a modern ambient (MA) control and was kept at ambient CO_2 levels and temperatures, characteristic of *parviglumis* environments today (Hufford, Martínez-Meyer, et al. 2012).

The EH chamber average CO_2 and temperature levels were 259.7 ppmv and 23.3°C and 260.8 ppmv and 23.2°C in 2013 and 2014, respectively. The MA average chamber temperature

was 25.1°C in both years, with an average CO_2 of 371 and 374 ppmv, respectively. Additional details on chamber environments can be found in (Piperno et al. 2015).

Plants were germinated from seed in five-gallon pots in natural topsoil from a local orchard and watered without fertilizer three to four times per week. In 2013 three plants were grown from each of the four *parviglumis* accessions in each chamber, followed by two replicates of each of the four maize inbreds in each chamber in 2014. While this leads to a confounding effect of year, we are unaware of any reason why the very small chamber differences between the two years would contribute to the observed gene expression differences between parviglumis and maize (see below). We recorded weekly measurements of plant height, branch length and number, and inflorescence characteristics. After plants were harvested at maturity they were air dried and we measured the total vegetative biomass (stems, leaf, sheaths, ear bracts), node number and plant height (S2 and S3 Tables).

RNAseq experiment

Plants were sampled for gene expression 50 or 60 days after germination by removing the first visible leaf on the plant and placing it immediately in liquid nitrogen. For the first year, teosintes were collected after 60 days. During the second grow-out, maizes were starting to flower after 50 days in both conditions and were therefore collected 10 days earlier. Samples were stored at -80°C until they were shipped overnight on dry ice to UC Davis and kept at -80°C until extraction. One teosinte plant of population 4 (pop4.B.1) didn't grown in the 400ppm chamber and was not collected for RNA extraction. Leaf tissue was ground in liquid nitrogen, and total RNA was isolated with the RNeasy mini Kit (Qiagen) following the manufacturer's protocol. RNA quality and concentration were verified using a Bioanalyzer (Agilent RNA Nano). Total mRNA

was extracted twice with Dynabeads oligo(dt)25 (Ambion) from 2μg of total RNA. We prepared libraries as previously described (Zhong et al. 2011), with minor modifications and without the strand specificity. Samples were multiplexed and sequenced in two lanes of an Illumina Hiseq 2500 at the UCDavis Genome Center sequencing facility, resulting in 50 bp single-end reads with an insert size of approximately 300 bases. After demultiplexing, 3.8–20 million reads were generated for each sample (S4 Table). The raw sequence data has been deposited in NCBI Sequence Read Archive with the BioProject ID PRJNA391707(https://www.ncbi.nlm.nih.gov/bioproject/PRJNA391707). Low quality bases (base quality < 33) were trimmed using FASTX-Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) and adapters were subsequently removed using fastq-mcf version 1.04 (https://code.google.com/archive/p/ea-utils/wikis/FastqMcf.wiki). Trimmed reads were mapped to the AGPv3.22 version of the maize genome using Gmap/Gsnap version 2014-05-15 with command line parameters of *-m 10 -i 2 -N 1 -w 10000 -A sam -t 8 -n 3* (Wu et al. 2016). Read counting was performed with biocLite GenomicAlignments (Lawrence et al. 2013) (S1 File, Maize_readcounts and Teosinte_readcounts); only reads with mapping quality 25 or higher were included in subsequent analyses. Differential gene expression was performed with DEseq2 1.10.1 (Love, Huber, and Anders 2014) using a linear model (~genotype + condition) accounting for both environment (EH and MA) and population of origin. The models were run separately for maize and teosinte. In both cases, we included multiple plants per population/genotype. Individual plants from each maize inbred line were treated as biological replicates. Teosinte, however, is an outbred and each plant was thus included separately but population was used as a covariate. The model is then constructed on n individuals, p genotype or population levels and 2 environmental levels, with the effect size of environment being the

measured effect. We used a false discovery rate (FDR) cutoff of 0.05 for determining differentially expressed genes (S1 File, Maize_DE and Teosinte_DE). We use intra-chamber variation as experimental error, so the statistical significances reported here are over-estimates. To mitigate these effects, we also applied a more stringent FDR cutoff of 0.01. Our results remained qualitatively identical suggesting that effects of such pseudo-replication would have to be substantial to impact our general conclusions. We removed 15 genes (5 from maize, 7 from teosinte, and 3 from both, list available in S1 File) identified as showing differential expression before and after flowering (Alter et al. 2016) to account for the difference of developmental stage between the two subspecies.

Co-expression networks

Co-expression analysis was conducted using the program WGCNA (Langfelder and Horvath 2008). Raw expression counts were normalized using the variance stabilizing transformation in DESeq2 (Love, Huber, and Anders 2014). Genes that were not expressed in both maize and teosinte across both environmental treatments were filtered from the dataset, leaving 29,611 genes. Co-expression networks were created for maize and teosinte individually based on expression values in the EH treatment. Pearson correlation values of expression were first assigned to all pairs of genes and then used to create adjacency matrices by raising the correlation value to a soft power as determined by the data and unique to each network (24 and 10, for maize and teosinte, respectively). Topological overlap matrices were then formed from the adjacency matrices. The adjacency matrix indicates the connection strength between two genes (edge weights within the network), while the topological overlap matrix indicates the degree of connectivity between two genes based on their interactions with other genes in the network as well as with each

other. Topological overlap matrices were used to create dissimilarity measures, which were then used to construct modules based on average linkage hierarchical clustering and the dynamic tree cut method (Langfelder and Horvath 2008). Modules with similar eigengenes were merged using a cut-off of 0.25, meaning modules with an overall similarity of 0.75 were merged. To compare modules between EH and MA environments, a module preservation analysis was performed (Langfelder et al. 2011) using EH as the reference and MA as the test for both maize and teosinte modules. Gene ontologies for each module in the maize and teosinte networks were calculated using AgriGo (https://bioinfo.cau.edu.cn/agriGO/). The top hub genes were identified for each module (Miller, Horvath, and Geschwind 2010) and visualized within the module using VisANT (Hu et al. 2004).

Enrichment analyses

We performed Gene ontology (GO) term enrichment analyses in AgriGo (https://bioinfo.cau.edu.cn/agriGO/), using a customized reference consisting of the genes expressed in leaf tissue according to our expression data in *parviglumis* or *mays*, depending on which subspecies was used for the enrichment analysis. GO terms of all differentially expressed genes were functionally classified into three major GO categories: molecular function (MF), biological process (BP) and cellular component (CC). Genes without GO terms were removed from the analysis. We identified significantly enriched GO terms using a Fisher's exact test and a p-value cut-off of ≤ 0.05 after applying the Yekutieli FDR correction. To test for enrichment between different categories of genes, we conducted Monte Carlo re-sampling, comparing the overlap of a particular category (e.g. teosinte-specific differentially expressed genes) with 10,000 equal-sized sets of randomly sampled genes expressed in leaf tissue (S1 File).

Additional data sets

We re-analyzed the data of Lemmon *et al*. (Lemmon et al. 2014), following their methods to identify candidate genes for differential expression between maize and teosinte. For categories included in the published data (Cis only, Cis + Trans), our reanalysis identified identical gene lists. In addition to these, we followed their filtering protocol to identify a list of top candidates in Trans only and Cis x Trans regulated genes. Because these data come from leaves at a different developmental stage from ours, we also ran analyses using ear and stem tissue from the same data to assess the robustness of our conclusions.

We used expression data from Hirsch and coauthors (Hirsch et al. 2014) to calculate the coefficient of variation of expression of 48,136 genes over 503 modern inbred lines of maize to compare them to our sets of genes. Finally, we included analysis of nucleotide diversity of genes in maize and teosinte, taken from Hufford *et al*. (Hufford, Xu, et al. 2012) and downloaded from https://figshare.com/articles/Gene_Popgen_Stats_from_Hufford_et_al_2012_Nat_Gen_/746968.

IV.4 Results

We grew four accessions of teosinte (*parviglumis*) and four inbred lines of domesticated maize in controlled environmental chambers simulating temperature and CO_2 conditions reflecting Early Holocene (EH) or Modern Ambient (MA) conditions. The largest difference in average temperature and average concentration of CO_2 within environmental treatments was 0.1°C and 3 ppmv respectively (see Methods). Many of the teosinte, particularly in the MA, had not developed inflorescences or complete branches at the time of harvest, preventing direct comparison of

inflorescence sexuality. Other phenotypic characteristics we observed were nonetheless consistent with our previous experiments under these conditions (Piperno et al. 2015), with teosinte plants grown in EH conditions exhibiting smaller stature and fewer axillary nodes—indicating fewer branches—than their counterparts grown in MA (S2 Table and S1 Fig). Maize grown in EH conditions was also smaller and less fecund than plants in MA conditions, but in contrast to teosinte grown in previous experiments (Piperno et al. 2015) we observed no variation in branching, inflorescence sexuality, or cob development, further indicating these traits are invariant in domesticated maize (S3 Table and S1 Fig).

To assess differences in gene expression plasticity between teosinte and maize, we sampled leaf tissue from 39 plants and extracted and sequenced total mRNA (see Methods). On average, we sampled 10 million reads per individual (S4 Table) and identified a total of 34,341 and 35,390 expressed genes in teosinte and maize, respectively, representing 87–90% of genes in the reference transcriptome. Analysis of differentially expressed (DE) genes under EH and MA conditions identified 3,953 and 3,355 DE genes in maize and teosinte at a false discovery rate (FDR) of 0.05 (Fig 1a; S1 File Maize_DE and Teosinte_DE). Many genes were differentially expressed in both taxa, and the observed 1,021 shared genes (Fig 1b) is significantly more than expected under a simple model of independence (p-value <1e-04).

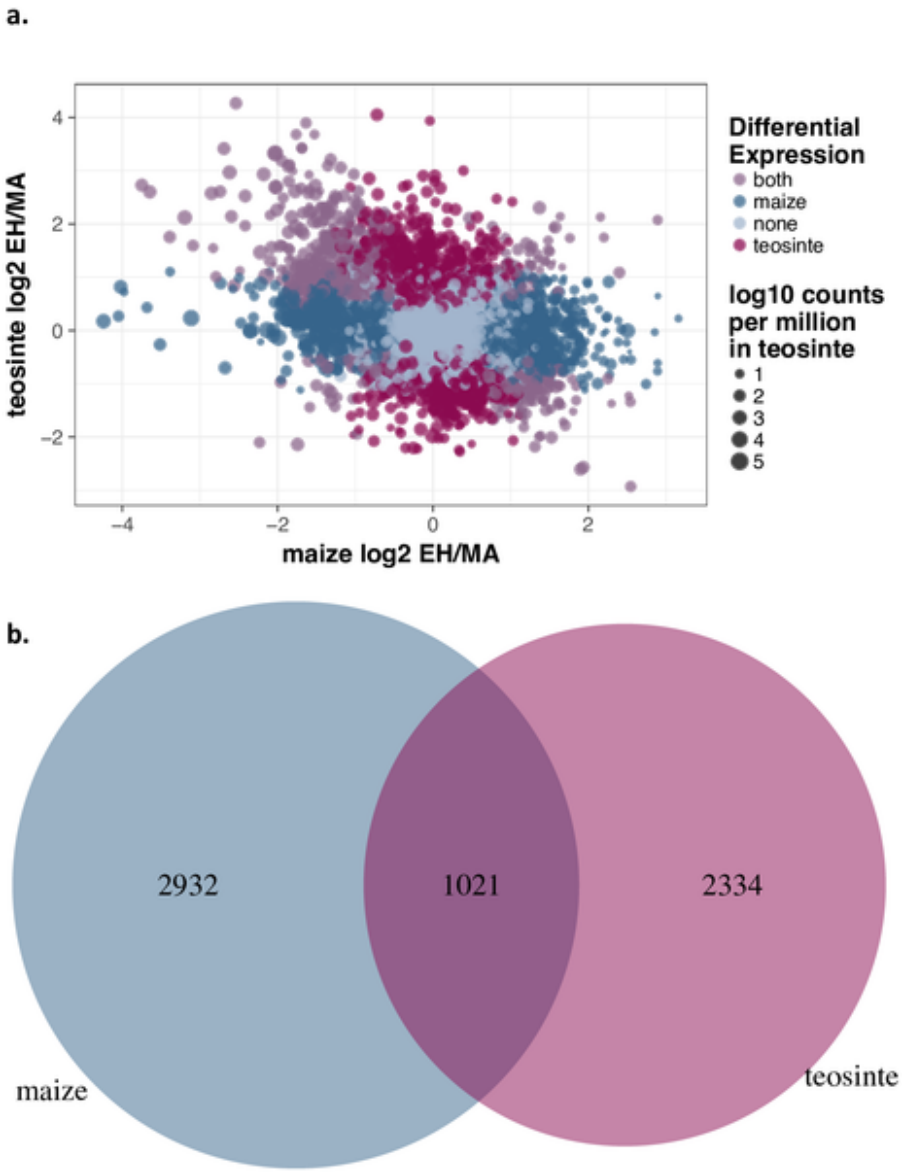


Fig 1. Differential expression in maize and teosinte under EH and MA conditions.

(a.) Categories of genes are shown in color (maize-specific DE genes in blue, teosinte-specific DE genes in red, shared DE genes in purple and non DE genes in gray), and point size represents the log mean counts per million in teosinte. (b.) Venn diagram of the overlap (purple), among DE genes of maize (blue) and teosinte (red) when exposed to the EH environment. https://doi.org/10.1371/journal.pone.0184202.g001

Co-expression analysis (see Methods) identified a total of 35 and 52 gene modules in maize and teosinte, respectively. Module preservation analysis indicated that gene networks were much

more highly conserved between MA and EH conditions in maize than in teosinte: while only 3% of modules showed no preservation in maize, over 35% were significantly changed in teosinte, indicating a much more labile co-expression response of teosinte to environment (Fig 2).

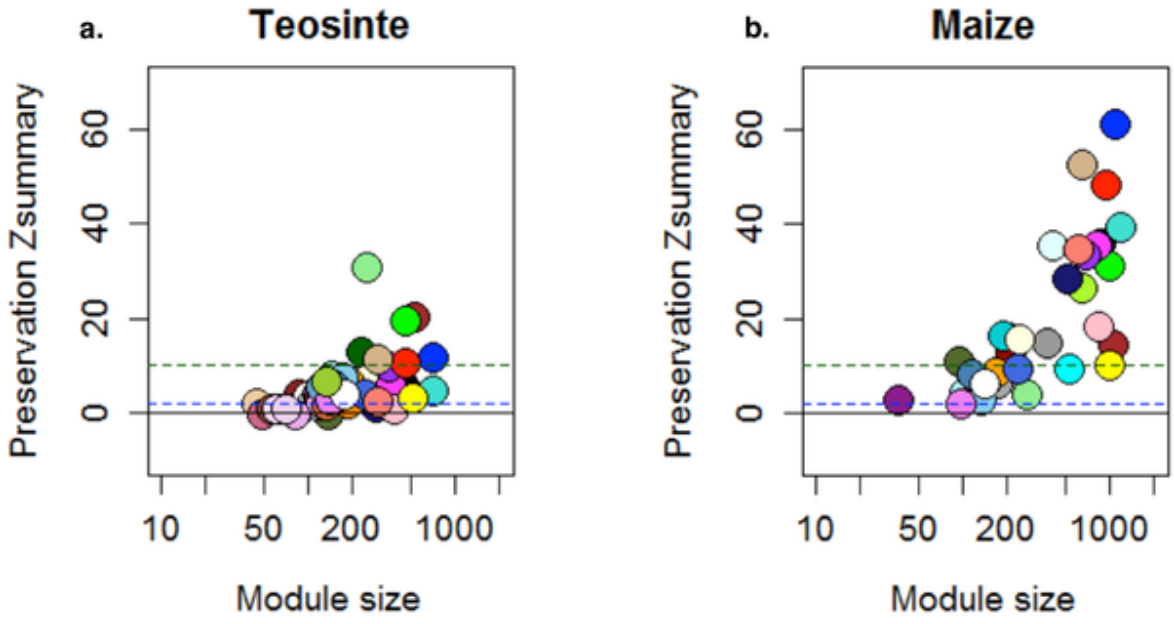


Fig 2. Module preservation in co-expression analysis.

WGCNA preservation scores for teosinte (a.) and maize (b.) modules across early Holocene and modern ambient environmental conditions. Modules with scores below 2 (blue dashed line) have no preservation across conditions, those between 2 and 10 (green dashed line) are moderately preserved, and those above 10 are highly preserved. https:/doi.org/10.1371/journal.pone.0184202.g002

We then investigated the role of selection during domestication in shaping the observed differences in expression across environments and between teosinte and maize by taking advantage of a number of published datasets. We first reanalyzed allele-specific expression data from Lemmon *et al*. (Lemmon et al. 2014) to generate lists of candidate genes with regulatory divergence between maize and teosinte in leaf tissue (see Methods). We identified sets of genes

differentially expressed in only one of the two taxa; we call these sets maize-specific and teosinte-specific DE genes. Both maize- and teosinte-specific DE gene sets were enriched for genes showing *cis*—but not *trans*—differences in expression between maize and teosinte (Fig 3). Genes differentially expressed in both maize and teosinte but in opposite directions were also similarly enriched in *cis* (p-value 0.029) but not *trans* (p-value 0.501), while shared DE genes showing similar direction in maize and teosinte were not enriched in any category.

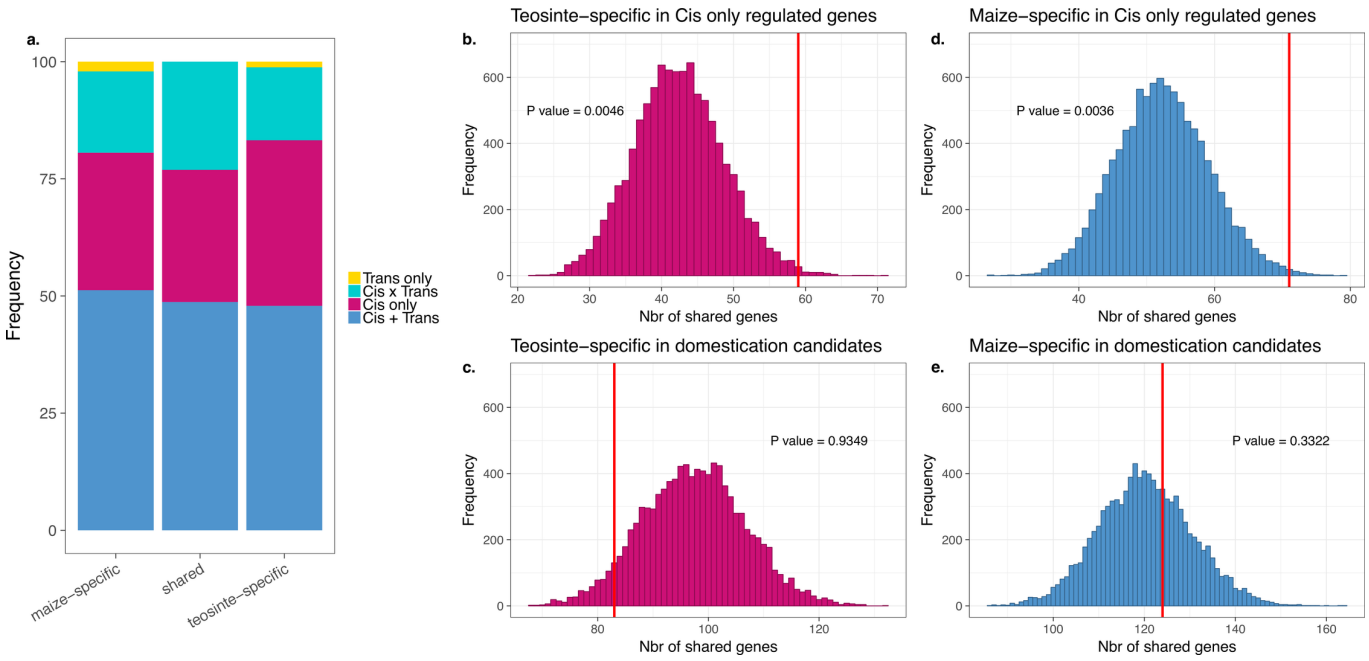


Fig 3. Overlap with domestication candidate genes.

(a) Patterns of expression shown as a proportion of genes differentially expressed between EH and MA conditions that are also differentially expressed between maize and teosinte. Monte Carlo re-sampling of DE genes in teosinte (b, c) and maize (d, e) for enrichment in genes showing *cis*-regulated differential expression between maize and teosinte (b, d) or evidence of selection during domestication (c, e). Maize and teosinte differential expression data are from Lemmon *et al*. (Lemmon et al. 2014), and selected gene lists are from Hufford *et al*. (Hufford, Xu, et al. 2012). https://doi.org/10.1371/journal.pone.0184202.g003

We next compared our set of taxon-specific DE genes (maize or teosinte -specific) to those showing evidence of selection during domestication (Hufford, Xu, et al. 2012), but found no evidence of enrichment for candidate loci (p-value >0.05 in all cases; Fig 3), and maize genes exhibit similar patterns of lower nucleotide diversity when compared to teosinte across both DE and non-DE genes (S3 Fig), consistent with overall patterns expected due to the demographic impacts of a domestication bottleneck (Hufford, Xu, et al. 2012). Results were similar when using data from ear or stem tissue as well, with the exception that teosinte-specific DE genes in our data also became enriched for *trans* differences when compared to expression from ear and stem tissues (p-value 0.0098 and 0.0135 for ear and stem, respectively). Finally, we asked whether taxon-specific DE genes show different patterns of variation in expression among modern maize lines. We find that both maize- and teosinte- specific genes show reduced variation in expression across a panel of more than 500 inbred lines (Hirsch et al. 2014), and teosinte-specific DE genes showed a small but statistically significant decrease in variation beyond that seen in maize-specific genes. (Fig 4).

We conducted GO enrichment analysis of both shared and taxon-specific DE genes (S1 File). DE genes shared between maize and teosinte are enriched in categories involved in photosynthesis, nitrogen and sugar synthesis, as well as response to stress, starvation or low phosphate conditions. Those unique to maize were mostly enriched in categories involved in photosynthesis, and these genes predominantly showed decreased expression in EH conditions; genes unique to maize also showed enrichment for biosynthesis categories. DE genes specific to teosinte were enriched for biological processes involving biosynthesis and metabolic pathways of numerous molecules including small molecules, amines, alcohols, sugars, amino acids, organic acids, and polyols. Of the few modules with co-expression showing changes in co-expression

patterns across environmental conditions in maize, one module showed enrichment for ontology classes related to membrane-bounded organelles. In contrast, modules changing co-expression in teosinte were enriched for diversity of ontology classes including phosphorus metabolism, protein kinase activity, organic and carboxylic acid biosynthesis, intracellular transport and localization, and amino acid ligase activity.

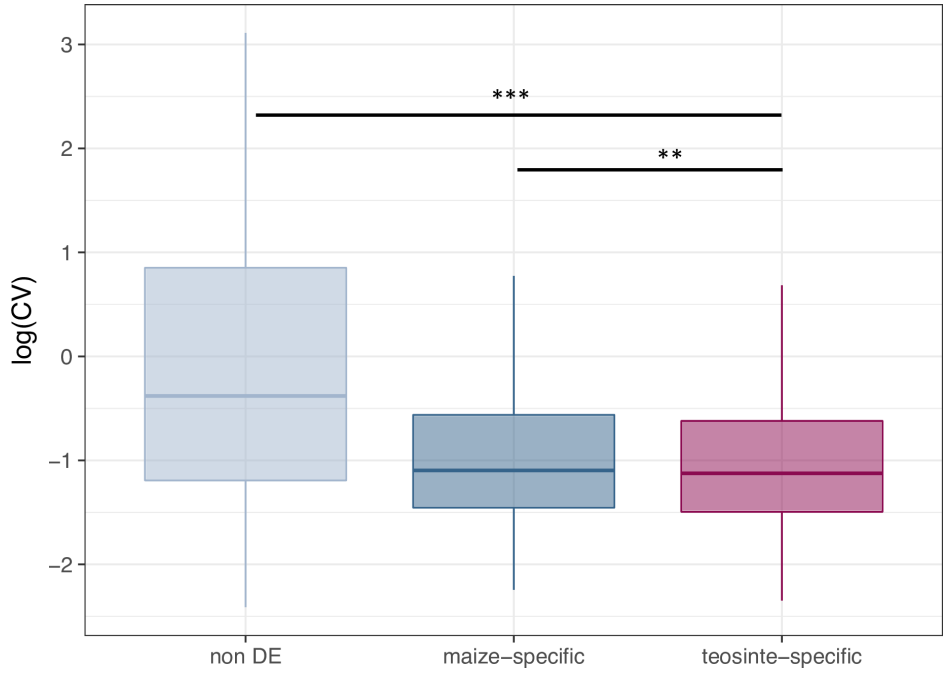


Fig 4. Box plot of the coefficient of variation.

Genes not differentially expressed are shown in gray, maize-specific DE genes in blue, and teosinte-specific DE genes in red. The significance of the Mann-Whitney U test is as shown with **<0.01, ***<0.001. https://doi.org/10.1371/journal.pone.0184202.g004

IV.5 Discussion

Phenotypic plasticity is a subject of growing importance in evolutionary biology (Ledón-Rettig, Pfennig, and Crespi 2010; Magalhaes et al. 2009; Pfennig and McGee 2010; Schneider et al. 2014) and recent research has shown that gene expression is key to understanding both plastic and adaptive responses of plants to varying environmental conditions (e.g., (Footitt et al. 2013; Des Marais, Hernandez, and Juenger 2013a; Munné-Bosch, Queval, and Foyer 2013)). Several studies have shown that selection on segregating genetic variation for environmentally-induced gene expression can decrease plasticity and result in constitutive expression and even the evolution of novel traits (Beldade, Mateus, and Keller 2011; Gilbert and Epel 2009; Pfennig and McGee 2010). This process of genetic assimilation has now been detailed in multiple taxa (Diggle and Miller 2013; Kuzawa and Bragg 2012; Standen, Du, and Larsson 2014) including in response to increased CO_2 (Walworth et al. 2016).

In this study we sought to evaluate the role of genetic plasticity in the evolution of maize during its domestication by growing both maize and its wild ancestor teosinte in environmental conditions reflecting both modern and ancient climates. Previous experiments had demonstrated dramatic phenotypic changes in teosinte when grown under ancient conditions, and our experiment found that nearly 10% of genes expressed in leaves are differentially expressed when grown in low temperature and CO_2 conditions reminiscent of the Early Holocene. A similar proportion of genes were also differentially expressed in maize, and the majority showed similar direction of change (Fig 1). Nonetheless, we saw much less change in overall modules of gene co-expression (Fig 2) and comparatively little change in plant morphology (S3 Table and S1 Fig).

Gene Ontology terms associated with shared and maize-specific DE genes reveal involvement in photosynthesis and are primarily down-regulated in the EH environment. Combined with GO-enrichment for stress-related genes across all candidates, these results suggest that decreases in temperature and CO_2 were likely stressful for both maize and teosinte, and we speculate that the stress associated with ongoing rapid climate change (Bassu et al. 2014) may lead to similarly significant changes in gene expression.

While many DE genes were shared between maize and teosinte, from the perspective of domestication those showing teosinte-specific expression are of most interest, as such genes are variable in the wild ancestor but appear canalized in domesticated maize. If genetic assimilation—selection on genetic changes that canalize a plastic response such as gene expression—played a predominant role genome-wide, we might expect to see the set of teosinte-specific DE genes enriched for genes previously identified as differentially regulated between maize and teosinte (Lemmon et al. 2014). While both maize- and teosinte- specific DE genes are enriched for genes showing cis-regulatory expression differences between maize and teosinte, this result is perhaps not surprising because taxon-specific DE genes were identified as genes with variable expression in one taxon and not the other. We thus expect *a priori* that these sets may have different cis-regulatory elements (and thus different response to experimental treatment) in maize and teosinte. For GA to play a genome-wide role in domestication, we also expect genes showing evidence of canalization in maize (teosinte-specific DE genes) to show population genetic evidence of selection. Instead, we find no enrichment for genes showing evidence of selection from genome scans (Hufford, Xu, et al. 2012) (Fig 3), and find that both maize- and teosinte-specific genes show decreased nucleotide diversity in maize (S3 Fig), likely the result of genetic drift during the maize domestication bottleneck. While the existing evidence does not support a genome-wide impact of

genetic assimilation, there are a number of reasons we might not observe such a pattern, including maladaptive plasticity (Sultan, Barton, and Wilczek 2009), selection on standing genetic variation (Barrett and Schluter 2008), and inbreeding during the development of modern maize lines.

Although GA may not have played a role genome-wide, our data hint at the possibility such a process may have been important for subsets of genes. For example, 83 teosinte-specific DE genes do show evidence of selection during domestication, and 6 of these have also been previously identified with a fixed regulatory difference between maize and teosinte (S1 File Teosinte_specific_in_domestication.xls). Moreover, a number of the differentially expressed genes we observed not identified as domestication candidates have previously been linked to morphological changes similar to those important for domestication—sometimes paralleling differences between maize and teosinte—and that were previously observed in our growth chamber experiments (Piperno et al. 2015; McSteen 2009; Kebrom, Spielmeyer, and Finnegan 2013; Gallavotti 2013; Hartwig et al. 2011; Lawit et al. 2010; Colebrook et al. 2014). These genes include various auxins; Brassinosteroids; a TCP transcription factor; gibberellin, absiccic acid (ABA), and cytokinin regulators; and genes implied in carbon and nitrogen fixation. Phenotypic attributes they may influence include vegetative architecture, inflorescence sexuality, plant height and biomass [e.g., (Piperno et al. 2015; McSteen 2009; Kebrom, Spielmeyer, and Finnegan 2013; Gallavotti 2013; Hartwig et al. 2011; Lawit et al. 2010; Colebrook et al. 2014)]. A relationship between sub-optimal conditions and plasticity in teosinte is in fact well known: poor growing conditions (shade, poor soils, crowding) induce plastic phenotypic response in teosinte that include suppression of branch elongation during growth (Doebley, Stec, and Gustus 1995; Hubbard et al. 2002; Whipple et al. 2011), resulting in plants with maize attributes in vegetative and inflorescence traits similar to those seen here and in previous experiments. This suggests that these and other DE

genes identified here may also lead to increased understanding of the maize domestication process by further informing the molecular basis of plasticity, phenotypic changes, and adaptation in past environments. Some genes were DE only in teosinte, suggesting genetic assimilation may have occurred. They include the following auxin and auxin response genes: *SAUR* 33 (GRMZM2G460861), auxin efflux carrier *PIN* 5a (GRMZM2G025742), *AUX_IAA* (GRMZM2G057067) and a *PAR* (GRMZM2G423863). Also with evidence of assimilation were *TCP* (TEOSINTE-BRANCHED1/CYCLOIDEA/PCF) transcription factor 44 (GRMZM2G089361), *ZOG* 3 (GRMZM2G338465), gibberellin and ABA regulators GRMZM2G301932 and GGRMZM2G150363, nitrate reductase *NADH* 1 (GRMZM2G568636) and *ferredoxin* 1 (GRMZM2G043162).

IV.6 Conclusion

Our experimental analysis of transcriptome change has identified a large number of genes showing differential expression in maize and teosinte when grown in environments reminiscent of the Early Holocene, the time period of maize domestication. We find greater changes in teosinte morphology and gene networks, and more than 2,000 genes showing differential expression only in teosinte, suggesting substantial loss of plasticity associated with maize domestication. To our knowledge, this is the first set of transcriptomic data showing evidence of a loss of plasticity linked to domestication. Though we find little evidence to support a genome-wide role of selection and genetic assimilation in patterning this loss of plasticity, we nevertheless identify a number of genes that show evidence of genetic assimilation including some linked to morphological changes related to domestication. Future studies should expand on the work presented here by investigating

additional environments (including modeled future climates) and providing more detailed, functional analysis of genes showing environmentally-induced plastic changes that may play important roles in patterning phenotypic variation in maize and teosinte.

IV.7 Acknowledgements

The authors wish to acknowledge funding support from the US National Science Foundation (IOS-0922703 and IOS-1238014), USDA (Hatch project CA-D-PLS-2066-H) and the Smithsonian Institution. We would like to acknowledge Milton Garcia for his technical assistance. We would also like to thank members of the Ross-Ibarra lab for comments on earlier drafts of the manuscript.

IV.8 References

Ahn, J. et al. 2004. "A Record of Atmospheric CO2 during the Last 40,000 Years from the Siple Dome, Antarctica Ice Core." *Journal of Geophysical Research: Atmospheres* 109(D13).

Alter, P., S. Bircheneder, L-Z. Zhou, U. Schlüter, M. Gahrtz, U. Sonnewald, and T. Dresselhaus. 2016. "Flowering Time-Regulated Genes in Maize Include the Transcription Factor ZmMADS1." *Plant Physiology* 172 (1). Am Soc Plant Biol:389–404.

Barrett, R. D H, and D. Schluter. 2008. "Adaptation from Standing Genetic Variation." *Trends Ecol Evol* 23 (1):38–44. https://doi.org/10.1016/j.tree.2007.09.008.

Bassu, S., N. Brisson, J-L. Durand, K. Boote, J. Lizaso, J. W. Jones, C. Rosenzweig, et al. 2014. "How Do Various Maize Crop Models Vary in Their Responses to Climate Change Factors?" *Glob Chang Biol* 20 (7):2301–20. https://doi.org/10.1111/gcb.12520.

Beadle, G. W. 1972. "Mystery of Maize." *Field Museum Natural History Bulletin*.

Beldade, P., A. R. A. Mateus, and R. A. Keller. 2011. "Evolution and Molecular Mechanisms of Adaptive Developmental Plasticity." *Mol Ecol* 20 (7):1347–63.

https://doi.org/10.1111/j.1365-294X.2011.05016.x.

Bomblies, K., and J. F. Doebley. 2006. “Pleiotropic Effects of the Duplicate Maize FLORICAULA/LEAFY Genes zfl1 and zfl2 on Traits under Selection during Maize Domestication.” *Genetics* 172 (1):519–31. https://doi.org/10.1534/genetics.105.048595.

Briggs, W. H., M. D. McMullen, B. S. Gaut, and J. Doebley. 2007. “Linkage Mapping of Domestication Loci in a Large Maize Teosinte Backcross Resource.” *Genetics* 177 (3):1915–28. https://doi.org/10.1534/genetics.107.076497.

Bush, M. B., A. Y Correa-metrio, D. A. Hodell, M. Brenner, F. S. Anselmetti, D. Ariztegui, A. D Mueller, et al. 2009. “Re-Evaluation of Climate Change in Lowland Central America during the Last Glacial Maximum Using New Sediment Cores from Lake PeténItzá, Guatemala.” In *Past Climate Variability in South America and Surrounding Regions*, 113–28. Springer.

Colebrook, E. H., S. G. Thomas, A. L. Phillips, and P. Hedden. 2014. “The Role of Gibberellin Signalling in Plant Responses to Abiotic Stress.” *J Exp Biol* 217 (Pt 1):67–75. https://doi.org/10.1242/jeb.089938.

Diggle, P. K., and J. S. Miller. 2013. “Developmental Plasticity, Genetic Assimilation, and the Evolutionary Diversification of Sexual Expression in Solanum.” *Am J Bot* 100 (6):1050–60. https://doi.org/10.3732/ajb.1200647.

Doebley, J, A Stec, and C Gustus. 1995. “Teosinte branched1 and the Origin of Maize: Evidence for Epistasis and the Evolution of Dominance.” *Genetics* 141 (1):333–46.

Doebley, J. 2004. “The Genetics of Maize Evolution.” *Annu Rev Genet* 38:37–59. https://doi.org/10.1146/annurev.genet.38.072902.092425.

Footitt, S., Z. Huang, H. A Clay, A. Mead, and W. E. Finch-Savage. 2013. “Temperature, Light and Nitrate Sensing Coordinate Arabidopsis Seed Dormancy Cycling, Resulting in Winter and Summer Annual Phenotypes.” *The Plant Journal* 74 (6). Wiley Online Library:1003–15.

Gallavotti, A. 2013. “The Role of Auxin in Shaping Shoot Architecture.” *J Exp Bot* 64 (9):2593–2608. https://doi.org/10.1093/jxb/ert141.

Gepts, P. 2012. *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*.

Cambridge University Press.

Gilbert, S. F, and D. Epel. 2009. “Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution.” *Sinauer Associates*.

Hartwig, T., G. S Chuck, S. Fujioka, A. Klempien, R. Weizbauer, D. P. V. Potluri, S. Choe, G. S. Johal, and B. Schulz. 2011. “Brassinosteroid Control of Sex Determination in Maize.” *Proc Natl Acad Sci U S A* 108 (49):19814–19. https://doi.org/10.1073/pnas.1108359108.

Hendry, A. P. 2016. “Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics.” *J Hered* 107 (1):25–41. https://doi.org/10.1093/jhered/esv060.

Hirsch, C. N., J. M. Foerster, J. M. Johnson, R. S. Sekhon, G. Muttoni, B. Vaillancourt, F. Peñagaricano, et al. 2014. “Insights into the Maize Pan-Genome and Pan-Transcriptome.” *Plant Cell* 26 (1):121–35. https://doi.org/10.1105/tpc.113.119982.

Hodell, D. A., F. S. Anselmetti, D. Ariztegui, M. Brenner, J. H. Curtis, A. Gilli, D. A. Grzesik, et al. 2008. “An 85-Ka Record of Climate Change in Lowland Central America.” *Quaternary Science Reviews* 27 (11). Elsevier:1152–65.

Hu, Z., J. Mellor, J. Wu, and C. DeLisi. 2004. “VisANT: An Online Visualization and Analysis Tool for Biological Interaction Data.” *BMC Bioinformatics* 5 (February):17. https://doi.org/10.1186/1471-2105-5-17.

Hubbard, L., P. McSteen, J. Doebley, and S. Hake. 2002. “Expression Patterns and Mutant Phenotype of Teosinte branched1 Correlate with Growth Suppression in Maize and Teosinte.” *Genetics* 162 (4):1927–35.

Hufford, M. B., E. Martínez-Meyer, B. S. Gaut, L. E. Eguiarte, and M. I. Tenaillon. 2012. “Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight.” *PLoS ONE* 7 (11). https://doi.org/10.1371/journal.pone.0047659.

Hufford, M. B, X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J-M. Chia, R. A Cartwright, R J Elshire, et al. 2012. “Comparative Population Genomics of Maize Domestication and Improvement.” *Nat Genet* 44 (7):808–11. https://doi.org/10.1038/ng.2309.

Kebrom, T. H., W. Spielmeyer, and E. J. Finnegan. 2013. “Grasses Provide New Insights into

Regulation of Shoot Branching." *Trends Plant Sci* 18 (1):41–48. https://doi.org/10.1016/j.tplants.2012.07.001.

Kuzawa, C. W, and J. M. Bragg. 2012. "Plasticity in Human Life History Strategy." *Current Anthropology* 53 (S6). JSTOR:S369--S382.

Langfelder, P., and S. Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9:559. https://doi.org/10.1186/1471-2105-9-559.

Langfelder, P., R. Luo, M. C. Oldham, and S. Horvath. 2011. "Is My Network Module Preserved and Reproducible?" *PLoS Comput Biol* 7 (1):e1001057. https://doi.org/10.1371/journal.pcbi.1001057.

Larson, G., D. R. Piperno, R. G. Allaby, M. D. Purugganan, L. Andersson, M. Arroyo-Kalin, L. Barton, et al. 2014. "Current Perspectives and the Future of Domestication Studies." *Proc Natl Acad Sci U S A* 111 (17):6139–46. https://doi.org/10.1073/pnas.1323964111.

Lawit, S. J., H. M. Wych, D. Xu, S. Kundu, and D. T. Tomes. 2010. "Maize DELLA Proteins Dwarf plant8 and Dwarf plant9 as Modulators of Plant Development." *Plant Cell Physiol* 51 (11):1854–68. https://doi.org/10.1093/pcp/pcq153.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J Carey. 2013. "Software for Computing and Annotating Genomic Ranges." *PLoS Comput Biol* 9 (8):e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

Ledón-Rettig, C. C, D. W Pfennig, and E. J. Crespi. 2010. "Diet and Hormonal Manipulation Reveal Cryptic Genetic Variation: Implications for the Evolution of Novel Feeding Strategies." *Proc Biol Sci* 277 (1700):3569–78. https://doi.org/10.1098/rspb.2010.0877.

Lemmon, Z. H., R. Bukowski, Q. Sun, and J. F. Doebley. 2014. "The Role of Cis Regulatory Evolution in Maize Domestication." *PLoS Genet* 10 (11):e1004745. https://doi.org/10.1371/journal.pgen.1004745.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biol* 15 (12):550. https://doi.org/10.1186/s13059-014-0550-8.

Magalhaes, I. S., S. Mwaiko, M. V. Schneider, and O. Seehausen. 2009. "Divergent Selection

and Phenotypic Plasticity during Incipient Speciation in Lake Victoria Cichlid Fish." *J Evol Biol* 22 (2):260–74. https://doi.org/10.1111/j.1420-9101.2008.01637.x.

Marais, D. L. Des, K. M. Hernandez, and T. E. Juenger. 2013. "Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment." *Annual Review of Ecology, Evolution, and Systematics* 44 (1):5–29. https://doi.org/10.1146/annurev-ecolsys-110512-135806.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. S. Buckler, and J. Doebley. 2002. "A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping." *Proc Natl Acad Sci U S A* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

McSteen, P. 2009. "Hormonal Regulation of Branching in Grasses." *Plant Physiol* 149 (1):46–55. https://doi.org/10.1104/pp.108.129056.

Miller, J. A., S. Horvath, and D. H. Geschwind. 2010. "Divergence of Human and Mouse Brain Transcriptome Highlights Alzheimer Disease Pathways." *Proc Natl Acad Sci U S A* 107 (28):12698–703. https://doi.org/10.1073/pnas.0914257107.

Moczek, A. P., S. Sultan, S. Foster, C. Ledón-Rettig, I. Dworkin, H. F. Nijhout, E. Abouheif, and D. W Pfennig. 2011. "The Role of Developmental Plasticity in Evolutionary Innovation." *Proc Biol Sci* 278 (1719):2705–13. https://doi.org/10.1098/rspb.2011.0971.

Munné-Bosch, S., G. Queval, and C. H. Foyer. 2013. "The Impact of Global Change Factors on Redox Signaling Underpinning Stress Tolerance." *Plant Physiol* 161 (1):5–19. https://doi.org/10.1104/pp.112.205690.

Pfennig, D. W, and M. McGee. 2010. "Resource Polyphenism Increases Species Richness: A Test of the Hypothesis." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365 (1540). The Royal Society:577–91.

Pigliucci, M. 2006. "Phenotypic Plasticity and Evolution by Genetic Assimilation." *Journal of Experimental Biology* 209 (12):2362–67. https://doi.org/10.1242/jeb.02070.

Piperno, D. R., and K. V. Flannery. 2001. "The Earliest Archaeological Maize (Zea Mays L.) from Highland Mexico: New Accelerator Mass Spectrometry Dates and Their

Implications." *Proc Natl Acad Sci U S A* 98 (4):2101–3. https://doi.org/10.1073/pnas.98.4.2101.

Piperno, D. R. 2011. "The Origins of Plant Cultivation and Domestication in the New World Tropics." *Current Anthropology* 52 (S4). JSTOR:S453--S470.

Piperno, D. R., I. Holst, K. Winter, and O. McMillan. 2015. "Teosinte before Domestication: Experimental Study of Growth and Phenotypic Variability in Late Pleistocene and Early Holocene Environments." *Quaternary International* 363. Elsevier:65–77.

Piperno, D. R., J. E. Moreno, J. Iriarte, I. Holst, M. Lachniet, J. G. Jones, A. J. Ranere, and R. Castanzo. 2007. "Late Pleistocene and Holocene Environmental History of the Iguala Valley, Central Balsas Watershed of Mexico." *Proceedings of the National Academy of Sciences* 104 (29). National Acad Sciences:11874–81.

Restrepo, A., P. Colinvaux, M. Bush, A. Correa-Metrio, J. Conroy, M. R. Gardener, P. Jaramillo, M. Steinitz-Kannan, and J. Overpeck. 2012. "Impacts of Climate Variability and Human Colonization on the Vegetation of the Galápagos Islands." *Ecology* 93 (8):1853–66.

Schlichting, C. D., and M. A. Wund. 2014. "Phenotypic Plasticity and Epigenetic Marking: An Assessment of Evidence for Genetic Accommodation." *Evolution* 68 (3):656–72. https://doi.org/10.1111/evo.12348.

Schneider, R. F., Y. Li, A. Meyer, and H. M. Gunter. 2014. "Regulatory Gene Networks That Shape the Development of Adaptive Phenotypic Plasticity in a Cichlid Fish." *Mol Ecol* 23 (18):4511–26. https://doi.org/10.1111/mec.12851.

Standen, E. M., T. Y. Du, and H. C. E. Larsson. 2014. "Developmental Plasticity and the Origin of Tetrapods." *Nature* 513 (7516):54–58. https://doi.org/10.1038/nature13708.

Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. "Identification of a Functional Transposon Insertion in the Maize Domestication Gene tb1." *Nature Genetics* 43 (11):1160–63. https://doi.org/10.1038/ng.942.

Sultan, S. E., K. Barton, and A. M Wilczek. 2009. "Contrasting Patterns of Transgenerational Plasticity in Ecologically Distinct Congeners." *Ecology* 90 (7):1831–39.

Swanson-Wagner, R., R. Briskine, R. Schaefer, M. B. Hufford, Je. Ross-Ibarra, C. L Myers, P.

Tiffin, and N. M. Springer. 2012. “Reshaping of the Maize Transcriptome by Domestication.” *Proc Natl Acad Sci U S A* 109 (29):11878–83. https://doi.org/10.1073/pnas.1201961109.

Vollbrecht, E., P. S. Springer, L. Goh, E. S. Buckler, and R. Martienssen. 2005. “Architecture of Floral Branch Systems in Maize and Related Grasses.” *Nature* 436 (7054):1119–26. https://doi.org/10.1038/nature03892.

Waddington, C. H. 1953. “Genetic Assimilation of an Acquired Character.” *Evolution*. JSTOR, 118–26.

Walworth, N. G., M. D Lee, F-X. Fu, D. A. Hutchins, and E. A. Webb. 2016. “Molecular and Physiological Evidence of Genetic Assimilation to High CO2 in the Marine Nitrogen Fixer Trichodesmium.” *Proc Natl Acad Sci U S A*, November. https://doi.org/10.1073/pnas.1605202113.

Wang, H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, M. Faller, K.Bomblies, L. Lukens, and J. F. Doebley. 2005. “The Origin of the Naked Grains of Maize.” *Nature* 436 (7051):714–19. https://doi.org/10.1038/nature03863.

West-Eberhard, M. J. 2003. *Developmental Plasticity and Evolution*. Oxford University Press.

Whipple, C. J., T. H. Kebrom, A. L Weber, F. Yang, D. Hall, R. Meeley, R. Schmidt, J. Doebley, T. P. Brutnell, and D. P. Jackson. 2011. “Grassy tillers1 Promotes Apical Dominance in Maize and Responds to Shade Signals in the Grasses.” *Proc Natl Acad Sci U S A* 108 (33):E506-12. https://doi.org/10.1073/pnas.1102819108.

Wills, D. M., C. J. Whipple, S. Takuno, L. E Kursel, L. M. Shannon, J. Ross-Ibarra, and J. F. Doebley. 2013. “From Many, One: Genetic Control of Prolificacy during Maize Domestication.” *PLoS Genet* 9 (6):e1003604. https://doi.org/10.1371/journal.pgen.1003604.

Wright, S. I., I. V. Bi, S. G Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. “The Effects of Artificial Selection on the Maize Genome.” *Science (New York, N.Y.)* 308 (5726):1310–14. https://doi.org/10.1126/science.1107891.

Wu, T. D., J. Reeder, M. Lawrence, G. Becker, and M. J. Brauer. 2016. “GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality.”

Methods Mol Biol 1418:283–334. https://doi.org/10.1007/978-1-4939-3578-9_15.

Zhong, S., J-G. Joung, Y. Zheng, Y. Chen, B. Liu, Y. Shao, J. Z. Xiang, Z. Fei, and J. J. Giovannoni. 2011. "High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation." *Cold Spring Harb Protoc* 2011 (8):940–49. https://doi.org/10.1101/pdb.prot5652.

IV.9 Supporting information

USDA accession ID	**Lat/Lon**	**Origin**
PI 384062, pop1	17.417 N, -99.5 W	Valle de Bravo, Mexico State
PI 384063, pop2	18.83 N, -100.16 W	Mexico, Mexico State
PI 384071, pop3	18.33 N, -100.31 W	Iguala, Guerrero State
PI 566692, pop4	19.06 N, -100.41 W	Zitacuaro, Michoacán State
Ames 19288, Oh43	NA	Ohio United States
PI 550473, B73	NA	Iowa United States
NSL 30053, W22	NA	Wisconsin United States
PI 558532, Mo17	NA	Missouri United States

S1 Table. Sources of the teosinte and maize seeds. https://doi.org/10.1371/journal.pone.0184202.s001

All vegetative biomass (g)		Plant height (cm)		# Nodes	
EHC	**MCC**	**EHC**	**MCC**	**EHC**	**MCC**
241.2 ± 114.9	265.3 ± 95.8	141.8 ± 57.6	239.9 ± 48.3	15 ± 4.7	22 ± 4.5

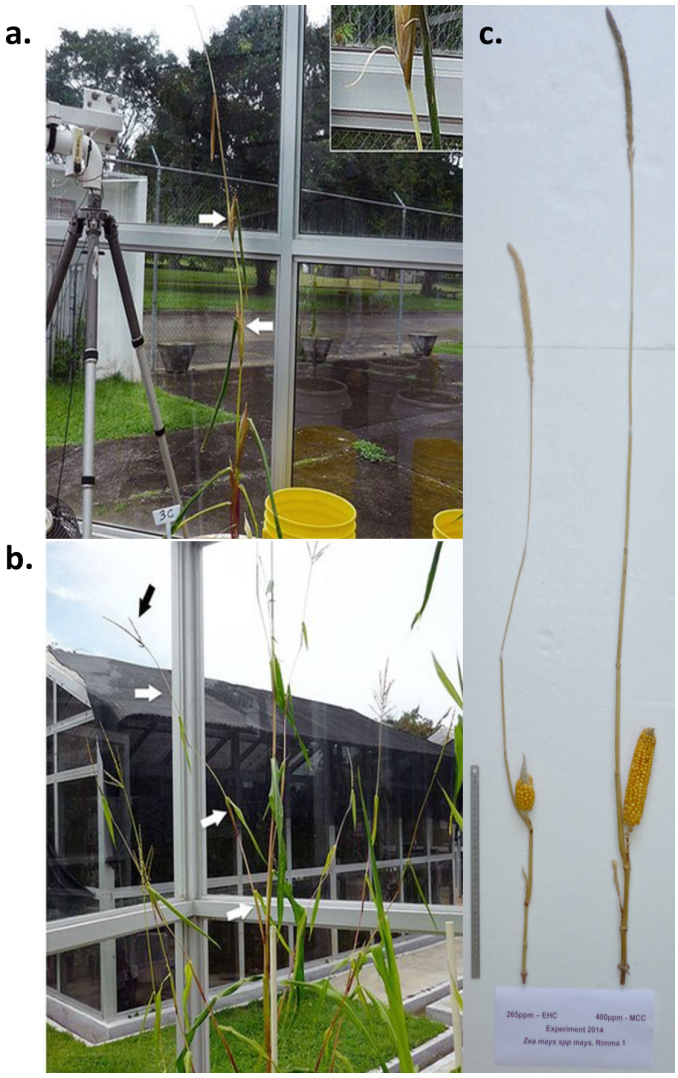
S2 Table. Teosinte phenotypes in 2013 experiment. https://doi.org/10.1371/journal.pone.0184202.s002

Vegetative biomass (g)		Cobs (g)		Total Biomass (g)		Height (cm)	
EHC	**MCC**	**EHC**	**MCC**	**EHC**	**MCC**	**EHC**	**MCC**
55.0 ± 18.3	69.5 ± 17.0	39.3 ± 14.9	64.5 ± 11.7	94.3 ± 32.6	134.0 ± 25.1	140.6 ± 19.3	176.4 ± 15.5

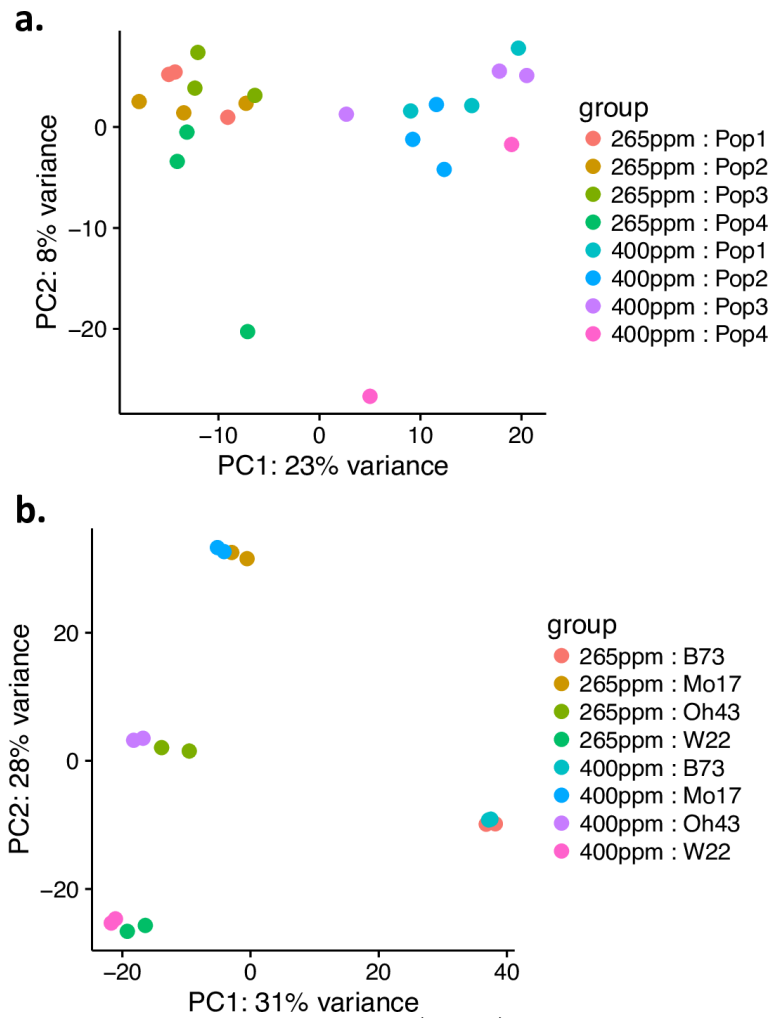
S3 Table. Maize phenotypes in 2014 experiment. https://doi.org/10.1371/journal.pone.0184202.s003

Sample	Reads	Maize-like mom
265ppm_pop1.A.1	8643790	X
265ppm_pop1.A_2.1	7648528	
265ppm_pop1.B.1	5031069	
265ppm_pop2.A.1	6732943	
265ppm_pop2.B.1	5063618	X
265ppm_pop2.B_2.1	6467938	
265ppm_pop3.A.1	4399187	
265ppm_pop3.B.1	5564170	
265ppm_pop3.C.1	4139946	X
265ppm_pop4.A.1	7427625	
265ppm_pop4.B.1	3814875	
265ppm_pop4.C.1	4746629	X
400ppm_pop1.A.1	4070011	X
400ppm_pop1.A_2.1	5989185	
400ppm_pop1.B.1	7570271	
400ppm_pop2.A.1	7882249	
400ppm_pop2.B.1	5812095	X
400ppm_pop2.B_2.1	7455893	
400ppm_pop3.A.1	8630267	
400ppm_pop3.B.1	5433424	
400ppm_pop3.C.1	5029499	X
400ppm_pop4.A.1	5857687	
400ppm_pop4.C.1	8836575	X
265ppm_Oh43.A	14416915	
265ppm_Oh43.B	13978009	
265ppm_B73.A	15911016	
265ppm_B73.B	14202484	
265ppm_W22.A	16932095	
265ppm_W22.B	15442245	
265ppm_Mo17.A	14819357	
265ppm_Mo17.B	14669370	
400ppm_Oh43.A	14114563	
400ppm_Oh43.B	14914449	
400ppm_B73.A	20128181	
400ppm_B73.B	18545881	
400ppm_W22.A	16442158	
400ppm_W22.B	15726969	
400ppm_Mo17.A	16940428	
400ppm_Mo17.B	17291888	

S4 Table. Number of reads per sample. Plants from a maternal source with a maize-like phenotype in a previous experiment are marked. https://doi.org/10.1371/journal.pone.0184202.s004

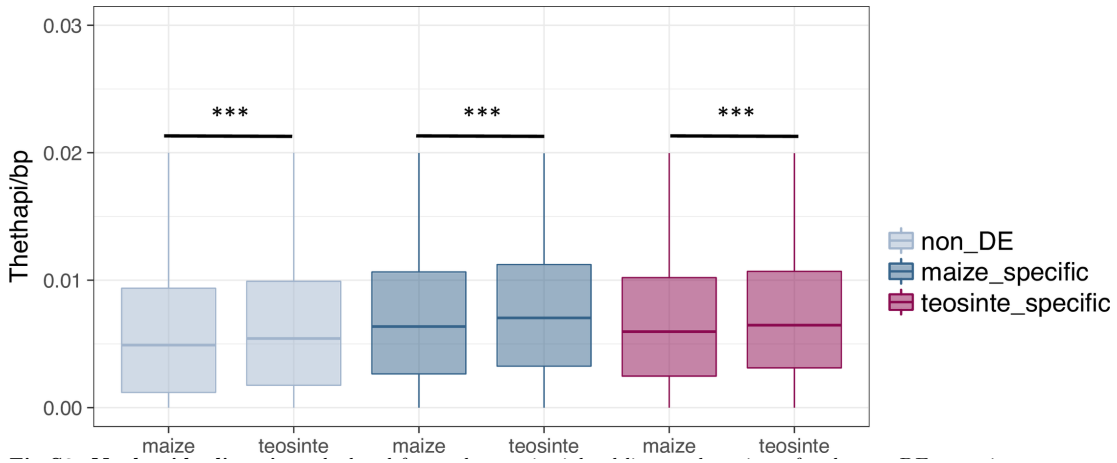


S1 Fig. Examples of phenotypic differences in EH chamber on the top and MA chamber on the bottom for teosinte ((a.) and (b.) from Piperno and coauthors (Piperno et al. 2015))and maize (c.). The teosinte plant in the EH chamber is a maize-like phenotype in vegetative architecture, inflorescence sexuality, and seed maturation, as described in the main text. The plant in the MA chamber is typical of teosinte today in those characteristics. These traits are unaltered for the maize plant between the EH chamber on the left and the MA chamber on the right. https://doi.org/10.1371/journal.pone.0184202.s005



S2 Fig. Principal component analysis (PCA) using rlog-normalized of the expression data for the principal components 1 (PC1) and PC2, for teosinte (a.) and maize (b.).

https://doi.org/10.1371/journal.pone.0184202.s006



S3 Fig. Nucleotide diversity calculated for modern maize inbred lines and teosintes for the non-DE genes in gray, maize-specific genes in blue and the teosinte-specific DE genes in red. Mann-Whitney U tests for all comparisons are significant (***P <0.001). https://doi.org/10.1371/journal.pone.0184202.s007

S1 File. Supplemental data files. Available online. Flowering_genes.csv, Go_terms.xls, Maize_DE.csv, Maize_readcounts.csv, Teosinte_DE.csv, Teosinte_readcounts.csv and Teosinte_specific_in_domestication.xls. https://doi.org/10.1371/journal.pone.0184202.s008

V. BOTH HARD AND SOFT SWEEPS CONTRIBUTE TO LOCAL ADAPTATION OF NATURAL POPULATIONS OF TEOSINTES

Anne Lorant[1,2], Maud Tenaillon[2] and Jeffrey Ross-Ibarra[1,3,4]

[1] : Dept. of Plant Sciences, University of California, Davis, CA 95616, USA.

[2]: Génétique Quantitative et Evolution – Le Moulon, Institut National de la Recherche agronomique, Université Paris-Sud, Centre National de la Recherche Scientifique, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, 91190, France.

[3] : Center for Population Biology, University of California, Davis, CA 95616, USA.

[4]: Genome Center, University of California, Davis, CA 95616, USA.

V.1 Introduction

Genome-wide datasets open new avenues to characterize the demographic and selective processes that shape natural variation in species. Methodological advances in population genomics have allowed the identification of putative adaptive traits and the mining of candidate loci via genome scans. First applied to species-wide samples (Akey et al. 2002; Payseur, Cutter, and Nachman 2002), these methods have more recently served to detect adaptation locally by comparing allele frequencies among populations (Eckert et al. 2009) and to associate allele frequencies to local ecological variation (Fournier-Level et al. 2011). Collectively, they are referred to as reverse ecology.

Local adaptation occurs because the environment is heterogeneous in space and time, and local conditions determine traits favored by natural selection. It results in populations more adapted, hence with higher fitness in their native environment, compared to non-native populations (Kawecki and Ebert 2004). Numerous studies based on reciprocal transplants between environments and common gardens, (i.e. when populations are grown in a common environment) in combination with fitness measures have shown that local adaptation is widespread in plants (Hereford 2009). Examples include flowering time in *Arabidopsis lyrata* (Leinonen, Remington, and Savolainen 2011) or seed size in wild barley (Volis, Mendlinger, and Ward 2002).

In the last decade, *Zea mays* ssp. *parviglumis* have emerged as useful model for dissecting long-term adaptation (Hufford, Bilinski, et al. 2012). While these plants are geographically limited to Mexico, their distributions span numerous environmental conditions, including a broad range of temperatures, precipitation levels and elevations. These factors, coupled with limited migration due to the complex landscape of Mexico (Pyhäjärvi et al. 2013) and a large effective population size (*Ne*) (Ross-Ibarra, Tenaillon, and Gaut 2009), allowed for extensive local adaptation of which

signatures persist in contemporary genomes. Only few studies investigated local adaptation in *parviglumis*. From these studies, we learned that candidate loci associated with environmental variables are mostly involved in flowering time adaptation to soil and defense loci against herbivory (Pyhäjärvi et al. 2013; Aguirre-Liguori et al. 2017; Fustier et al. 2017; Moeller and Tiffin 2008). They also showed that four large inversions play an important role in local adaptation (Fang et al. 2012; Pyhäjärvi et al. 2013) and loci involved in local adaptation are enriched in non-genic regions (Fustier et al. 2017; Pyhäjärvi et al. 2013).

During local adaptation, natural selection shifts allele frequencies over time in populations, moving traits toward phenotypic optima favored by local biotic and abiotic conditions. When an environmental change alters the selective optimum, a population can adapt in three ways. First, a single de-novo beneficial mutation may arise after the onset of selection and sweeps through the population, leaving the discernable footprint of a hard genetic sweep (Hermisson and Pennings 2017). This mechanism shifts the beneficial allele of the target site and linked neutral variants i.e. the haplotype containing the beneficial variant, to high frequency. Second, neutral variants segregating in the population (standing genetic variation) before the onset of selection, may become advantageous and be selected. Third, recurrent mutations within the same locus arising on different haplotypes may occur in the population during the course of selection. In contrast to the first scenario, these two last scenarios leave soft sweep signatures, where multiple haplotypes sweep in the population (Hermisson and Pennings 2017). This results in less reduction of diversity when compared to hard sweeps, making them more difficult to identify. However, a significant mass of empirical and theoretical evidence has accumulated over the past ~10 years which support soft sweeps as a frequent mode of adaptation in many populations (Cutter and Payseur 2013;

Messer and Petrov 2013; Pritchard, Pickrell, and Coop 2010; Schrider and Kern 2017; Sheehan and Song 2016).

Although previous work on teosintes found extensive local adaptation, most of them used a 55k SNPs array data (Aguirre-Liguori et al. 2017; Pyhäjärvi et al. 2013) and calculated outliers on differentiation of allele frequencies between subpopulation (F_{ST}) associated with environmental variables. Methods based on SNP arrays place strong assumptions on both the number and scale of adaptive sites. One recent study, however, used whole genome sequencing at very low depth and employed an haplotype-based method for 47 candidate regions, defined by more classical outlier detection. Haplotype analysis revealed interesting features including signatures of soft sweeps and some degree of convergence for pairs of populations sampled at similar altitudes (Fustier et al. 2017).

In the present study, we assessed the relative contribution of hard and soft sweeps and the degree of convergence at a whole genome scale by relying on high depth coverage (20-25x) of 60 individuals from 6 distinct populations. In order to avoid confusion between teosinte subspecies differentiation and local adaptation, we also focused solely on *Zea mays* ssp. *parviglumis*, the closest wild relative of cultivated maize. Using populations from differing environments, we determined the geographic scale of local adaptation in *parviglumis*, the demographic processes that have affected the efficacy of selection, and phenotypic proxies that selection may have acted on.

V.2 Results and discussion

We sequenced at high depth (20-25X), 60 *Zea mays* ssp. *parviglumis* (thereafter *parviglumis*) individuals from 6 populations covering the geographical range of the subspecies (Figure 1A). We identified a quality filtered set of 105,109,679 SNPs. We used principal components analysis on genotypes to determine the population structure. Along the first 3 Principal Components (PCs), individuals from each of the six populations formed six distinct clusters, which pointed to mark population structure (Fig. 1A and B). This was further confirmed by very limited signal of admixture at K=6 using the software Admixture (Fig. 1C). Overall, projection of principal components onto a 2D (latitude, longitude) geographic map, closely matched the locations of the sampled populations (Fig 1A), consistent with isolation by distance. Interestingly, when the elevation was accounted for as an additional dimension, projected genetic distances from El Rodeo and Crucero Lagunitas populations better co-localized with geographical coordinates of the corresponding populations (figure 1B). This indicated that specific altitudes of these two populations (Fig. 1D) may play a role in their isolation.

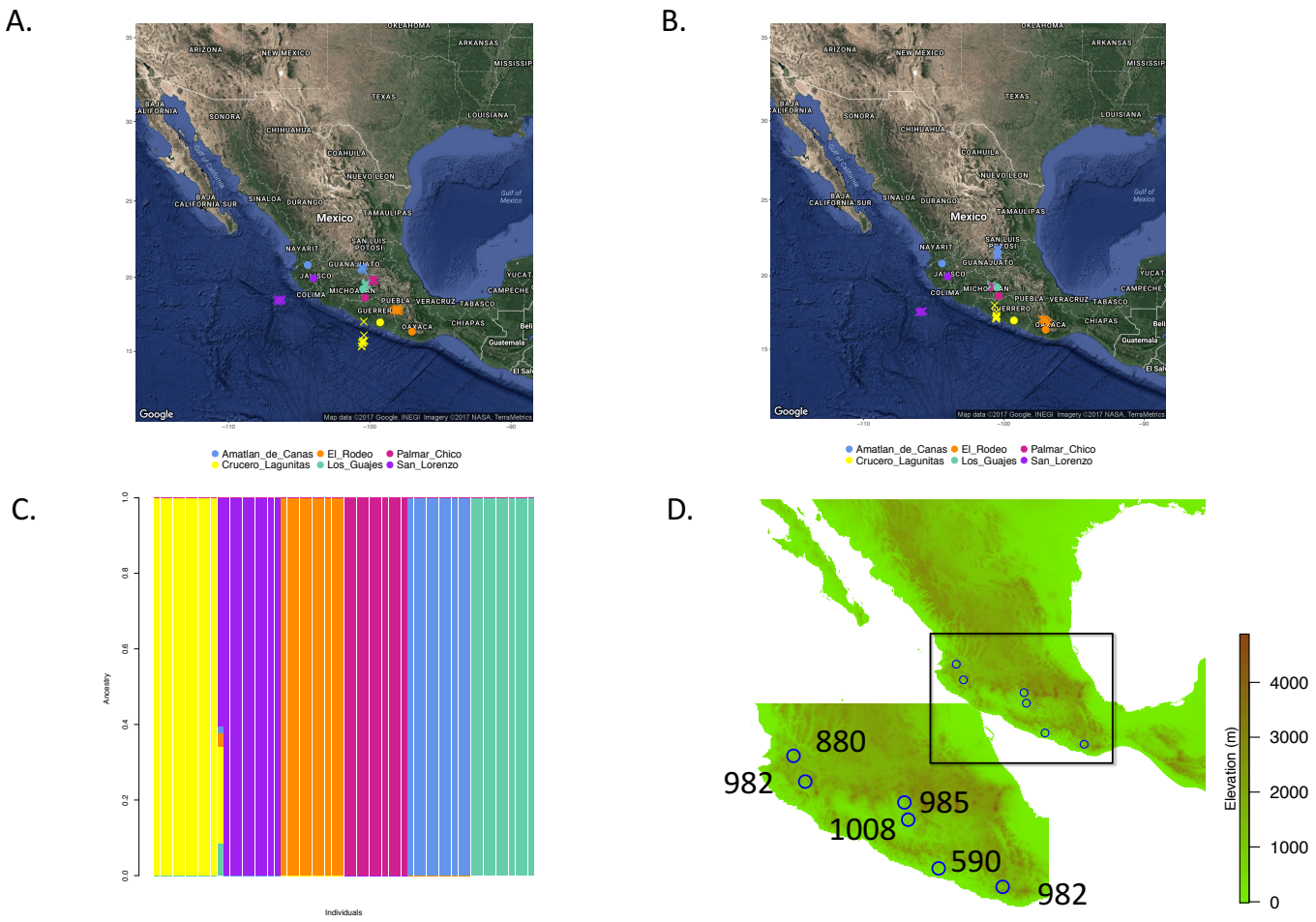


Figure 1: Population structure of 6 *parviglumis* populations originating from Mexico

(A) Projection of the first three genotype PC scores onto a 2D geolocalization map (longitude, latitude) and **(B)** 3D geolocalization map (longitude, latitude and altitude). Dots represent geographical coordinates of 6 *parviglumis* populations and crosses denotes the projected score for each individual. **(C)** Genomic admixture with 6 founding populations. **(D)** Elevation map of the 6 populations.

Both variant and invariant sites were utilized to calculate standard population genetic metrics, such as π, θ, Tajima's D and pairwise F_{ST} (see material and methods).

To calculate the unfolded site frequency spectrum (uSFS), we oriented SNPs with two outgroups, *Z. diploperennis* and *T. dactyloides,* with an estimated divergence time from

parviglumis of 500,000 and 1M years, respectively (Ross-Ibarra, Tenaillon, and Gaut 2009). Incomplete lineage sorting makes inference of ancestral and derived alleles difficult. To mitigate this issue, we added an extra outgroup and used the method described in Keightley et al. 2016, which corrected the estimation of high frequency variants in our data. Interestingly, we found that Palmar Chico and Los Guajes, the populations the closest from the geographic center, defined as median longitude and latitude of the number of 329 occurrence records (Aguirre-Liguori et al. 2017; Table 1), have an excess of rare alleles, consistent with a recent population expansion (Fig. 2A). This observation was further supported by genome-wide Tajima's D distributions shifted toward negative values (Fig.2B). Other populations in contrast displayed a uSFS containing fewer singletons and Tajima's D values closer to zero, which is the expectation under neutral equilibrium (Fig2A-B).

We further examined Tajima's D distribution on a per-chromosome basis. In Crucero Lagunitas, although the population is not deviating from neutrality, with an average Tajima's D value close to zero, the median Tajima's D values of chromosome 6 is negative (Mann-Whitney *U* test p-value < 2.2e-16). This is consistent with positive selection acting on that particular chromosome as demographic processes would impact each chromosome equally (Fig2-C, yellow population). Previously identified inversions in teosintes on chromosome 1 and 9 are associated with environmental adaptation (Fang et al. 2012; Pyhäjärvi et al. 2013), and the sweeping of one of these large regions could explain the negative value of a whole chromosome (Andersen et al. 2012). We investigated linkage disequilibrium (LD) patterns in each population to test for the presence of inversions. Inversions create large LD blocks where consecutive alleles are transmitted together. Although we identified numerous blocks of LD in our data set, chromosome 6 in Crucero Lagunitas did not show any long stretch of LD, making it unlikely that there is a recent inversion.

Further analyses and experiments will be required to elucidate how selection acted to generate this pattern on chromosome 6.

We also assessed pairwise F_{ST} values for each pairs of populations, which revealed that the population from Balsas form a group with a F_{ST} of 0.04 (Fig.S1). These populations (Los Guajes and Palmar Chico) showed the highest diversity, a negative median value of inbreeding and relatively short runs of homozygosity (ROHs) compared with other populations (Fig.S2 and S3), as one would expect from large outcrossing populations. In contrast, other populations had a lower diversity as found in San Lorenzo and El Rodeo, respectively between 1.3 and 1.54 times lower than the Balsas populations, and long ROHs consistent with their higher level of inbreeding (Fig. S2 and S3).

Populations	Distance from Geographic Centroid	Distance from Niche Centroid
Los Guajes	0.60	1.41
Crucero Lagunitas	2.22	6.77
El Rodeo	4.36	1.94
Amatlan de Canas	4.28	4.16
San Lorenzo	3.53	3.05
Palmar Chico	0.36	2.13

Table 1: Distance from the Geographic Centroid and the Niche Centroid as describe in Aguirre-Liguori et al. 2017

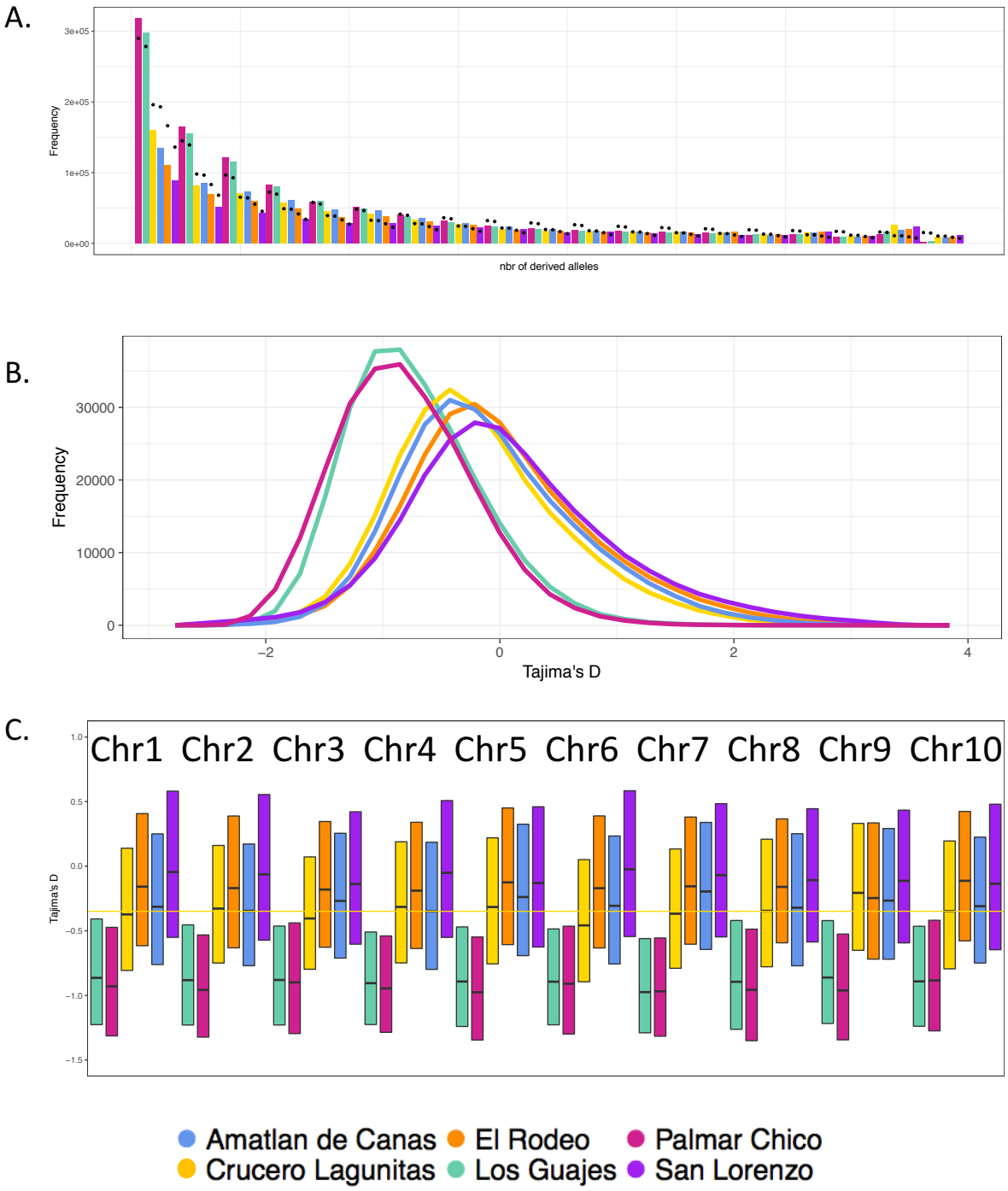


Figure 2: unfolded SFS and Tajimas'D

(A) Unfolded site frequency spectrum (uSFS) with two outgroups (*Zea diplo-perennis* and *Tripsacum dactyloides*). Black dots represent expected SFS **(B)** Distribution of Tajima's D statistic calculated on

10kb non-overlapping windows along the genome **(C)** Boxplot of Tajima'D value per chromosome (1 to 10) for each population, the yellow line represents the median Tajima's D value for Crucero Lagunitas.

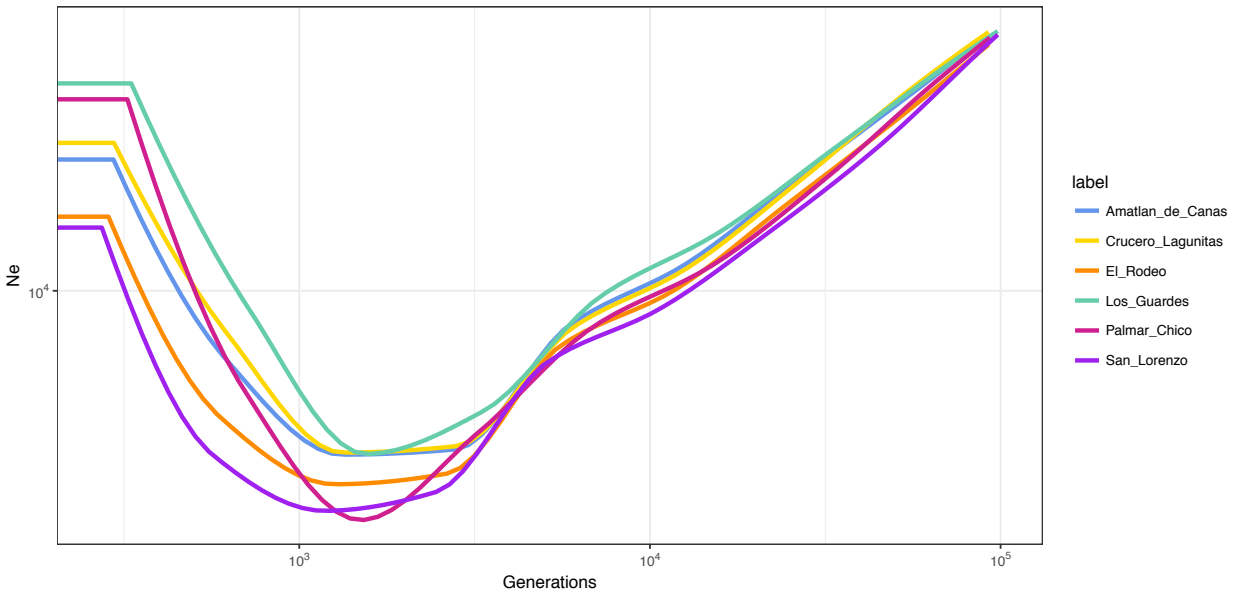


Figure 3: Demography

Estimation of the populations size (*Ne*) over time of each population with a generation time of 1 year and a mutation rate of $3x10^{-8}$.

We inferred historical population size changes (*Ne*), using the SMCPP software (Terhorst, Kamm, and Song 2017). This software infers demographic histories from unphased data using coalescent HMMs that make use of SFS and LD. All populations showed a gradual decrease in Ne between 100,000 and 1,500 years before present. This decrease in population size up until the recent past is then followed by a rapid population expansion. Although all populations experienced a bottleneck followed by an expansion phase, the Balsas River populations (Los Guajes and Palmar Chico) expanded to the largest current size. Note that this graph should be taken with a grain of salt due to the limitation of the algorithm to generate simulations reflecting the population genetic parameters measured on the populations (see VI. Discussion and Perspectives).

To investigate selective forces on the genome, we plotted the SFS for synonymous vs non-synonymous sites for each population (Fig. S5-S10). We observed an excess of low frequency non-synonymous variants in the Balsas populations consistent with purifying selection keeping deleterious variant at low frequency (Fu 1997). In the other populations, the lack of excess low frequency non-synonymous variants suggests that selection is less effective in these populations (Fig. S6-S9) and was the most extreme for the San Lorenzo population (Fig S9).

To assess positive selection across the entire genome including non-coding sequences, we quantified the number of hard sweeps in each population. Genetic sweeps are characterized by a reduction in diversity around an allele under selection caused by linkage, which triggers a shift of the SFS toward high and low frequencies in the sweep region. To identify such regions, we conducted genome-wide scans using SweeD (Pavlidis et al. 2013) on non-overlapping windows of 10kb. This program utilizes a composite likelihood ratio (CLR) to estimate the probability of observing the SFS within a sweep region given a neutral model estimated from the SFS observed across the whole chromosome. It should be noted that this method is relatively robust to demographic changes in population size (Nielsen et al. 2005). Regions under selection were identified by CLR values exceeding a threshold computed using simulations conducted with the ms software under neutrality (Hudson 2002) (see Material and Methods section for details). We identified numerous regions under selection in the populations (Fig.4).

To investigate the contribution of genic versus non-genic variants in identified hard sweeps, we compared the percentage of genic base pairs in sweep regions compare to the whole genome percentage (Table 2). This test revealed that four of the six populations showed lower percentage of genic regions compare to the whole genome average, one population was close to the genome percentage and one showed enrichment in genic regions. This also showed that

between 86 and 97% of hard sweeps are located in intergenic regions which is consistent with previous observations (Fustier et al. 2017; Pyhäjärvi et al. 2013). However, the different level of genic regions probably reflects different traits and genetic architecture selected on and should be further investigated.

	Percentage of genic regions in hard sweeps
Los Guajes	5.24 %
Crucero Lagunitas	6.25 %
El Rodeo	2.32 %
Amatlan de Canas	13.32 %
San Lorenzo	5.95 %
Palmar Chico	3.88 %
Whole genome	6.89 %

Table 2: Percentage of genic regions in hard sweep regions and the whole genome

We then quantified the overlap of selective sweeps across different populations and found that selective regions are, for the most part, unique to each population, suggesting considerable local adaptation (Fig.4). Moreover, the number of hard sweeps varied considerably across populations, with Los Guajes and Palmar Chico populations harboring the greatest number of hard sweeps (1499 and 1803, respectively), and San Lorenzo and El Rodeo the lowest number (246 and 405 respectively). Populations for which purifying selection is the most pervasive (Fig. S5-S10) are also the ones harboring the highest number of hard sweeps. While this could result from the model's inability to disentangle signatures of purifying and positive selection, this is unlikely because most sweeps are in intergenic regions.

Inferred past niches, as modelled in Hufford *at al.* 2012 (Hufford, Martínez-Meyer, et al. 2012), show that populations from San Lorenzo, El Rodeo, Amatlan de Canas and Crucero Lagunitas are recent arrivals to their locations while the environment in Palmar Chico and Los

Guajes have been static since the last glacial maximum (~21,000 years before present). Based on these results that suggest large difference in time for adaptation, we investigated the number of sweeps relative to niches and geographic centers. Interestingly, we found that the number of sweeps strongly negatively correlates with the geographic distance from the center of the niche (Pearson r^2=0.75; Fig S11 and S12). Moreover, the median level of inbreeding as well as the number of ROHs (runs of homozygosity) are also strongly positively correlated with the geographic distance from the center of the distribution (Pearson r^2=0.68 and 0.29, respectively Fig.S13 and S16), further reinforcing that isolation by distance is important for structuring genetic diversity in *parviglumis*. This suggests that higher inbreeding and drift associated with a restricted gene flow in these populations with smaller *Ne* (Fig.4) contribute to a reduction in diversity compared to populations from the center of the geographic distribution. This is in agreement with similar patterns observed in maize in the first chapter (Wang et al. 2017). Notably, we observed a negative correlation between the median cumulative length of ROHs and the number of hard sweeps per population suggesting that the signal we are picking-up is due to selection and not a reduction of diversity due to inbreeding (Pearson r^2=0.56).

Interestingly, when the species is considered as a whole, i.e. with all populations analyzed together, only a total of 34 sweeps are observed which is perhaps not too surprising given the low overlap between sweep regions, identified at the population-level, across populations. This suggests that they are locally adapted via distinct mechanisms to different environments. It would be interesting to further investigate biological processes impacted by these sweeps since they may be important for general adaptation in *parviglumis* species. This might also explain why other studies found only a few fixed differences separating maize and teosinte and hard sweeps not contributing to genome-wide patterns of diversity in maize (Beissinger et al. 2016). Our results

indicate that in *parviglumis*, strong population structure and local adaptation most likely drive this observed pattern. Most of the adaptation appears at a local scale while only few loci act at the level of the whole subspecies.

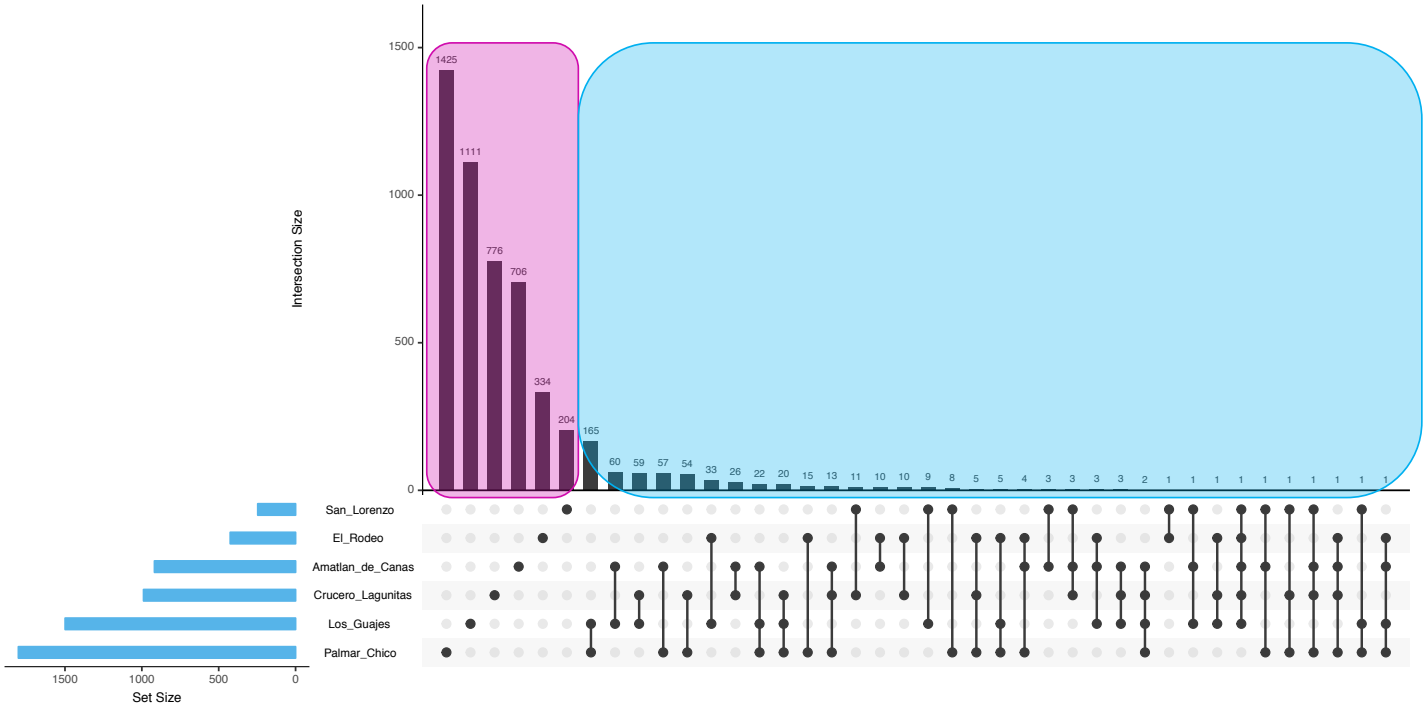


Figure 4: Hard sweeps.
Plot summarizing the number of selective sweeps private (circled in purple or shared between populations circled in blue). Filled circles denote a population, lines represent an overlap between populations. The vertical bar plot shows the number of regions per groups while the horizontal bar plot represents the total number of sweeps per populations.

Next, we used the 4000 RILs of the maize NAMs panel's public phenotype and genotype data (Bukowski et al. 2015; Wallace et al. 2014) to quantify how much additive genetic variation for various phenotypes can be explained by sweep regions using a residual maximum likelihood model as implemented in the LDAK software (Speed et al. 2012).

The sweep regions explained up to 40% of the heritability across 41 traits. Interestingly, sweep regions disproportionately contribute to phenotypic traits that may have been the targets of

natural selection, including fructose (involved in cold tolerance) (Bogdanović et al. 2008), fumarate (involved in multiple processes including nitrogen assimilation, pollen and seed germination) (Araújo, Nunes-Nesi, and Fernie 2011), glutamate (involved in root response to organic nitrogen in the soil) (Forde and Lea 2007) and plant height.

We further investigated the presence of soft sweeps in the populations. Soft sweeps appear when multiple haplotypes are sweeping at the same time in the population due to selection on standing variation or multiple mutations in the same locus (Hermisson and Pennings 2017). For this we used the H12 method that scans for haplotypes under selection in the genome (Garud et al. 2015). Interestingly, we found that the Balsas populations, which contained the largest number of hard sweeps, showed the lowest number of soft sweeps (Fig.5). Inversely, San Lorenzo and El Rodeo populations, which contained the lowest number of hard sweeps, are the population with the highest number of significant windows identified by the H12 scan, with 10 to 14 times more windows identified when compared to Palmar Chico and Los Guajes, respectively (Fig.5). The Balsas populations showed the least overlap with about 0.4 to 1% shared soft sweep nucleotides with other populations (Fig.S17). Crucero Lagunitas, El Rodeo, Amatlan de Canas and San Lorenzo had many more soft sweeps in common, with an estimated overlap of 11 to 14% which is about the maximum overlap in hard sweeps (10% between Los Guajes and Palmar Chico; Fig.4, S17). Although it should be noted that the H12 scan is possibly sensitivity to demography and bottlenecks these result in a diminution of rare haplotypes in the populations and lead to false positives. This motivates our decision to utilize machine learning models which can detect both hard and soft sweeps while correcting for demography (see VI. Discussion and Perspectives). It should also be noted, that these results contradict our observations in the introduction chapter, where hard sweeps and soft sweeps are respectively observed in the populations with the lowest

and the highest *Ne*. Although, this is most likely do to other factors playing an important role during local adaptation as strength of selection and time for adaptation. Indeed, it was theoretically showed that after an environmental change or a habitat expansion, adaptation from standing variation is prevalent for alleles of small effect. This effect is enhanced when the environmental change is followed by a bottleneck (Hermisson and Pennings 2005). Moreover, soft sweeps for large effect alleles, are also expected during such scenario as standing genetic variation is directly available and new mutations need time to arise (Hermisson and Pennings 2005). This point out that *Ne*, while an important factor influencing soft versus hard sweeps, is not the unique contributor to the genomic landscape shaping natural populations.

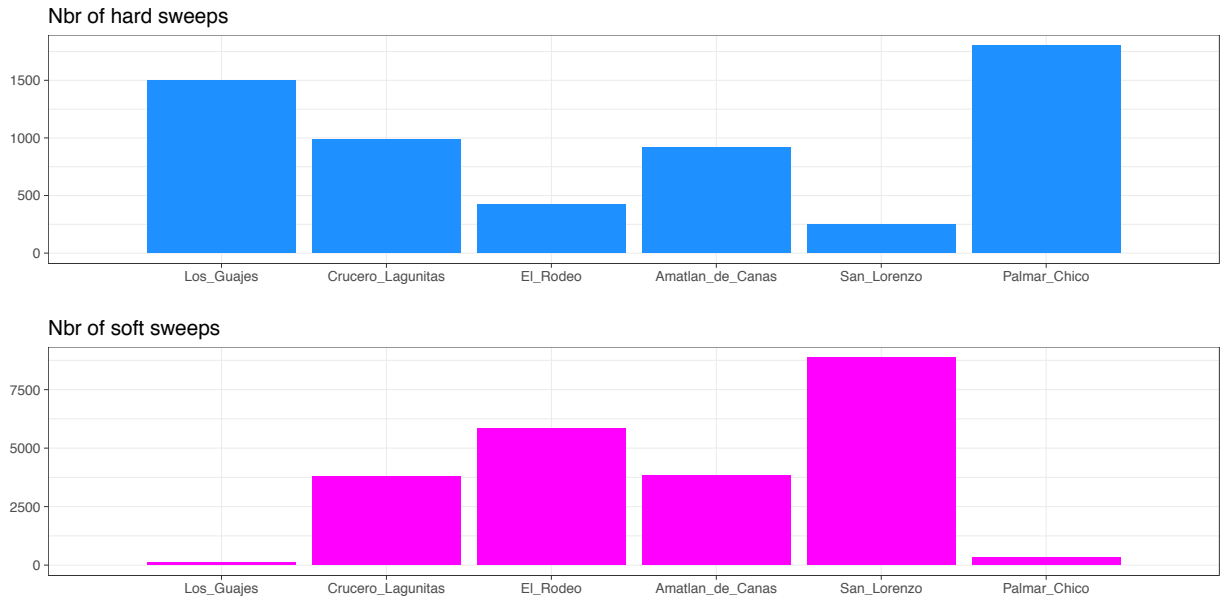


Figure 5: Total number of hard and soft sweeps.

Total number of hard sweeps per population in blue (top panel) and total number of soft sweeps per populations in pink (bottom panel).

Similarly, to our findings with hard sweeps, it appears that most soft sweeps are unique to populations, likely as a response to particular environmental conditions. This further strengthens evidence pointing to strong local adaptation in *parviglumis.* These differences in number of soft and hard sweeps between populations are in agreement with the histories of these populations. Indeed, the Balsas Valley populations are established since a longer time at the same location, while other populations had to adapt from the glaciation refuge to new conditions during a shorter time frame of ~20,000 years before present and therefore harbor more soft sweeps. When we investigated the patterns of soft sweeps across the whole subspecies, we observed a value that lies in the lower half of the range observed for single populations. Although the pattern of soft sweeps is reduced when the subspecies is analyzed as a whole, there are significantly more soft sweeps than hard sweep (30x more). This suggests some level of parallel adaptation and further investigations of convergent evolution is warranted.

V.3 Conclusion

While the genetics of maize domestication and differences between maize (*Zea mays* ssp. *mays*) and its ancestor, teosinte (*Zea mays* ssp. *parviglumis*) have been extensively studied and understood, little is known about the evolution of teosinte in natural populations (Beadle 1972; John Doebley 2004; Fukunaga et al. 2005). Although, p*arviglumis* is a good model to study local adaptation since it harbors high genetic diversity due to a large effective population size (Ross-Ibarra, Tenaillon, and Gaut 2009) and is found in various habitats (Hufford, Bilinski, et al. 2012). To better understand the evolutionary forces shaping genetic diversity in teosinte, we sequenced, with high deep coverage, the genomes of 60 wild-collected individuals from six distinct natural

populations spanning much of the growing range of teosinte in southern Mexico. This dataset enables us to investigate simultaneously sequence diversity and selection both globally across populations, and locally within populations.

Adaptation of the genome under positive selection as long been seen as exclusively arising from single beneficial mutations increasing in frequency in the population, a phenomenon known as a hard sweep. However, this scenario is unlikely to occur as evidences from quantitative genetic studies showed that adaptation is often fast and involves both standing genetic variation and recurrent mutations (as discussed in the introduction chapter), or in other terms, can involve both hard and soft sweeps.

Parviglumis is a spatially structured subspecies with limited migration (Pyhäjärvi et al. 2013), it therefore follows that adaptation had to mostly occur locally. Under this scenario, we expect to observe high rates of hard sweeps within single populations, and soft sweep signals across the global species (Messer and Petrov 2013). Our results are in agreement with this expectation as we observed a considerable number of hard sweeps when the populations are considered separately, while only a few when all the samples are analyzed together. The same pattern held for soft sweeps, although the number of soft sweeps for the whole subspecies is 30 times higher than the number of observed hard sweeps. Interestingly the different number of hard and soft sweeps among populations suggest that both selective processes are important during local adaptation and most likely reflect different time of adaptations. Finally, the low overlap between soft sweeps or hard sweeps among populations further support the importance of local adaptation in *parviglumis*.

V.4 Material and methods

Samples and whole genome re-sequencing

We obtained from USDA seeds for five populations of *Zea mays* ssp. *parviglumis*, and for the species *Tripsacum dactyloides* and *Zea diplo-perennis*. The seeds from the sixth populations (from Palmar Chico) came from the personal collection of John Doebley. The *parviglumis* populations spanned its natural range in southern Mexico (Figure 1). For *parviglumis*, ten individuals from each population were grown, while one individual of *Tripsacum dactyloides* and *Zea diplo-perennis* was grown. Genomic DNA was extracted from leaf tissue using the E.Z.N.A.® Plant DNA Kit (Omega Biotek), following manufacturer's instructions. DNA was quantified using Qubit (Life Technologies). 1ug of DNA per individual was fragmented using a bioruptor (Diagenode) with 30 seconds on/off cycles. DNA fragments were then prepared for Illumina sequencing. First, DNA fragments were repaired with the End-Repair enzyme mix (New England Biolabs). A deoxyadenosine triphosphate was added at each 3'end with the Klenow fragment (New England Biolabs). Illumina Truseq adapters (Affymetrix) were then added with the Quick ligase kit (New England Biolabs). Between each enzymatic step, DNA was washed with sera-mags speed beads (Fisher Scientific). The libraries were sequenced with an average coverage of 20 to 25x PE150 on the Xten at Novogene (Sacramento, USA). *Zea diplo-perennis* and *Tripsacum dactyloides* where sequenced with a coverage of 10 to 20x PE250 on 3 lanes of Illumina 2000 rapid run (UC Davis Genome Center, Davis, USA).

Read mapping and SNP calling

BWA-MEM v. 0.7.9a (H. Li and Durbin 2010) was used to map reads to the maize B73 reference genome v4 (Jiao et al. 2017), with the -M option enabled for compatibility with Picardtools. Duplicate reads were removed with the MarkDuplicates implementation in Picardtools v. 2.6.0 (https://broadinstitute.github.io/picard/). The resulting BAM files were locally realigned using IndelRealigner in GATK v. 3.6-0 (Depristo et al. 2011) and regions to be realigned were identified with RealignerTargetCreator. SNPs and indels were called in the cohort with the HaplotypeCaller from GATK creating two files containing variant and invariant sites (with the option --includeNonVariantSites).

SNPs were filtered after quality control with the VariantFiltration GATK across all samples (QD < 2.0; FS > 60.0; MQ < 40.0; MQRankSum < -12.5; ReadPosRankSum < -8.0). We retained only Biallelic SNPs (Bcftools https://samtools.github.io/bcftools/bcftools.html) with fewer than 20% missing genotypes over the 60 *parviglumis* individuals. Finally, we excluded SNPs that departed from Hardy-Weinberg equilibrium (HWE) within each population with VCFtools (Danecek et al. 2011). To determine the appropriate p-value threshold for the HWE test, we first performed a Hardy-Weinberg equilibrium exact test on the whole set of SNPs in PLINK 1.9 (Purcell et al. 2007) and plotted a subsample of one million SNPs at varying HWE thresholds (0.001, 0.01 and 0.1); We then determined the optimal p-value threshold to be 0.1 by minimizing the Euclidean distance between the expected and observed genotype frequencies at HWE on a QQ plot of the expected genotype frequencies at HWE versus the observed.

For *Tripsacum dactyloides* and *Zea diplo-perennis*, SNPs were called as described above but without quality filtering for missing data and HWE since each VCF file contained only one individual.

Population genetics parameters

Standard metrics of population genetics, which provide insights into selection processes, differentiation between populations and demographic histories, were calculated within 10Kb windows using VCFtools. These included Tajima's D, pairwise nucleotide diversity (π) and pairwise F_{ST}. After computing these metrics, the number of mapped base pairs, including both variant and invariant sites, where calculated for each window using a custom Python script. Windows containing less than 15% sequenced nucleotides were excluded when estimating those parameters. Because the estimation of π is sensitive to the number of available sites within each window, we adjusted π for the number of available base pairs per window using a custom python script.

Linkage disequilibrium decay provides information about the average haplotype length of the population. Specifically, a higher decay rate implies shorter haplotypes compared to a slower one.

To estimate LD decay in our populations, we computed the pairwise Pearson correlation coefficient (r^2) between SNPs in non-overlapping windows of 1 Mb with using PLINK 1.9. Note that we set the option -ld-window-r2 to zero to report all pairwise comparisons. We then plotted r^2 against the physical distance between each pair of SNPs and fitted the data using a decreasing exponential model. The LD analysis showed a rapid decay occurring within 300bp reflecting the outcrossing nature of *parviglumis* (Fig.S18). Large LD blocks were identified visually by plotting r^2 values along chromosomal positions.

Individual inbreeding coefficients were calculated using the method-of-moments estimator (MME) as previously described in Ritland 1996. This method estimates relatedness between individuals using all variant sites in the dataset.

To calculate runs of homozygosity (ROHs), variants were down-sampled to a single site per 2kb window in PLINK 1.9 within windows containing at least 20 SNPs, a maximum of 2 missing SNPs and one heterozygous site.

Genetic structure

To analyze population structure and genetic diversity, we first pruned SNPs based on LD to obtain a set of uncorrelated sites. For this, we excluded correlated SNPs (Pearson correlation coefficient > 0.185) within sliding windows of 1Kb, with a 100bp step-size using PLINK 1.9. We computed population structure between the 60 individuals in our sample, using principal component analysis (PCA) on the genotype implemented in PLINK 1.9. The three first principal components (PCs) were projected onto the geographical coordinates of our populations using a rigid transformation with a custom script in R. Briefly, we rotated, scaled and translated, minimizing the mean squared error between projected principal components and population geographical coordinates were found using a singular value decomposition of the covariance matrix as described in Umeyama et al., 1991.

The same set of SNPs was then used to evaluate the population structure admixture using the software Admixture (Alexander, Novembre, and Lange 2009).

To select the number of clusters (K), we ran a cross-validation procedure and found no differences in likelihood for K between 2 and 10. We therefore resorted to using 6 clusters (K=6), as this corresponded to the number of populations.

Site frequency spectrums

SNPs were polarized based on *T. dactyloides* and *Z. diplo-perennis* genomes to determine the frequency of derived sites for each population of *parviglumis*, using the method described in Keightley *et al.* (Keightley et al. 2016). The input file was generated using a custom python script. Only the SNPs shared between the focal population and the 2 outgroups were kept. Shared SNPs were divided into groups of 20, 18 and 16 alleles observed in the focal population to calculate each uSFS, as recommended in Keightley *et al.*. Site frequencies were then combined.

We annotated synonymous and non-synonymous sites in VCF files using SNPeff (Cingolani et al. 2012). These annotated sites were then extracted to generate folded SFS in each population and for each class.

Demography

smcpp (Terhorst, Kamm, and Song 2017) was used to infer changes in historical effective population sizes in the 6 populations of *parviglumis*. A mutation rate (number of mutations per nucleotide per year) of 3E-8 and a generation time of 1 year was used. The analysis was conducted for each population separately using all individuals. The analysis was conducted on the whole set of variant sites with masked centromeres and with both inclusion or exclusion of invariant sites since the program utilizes the haplotypes and the physical distance between them. We observed similar results between results obtained using analyses that included or excluded invariant sites.

Hard sweeps and soft sweeps

Sweed3.0 (Pavlidis et al. 2013) was used to detect candidate hard sweeps in the *parviglumis* populations. This software uses a composite likelihood ratio (CLR), as implemented in Sweepfinder (Nielsen et. al), which identifies regions of the genome with significant deviations from the neutral SFS of the chromosome. The null hypothesis relies on the SFS across the whole chromosome instead of a standard neutral model, making it more robust to demographic events such as recent expansion. The number of grids per chromosome was determined such as to obtain windows of 10kb. The program was run to detect both population specific hard sweeps and, on all individuals, together to detect species-wide selective sweeps.

To determine the significance of the data, 10,000 simulations of 20kb windows were performed under a standard neutral model using the ms program (Hudson 2002). For simulations, we used a mutation rate (θ of 0.014/bp, calculated on the data as described in Watterson 1975. The simulated regions were run in Sweed3.0 with a grid number of 2 to match our original analysis that used 10kb windows. We used the 95th percentile of CLR values obtained from the neutral analysis to declare significance.

To investigate soft sweeps, we used the haplotype homozygosity scan H12 as implemented in SelectionHapStats (Garud et al. 2015). This metric detects regions with two common haplotypes segregating at high frequency in the population. To do so, first the genomes were phased using Eagle2 (Browning and Browning 2011) and an unpublished recombination map of teosinte constructed from GBS data of ~5000 progenies of *parviglumis* out of 50 parents from Palmar Chico (unpublished data). The VCF format was parsed, using a custom python script, for compatibility with the software. For each chromosome, each line in the parsed file started with the sorted coordinate of the chromosome followed by alleles for each individual (A, T,G,C or N if

missing). Analyzes were performed using sliding windows of 400 sites with steps of 50 sites. The median value of H12 was used as a threshold and values resulting from one haplotype were excluded. We estimated the overlap of soft sweeps between populations by counting the number of shared base pairs in each window. To compare soft sweeps and hard sweeps, we concatenated and divided soft sweeps regions into 10kb regions to match the detection method for hard sweeps.

Percentage of genic regions in hard sweeps

To quantify the overlap between hard sweep regions unique to populations and genic regions, we calculated the overlap between hard sweep regions with all the genes in the V4 reference genome (using only the first transcript), including intron, exons, 3' and 5' UTRs and compared it to the whole genome. Both values were calculated as percentages of total bp.

Modeling heritabilities explained by selective sweeps

To assess how variant sites in selective sweep regions contribute to various phenotypic traits, we used published phenotypic and genotypic data from the maize Nested Association Mapping (NAM) population (Bukowski et al. 2015; Swarts et al. 2014; Wallace et al. 2014). Briefly, we estimated, for each selective sweep, the additive genetic variance explained by kinship matrices created from sets of SNPs contained within the region. For each of these SNP data sets, we generated a square kinship matrix, using the scaled IBS method implemented in TASSEL 5 (Bradbury et al. 2007). We then used the Restricted Maximum Likelihood (REML) model implemented in LDAK (Speed et al. 2012) to jointly estimate heritability explained by SNPs in hard sweep regions compared to the rest of the genome. To determine the significance of explained

heritabilities, we built the null distribution by repeating the analysis on randomly sampled regions of equal size compared to detected hard sweeps.

V.5 References

Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. "Connecting Genomic Patterns of Local Adaptation and Niche Suitability in Teosintes." *Molecular Ecology* 26 (16):4226–40. https://doi.org/10.1111/mec.14203.

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver. 2002. "Interrogating a High-Density SNP Map for Signitures of Natural Selection." *Genome Research*. https://doi.org/10.1101/gr.631202.

Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9):1655–64. https://doi.org/10.1101/gr.094052.109.

Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., … Kruglyak, L. (2012). Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature Genetics*, *44*(3), 285–290. https://doi.org/10.1038/ng.1050

Araújo, W. L., A. Nunes-Nesi, and A. R. Fernie. 2011. "Fumarate: Multiple Functions of a Simple Metabolite." *Phytochemistry* 72 (9):838–43. https://doi.org/10.1016/j.phytochem.2011.02.028.

Beadle, G. W. 1972. "Mystery of Maize." *Field Museum Natural History Bulletin*.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, and J. Ross-Ibarra. 2016. "Recent Demography Drives Changes in Linked Selection across the Maize Genome." *Nature Plants* 2 (7):16084. https://doi.org/10.1038/nplants.2016.84.

Bogdanović, J., M. Mojović, N. Milosavić, A. Mitrović, Ž Vučinić, and I. Spasojević. 2008. "Role of Fructose in the Adaptation of Plants to Cold-Induced Oxidative Stress." In *European Biophysics Journal*, 37:1241–46. https://doi.org/10.1007/s00249-008-0260-9.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics* 23 (19):2633–35. https://doi.org/10.1093/bioinformatics/btm308.

Browning, Sr, and Bl Browning. 2011. "Haplotype Phasing: Existing Methods and New

Developments." *Nature Reviews Genetics* 12 (10):703–14. https://doi.org/10.1038/nrg3054.Haplotype.

Bukowski, R., X. Guo, Y. Lu, C. Zou, B. He, Z. Rong, B. Wang, et al. 2015. "Construction of the Third Generation Zea Mays Haplotype Map." *bioRxiv*, 26963. https://doi.org/10.1101/026963.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; Iso-2; Iso-3." *Fly* 6 (2):80–92. https://doi.org/10.4161/fly.19695.

Cutter, A. D., and B. A. Payseur. 2013. "Genomic Signatures of Selection at Linked Sites: Unifying the Disparity among Species." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3425.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15):2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, et al. 2011. "A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data." *Nature Genetics* 43 (5):491–501. https://doi.org/10.1038/ng.806.

Doebley, J. 2004. "The Genetics of Maize Evolution." *Annu Rev Genet* 38:37–59. https://doi.org/10.1146/annurev.genet.38.072902.092425.

Eckert, A. J., A. D. Bower, J. L. Wegrzyn, B. Pande, K. D. Jermstad, K. V. Krutovsky, J. B. St. Clair, and D. B. Neale. 2009. "Association Genetics of Coastal Douglas Fir (Pseudotsuga Menziesii Var. Menziesii, Pinaceae). I. Cold-Hardiness Related Traits." *Genetics* 182 (4):1289–1302. https://doi.org/10.1534/genetics.109.102350.

Fang, Z., T. Pyhäjärvi, A. L. Weber, R. K. Dawe, J. C. Glaubitz, J. de Jesus Sánchez González, C. Ross-Ibarra, J. Doebley, P. L. Morrell, and J. Ross-Ibarra. 2012. "Megabase-Scale Inversion Polymorphism in the Wild Ancestor of Maize." *Genetics* 191 (3):883–94. https://doi.org/10.1534/genetics.112.138578.

Forde, B. G., and P. J. Lea. 2007. "Glutamate in Plants: Metabolism, Regulation, and Signalling." In *Journal of Experimental Botany*, 58:2339–58. https://doi.org/10.1093/jxb/erm121.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. 2011. "A Map of Local Adaptation in Arabidopsis Thaliana." *Science* 334 (6052):86–89. https://doi.org/10.1126/science.1209271.

Fu, Y. X. 1997. “Statistical Tests of Neutrality of Mutations against Population Growth, Hitchhiking and Background Selection.” *Genetics* 147 (2):915–25. https://doi.org/genetics.org//147/2/915.

Fukunaga, K., J. Hill, Y. Vigouroux, Y. Matsuoka, J. Sanchez, K. J. Liu, E. S. Buckler, and J. Doebley. 2005. “Genetic Diversity and Population Structure of Teosinte.” *Genetics* 169 (4):2241–54. https://doi.org/genetics.104.031393 [pii]\r10.1534/genetics.104.031393.

Fustier, M. A., J. T. Brandenburg, S. Boitard, J. Lapeyronnie, L. E. Eguiarte, Y. Vigouroux, D. Manicacci, and M. I. Tenaillon. 2017. “Signatures of Local Adaptation in Lowland and Highland Teosintes from Whole-Genome Sequencing of Pooled Samples.” *Molecular Ecology* 26 (10):2738–56. https://doi.org/10.1111/mec.14082.

Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov. 2015. “Recent Selective Sweeps in North American Drosophila Melanogaster Show Signatures of Soft Sweeps.” *PLoS Genetics* 11 (2):e1005004. https://doi.org/10.1371/journal.pgen.1005004.

He, S., S. Yan, P. Wang, W. Zhu, X. Wang, Y. Shen, K. Shao, H. Xin, S. Li, and J. Li. 2014. “Comparative Analysis of Genome-Wide Chromosomal Histone Modification Patterns in Maize Cultivars and Their Wild Relatives.” *PLoS ONE* 9 (5). https://doi.org/10.1371/journal.pone.0097364.

Hereford, J. 2009. “A Quantitative Survey of Local Adaptation and Fitness Trade-Offs.” *The American Naturalist* 173 (5):579–88. https://doi.org/10.1086/597611.

Hermisson, Joachim, and Pleuni S. Pennings. 2005. “Soft Sweeps: Molecular Population Genetics of Adaptation from Standing Genetic Variation.” *Genetics* 169(4): 2335–52.

Hermisson, J., and P. S. Pennings. 2017. “Soft Sweeps and beyond: Understanding the Patterns and Probabilities of Selection Footprints under Rapid Adaptation.” *Methods in Ecology and Evolution* 8 (6):700–716. https://doi.org/10.1111/2041-210X.12808.

Hudson, R. R. 2002. “Generating Samples under a Wright-Fisher Neutral Model of Genetic Variation.” *Bioinformatics* 18 (2):337–38. https://doi.org/10.1093/bioinformatics/18.2.337.

Hufford, M. B., P. Bilinski, T. Pyhäjärvi, and J. Ross-Ibarra. 2012. “Teosinte as a Model System for Population and Ecological Genomics.” *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.08.004.

Hufford, M. B., E. Martínez-Meyer, B.S. Gaut, L. E. Eguiarte, and M. I. Tenaillon. 2012. “Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight.” *PLoS ONE* 7 (11). https://doi.org/10.1371/journal.pone.0047659.

Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, et al. 2017.

"Improved Maize Reference Genome with Single-Molecule Technologies." *Nature* 546 (7659):524–27. https://doi.org/10.1038/nature22971.

Kawecki, T J, and D Ebert. 2004. "Conceptual Issues in Local Adaptation." *Ecology Letters* 7:1225–41. https://doi.org/10.1111/j.1461-0248.2004.00684.x.

Keightley, P. D., J. L. Campos, T. Booker, and B. Charlesworth. 2016. "Inferring the Frequency Spectrum of Derived Variants to Quantify Adaptive Molecular Evolution in Protein-Coding Genes of Drosophila Melanogaster." *Genetics* 203 (2):975–84. https://doi.org/10.1534/genetics.116.188102.

Leinonen, P. H., D. L. Remington, and O. Savolainen. 2011. "Local Adaptation, Phenotypic Differentiation, and Hybrid Fitness in Diverged Natural Populations of Arabidopsis Lyrata." *Evolution* 65 (1):90–107. https://doi.org/10.1111/j.1558-5646.2010.01119.x.

Li, H., and R. Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 26 (5):589–95. https://doi.org/10.1093/bioinformatics/btp698.

Messer, P. W., and D. A. Petrov. 2013. "Population Genomics of Rapid Adaptation by Soft Selective Sweeps." *Trends in Ecology and Evolution* 28 (11):659–69. https://doi.org/10.1016/j.tree.2013.08.003.

Moeller, D. A., and P. Tiffin. 2008. "Geographic Variation in Adaptation at the Molecular Level: A Case Study of Plant Immunity Genes." *Evolution* 62 (12):3069–81. https://doi.org/10.1111/j.1558-5646.2008.00511.x.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark, and C. Bustamante. 2005. "Genomic Scans for Selective Sweeps Using SNP Data." *Genome Research* 15 (11):1566–75. https://doi.org/10.1101/gr.4252305.

Pavlidis, P., D Živković, A. Stamatakis, and N. Alachiotis. 2013. "SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes." *Molecular Biology and Evolution* 30 (9):2224–34. https://doi.org/10.1093/molbev/mst112.

Payseur, B.A., A.D. Cutter, and M.W. Nachman. 2002. "Searching for Evidence of Positive Selection in the Human Genome Using Patterns of Microsatellite Variability." *Molecular Biology and Evolution* 19 (7):1143–53. https://doi.org/10.1093/oxfordjournals.molbev.a004172.

Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. "The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation." *Current Biology*. https://doi.org/10.1016/j.cub.2009.11.055.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A.R. Ferreira, D. Bender, J. Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage

Analyses." *The American Journal of Human Genetics* 81 (3):559–75. https://doi.org/10.1086/519795.

Pyhäjärvi, T., M. B Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biol Evol* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Ritland, K. 1996. "Estimators for Pairwise Relatedness and Individual Inbreeding Coefficients." *Genetical Research* 67 (2):175. https://doi.org/10.1017/S0016672300033620.

Ross-Ibarra, J., M. Tenaillon, and B. S. Gaut. 2009. "Historical Divergence and Gene Flow in the Genus Zea." *Genetics* 181 (4):1399–1413. https://doi.org/10.1534/genetics.108.097238.

Schrider, D. R., and A. D Kern. 2017. "Machine Learning for Population Genetics: A New Paradigm." *Doi.org*, 206482. https://doi.org/10.1101/206482.

Sheehan, S., and Y. S. Song. 2016. "Deep Learning for Population Genetic Inference." *PLoS Computational Biology* 12 (3). https://doi.org/10.1371/journal.pcbi.1004845.

Speed, D., G. Hemani, Michael R. Johnson, and David J. Balding. 2012. "Improved Heritability Estimation from Genome-Wide SNPs." *American Journal of Human Genetics* 91 (6):1011–21. https://doi.org/10.1016/j.ajhg.2012.10.010.

Swarts, K., H. Li, J. A. R. Navarro, D. An, M. C. Romay, S. Hearne, C. Acharya, et al. 2014. "Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants." *The Plant Genome* 7 (3):0. https://doi.org/10.3835/plantgenome2014.05.0023.

Terhorst, J., J. A. Kamm, and Y. S. Song. 2017. "Robust and Scalable Inference of Population History from Hundreds of Unphased Whole Genomes." *Nature Genetics* 49 (2):303–9. https://doi.org/10.1038/ng.3748.

Umeyama, S. 1991. "Least-Squares Estimation of Transformation Parameters Between Two Point Patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (4):376–80. https://doi.org/10.1109/34.88573.

Volis, S., S. Mendlinger, and D. Ward. 2002. "Differentiation in Populations of Hordeum Spontaneum along a Gradient of Environmental Productivity and Predictability: Life History and Local Adaptation." *Biological Journal of the Linnean Society* 77 (4):479–90. https://doi.org/10.1046/j.1095-8312.2002.00120.x.

Wallace, J. G., P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E.S. Buckler. 2014. "Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize." *PLoS Genetics* 10 (12). https://doi.org/10.1371/journal.pgen.1004845.

Wang, L., Beissinger, T. M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J., & Hufford, M. B.

(2017). The interplay of demography and selection during maize domestication and expansion. *Genome Biology*, *18*(1), 215. https://doi.org/10.1186/s13059-017-1346-4

Watterson, G. A. 1975. "On the Number of Segregating Sites in Genetical Models without Recombination." *Theoretical Population Biology* 7 (2):256–76. https://doi.org/10.1016/0040-5809(75)90020-9.

V.6 Supplementary information

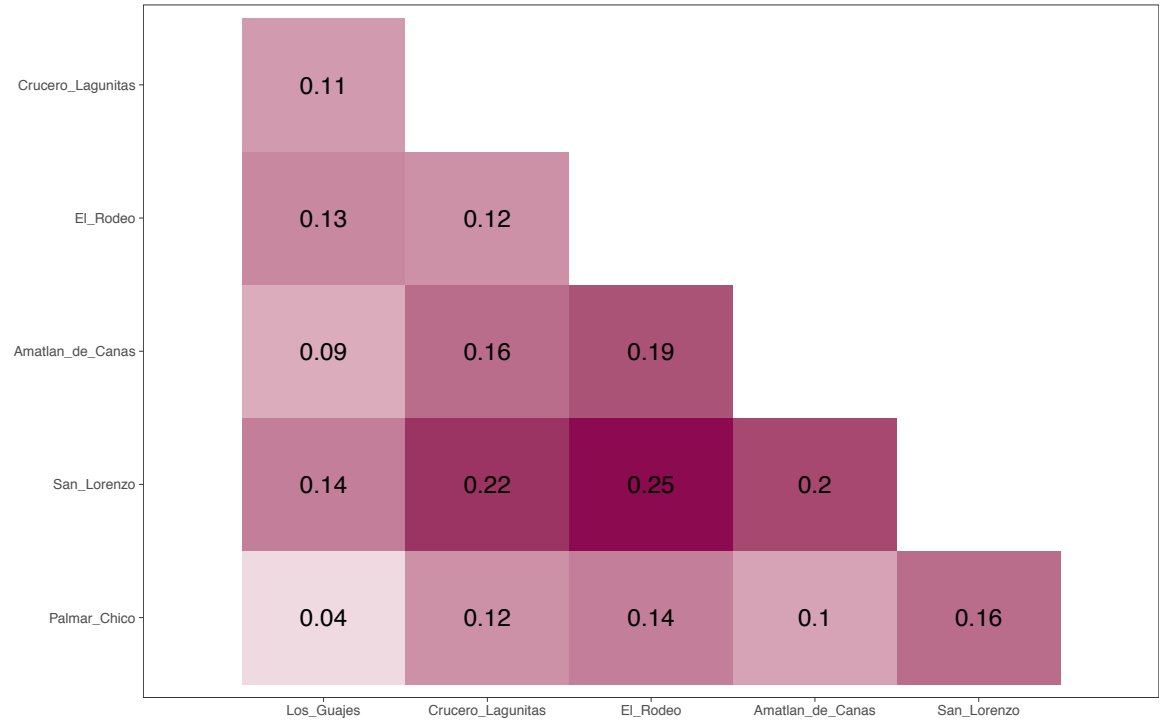


Figure S1: Median F_{ST} values of all pairs of populations across all samples.

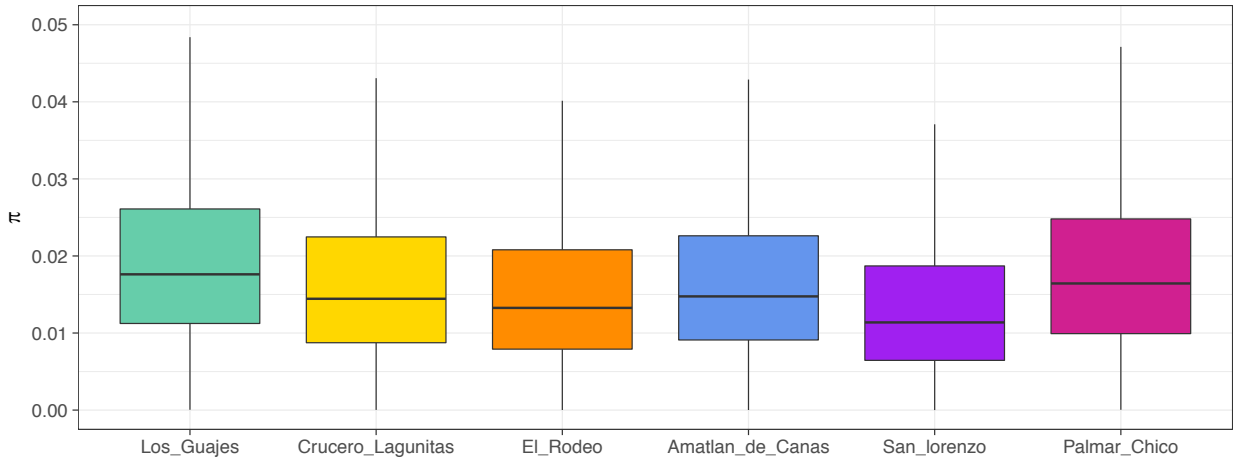


Figure S2: Median pi value across populations in each population.

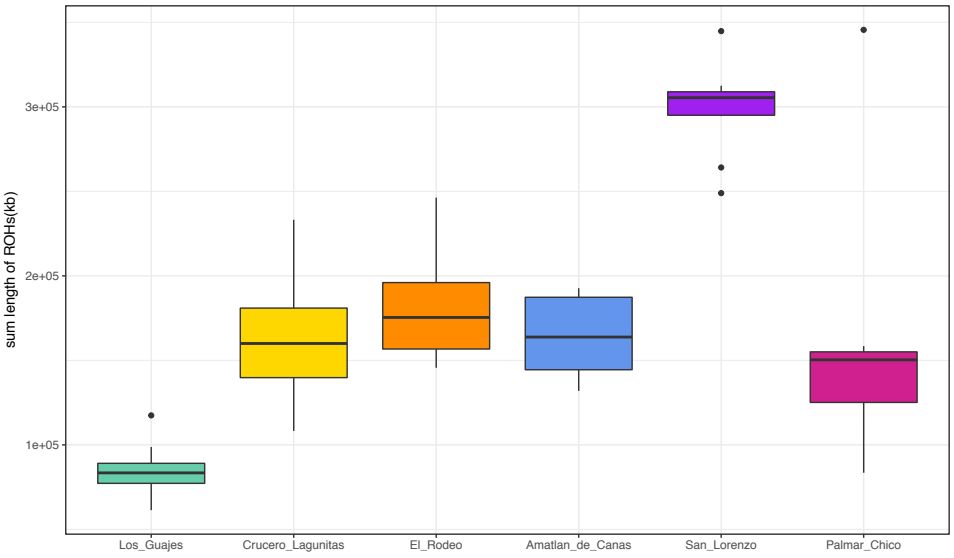


Figure S3: Boxplot of the cumulative length of runs of homozygosity (ROHs) in each population.

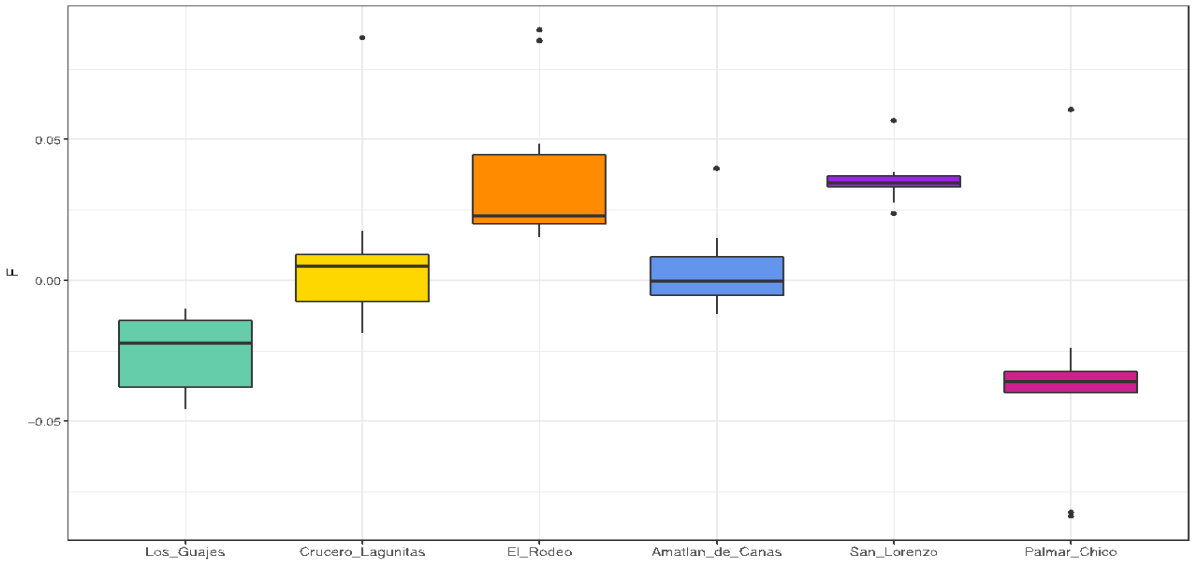


Figure S4: Boxplot of the individual inbreeding coefficients calculated with the method of moments estimators (MMEs).

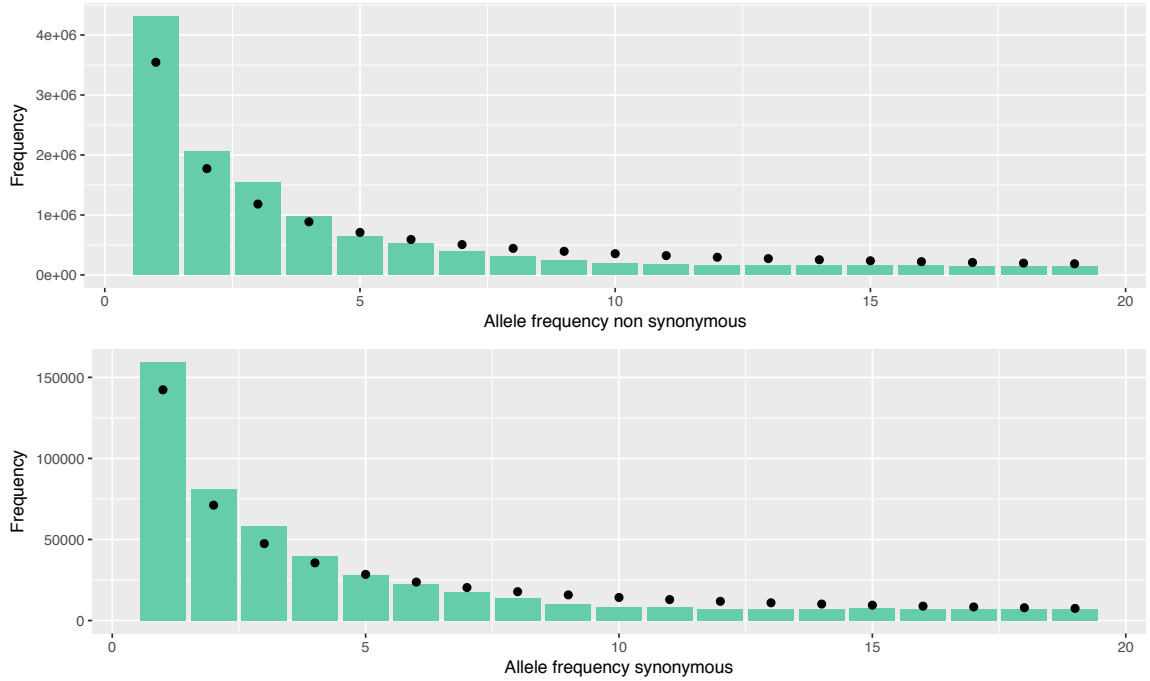


Figure S5: Folded site frequency spectrum at non-synonymous variants (top panel) and synonymous variants (bottom panel) for **Los Guajes**. Black dots represent expected SFS.

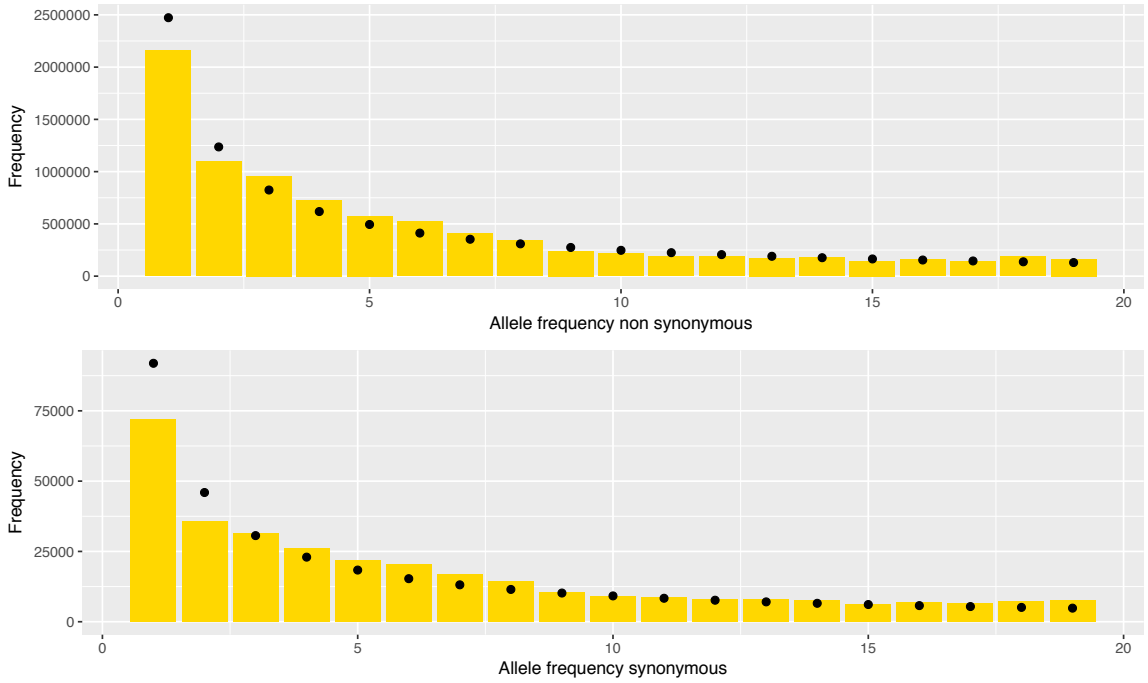


Figure S6: Folded site frequency spectrum at non-synonymous variants (top panel) and synonymous variants (bottom panel) for **Crucero Lagunitas**. Black dots represent expected SFS.

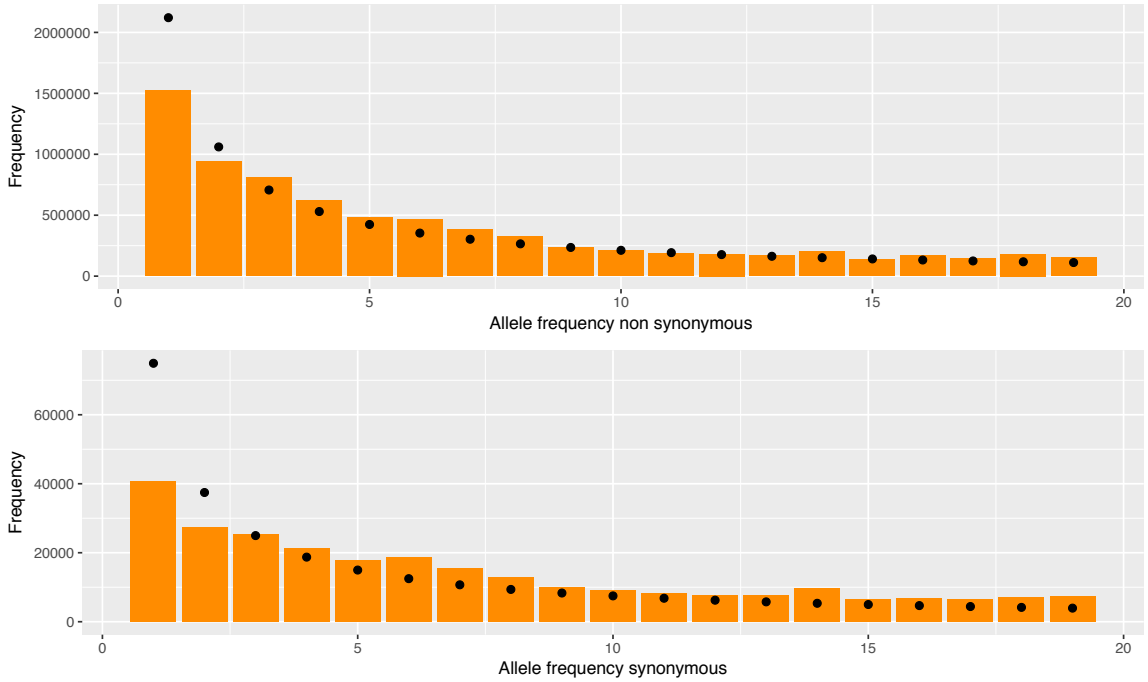


Figure S7: Folded site frequency spectrum at non-synonymous variants (top panel) and synonymous variants (bottom panel) for **El Rodeo**. Black dots represent expected SFS.

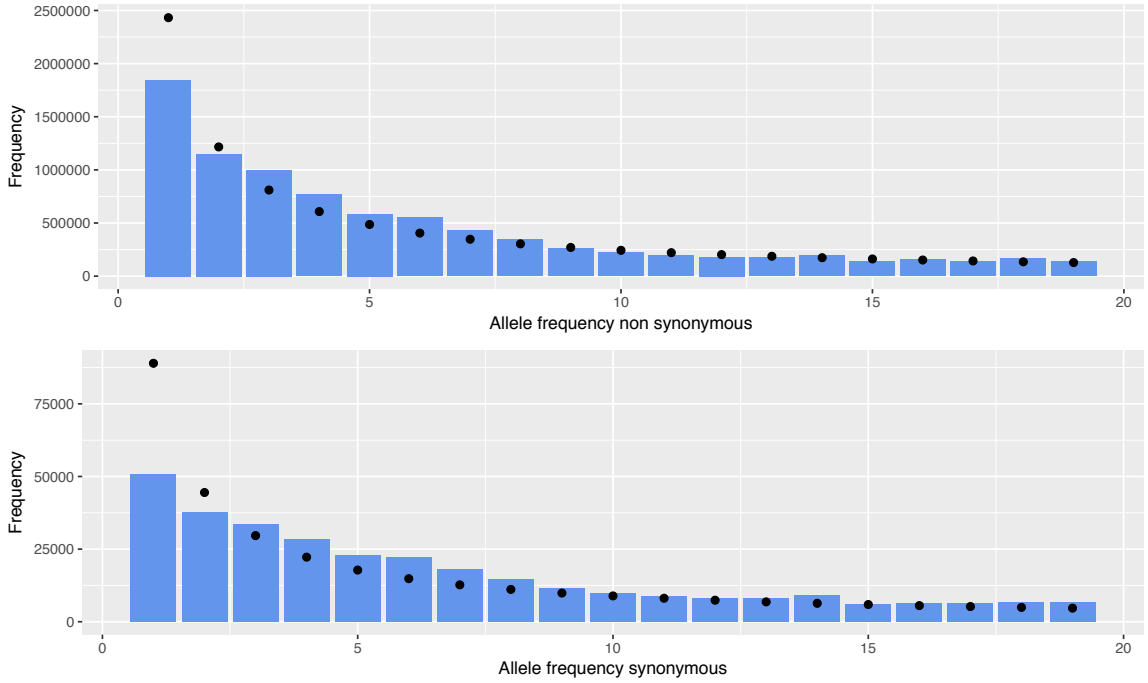


Figure S8: Folded site frequency spectrum at non-synonymous variants (top panel) and synonymous variants (bottom panel) for **Amatlan de Canas**. Black dots represent expected SFS.

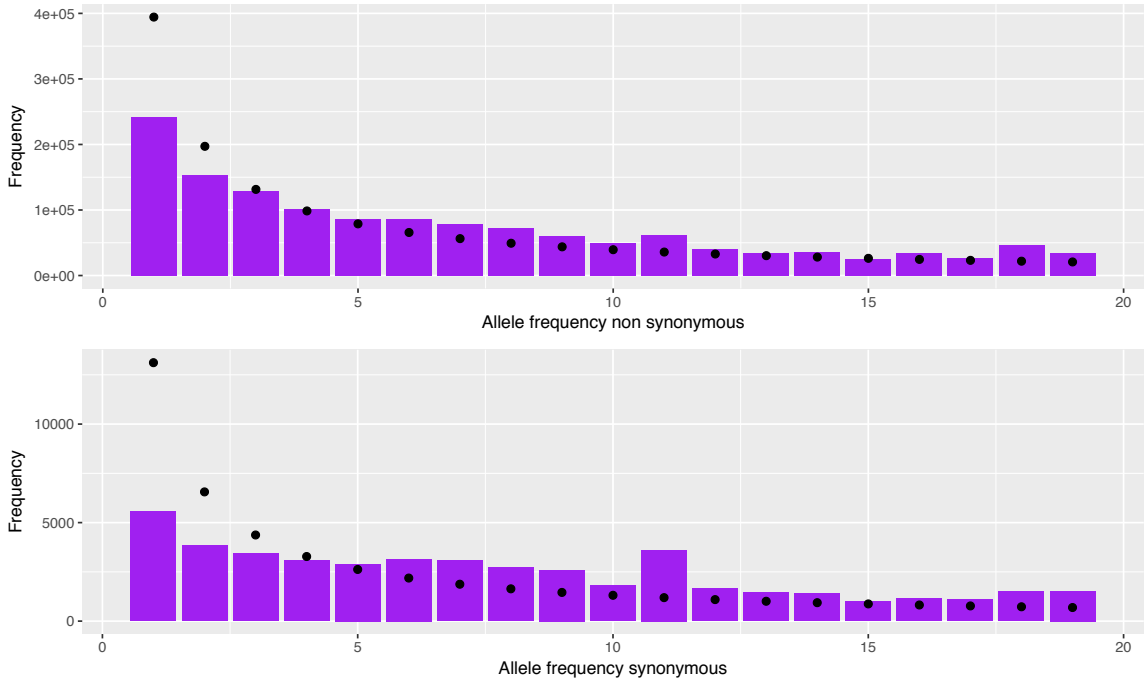


Figure S9: Folded site frequency spectrum at non-synonymous variants (top panel) and synonymous variants (bottom panel) for **San Lorenzo**. Black dots represent expected SFS.

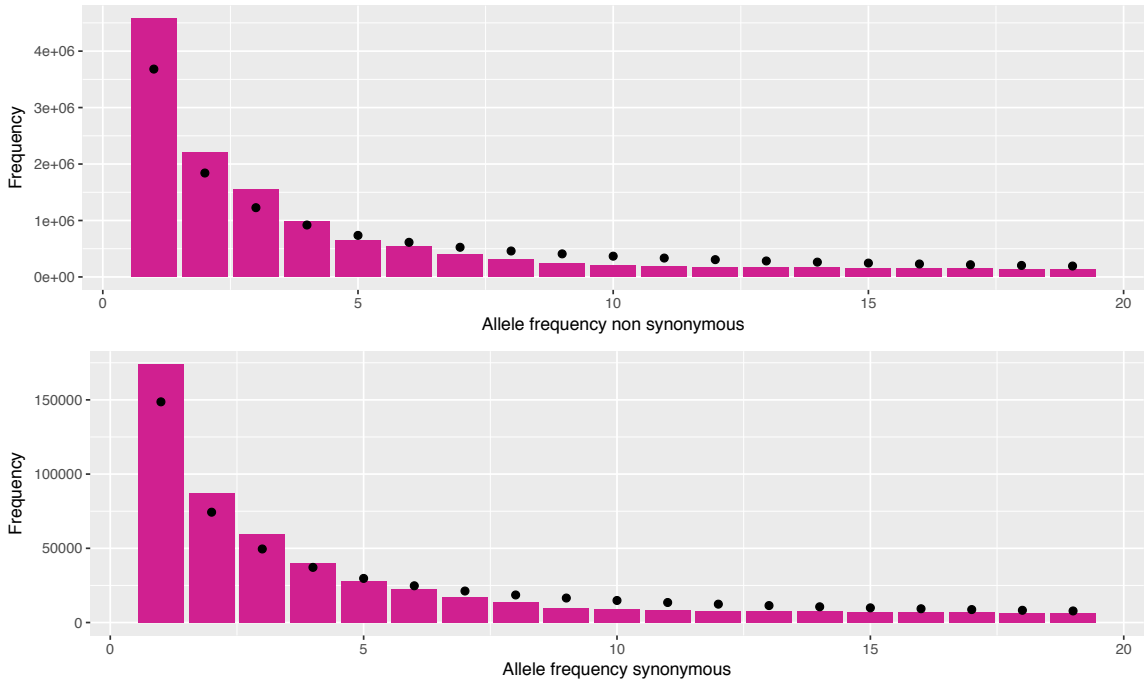


Figure S10: Folded site frequency spectrum of non-synonymous variants on top and synonymous variants on the bottom of **Palmar Chico** Black points represent expected SFS.

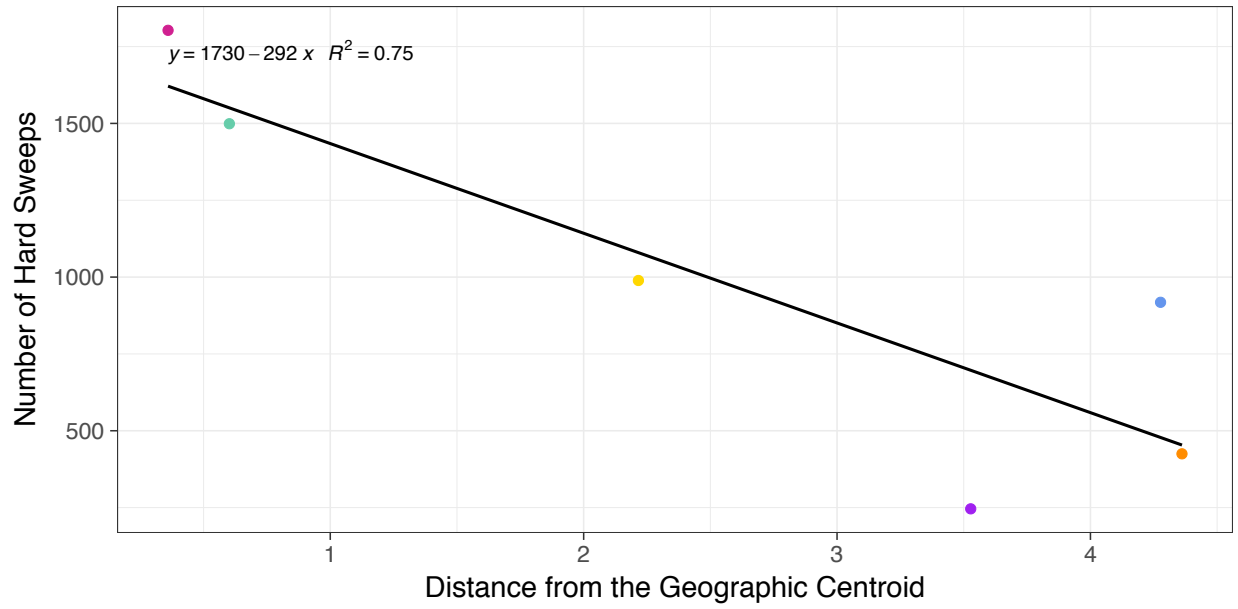


Figure S11: The number of hard sweeps versus geographic distance from the center of the niche. The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

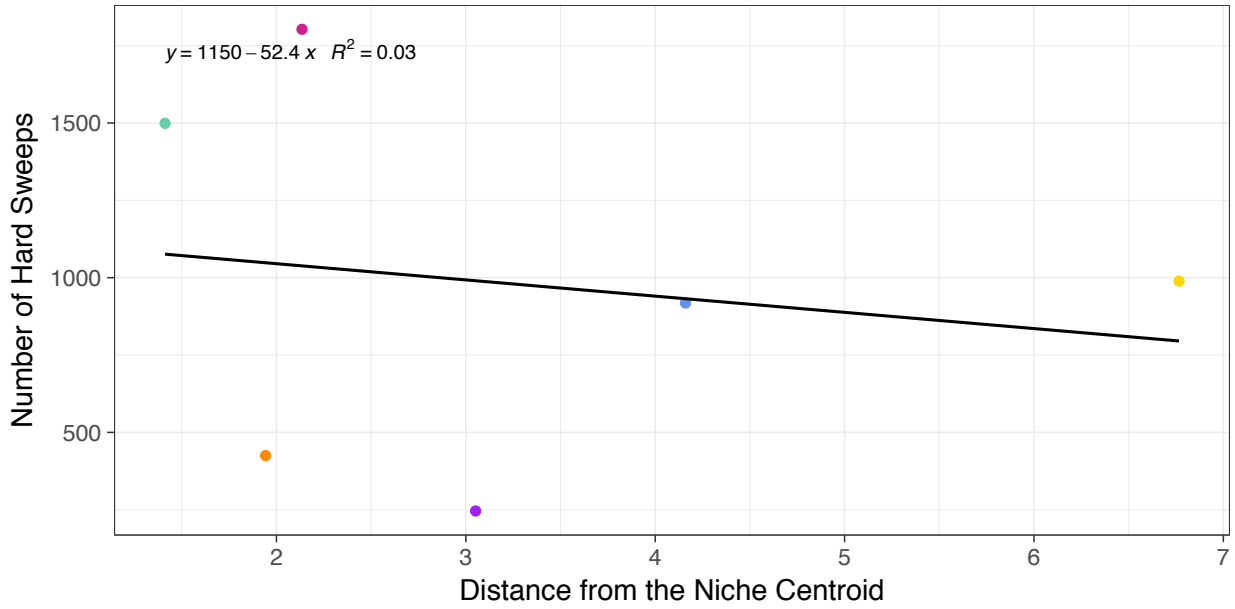


Figure S12: The number of hard sweeps versus distance from the niche centroid.

The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

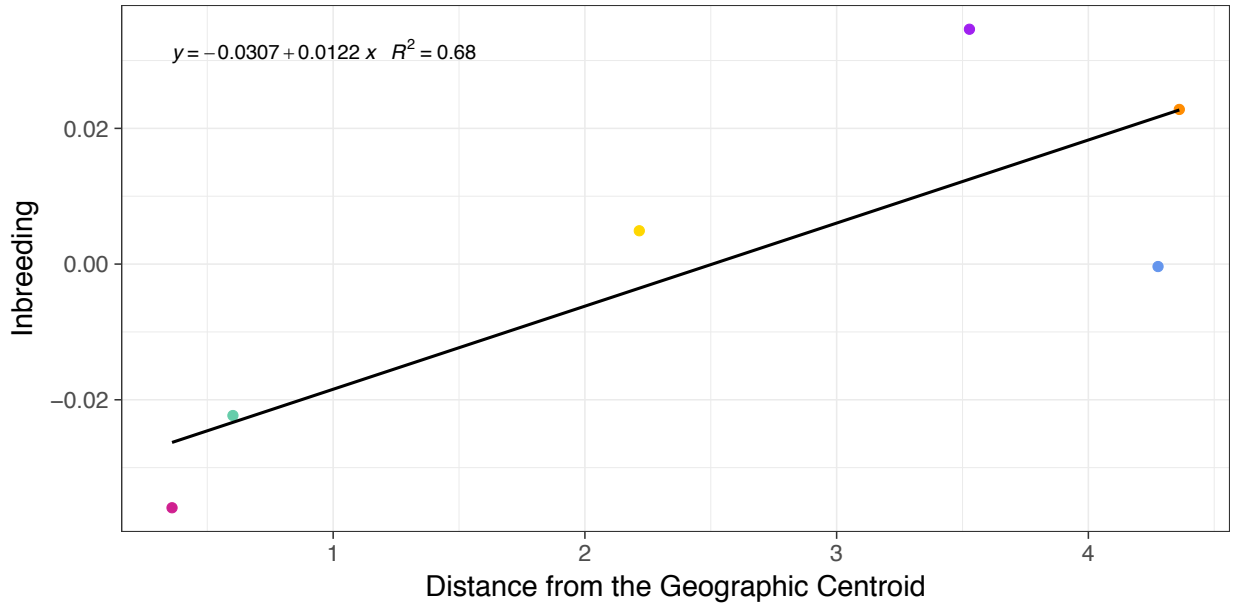


Figure S13: The median inbreeding values versus geographic distance from the center of the niche.

The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

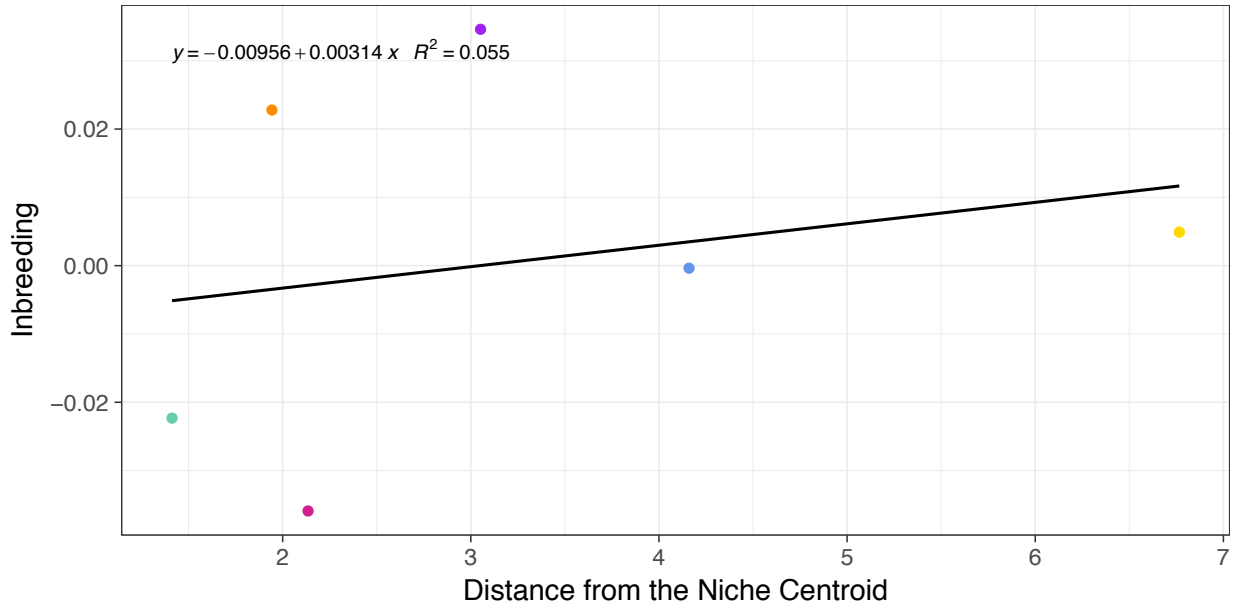


Figure S14: The median inbreeding values versus distance from the niche centroid.

The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

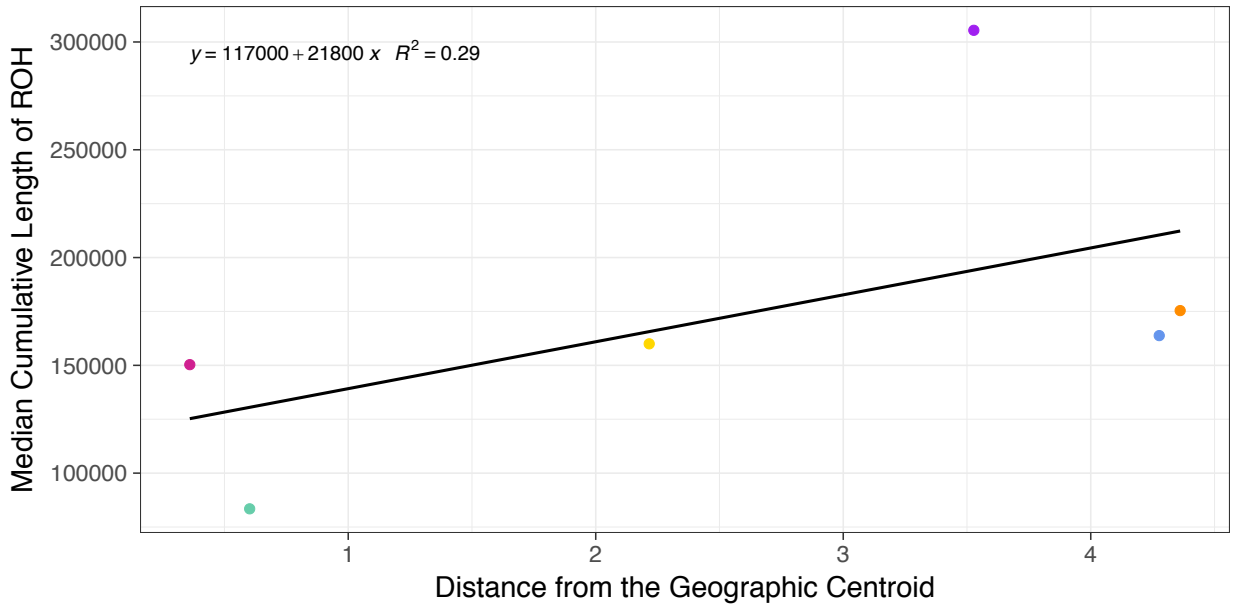


Figure S15: The median cumulative length of ROHs versus geographic distance from the center of the niche. The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

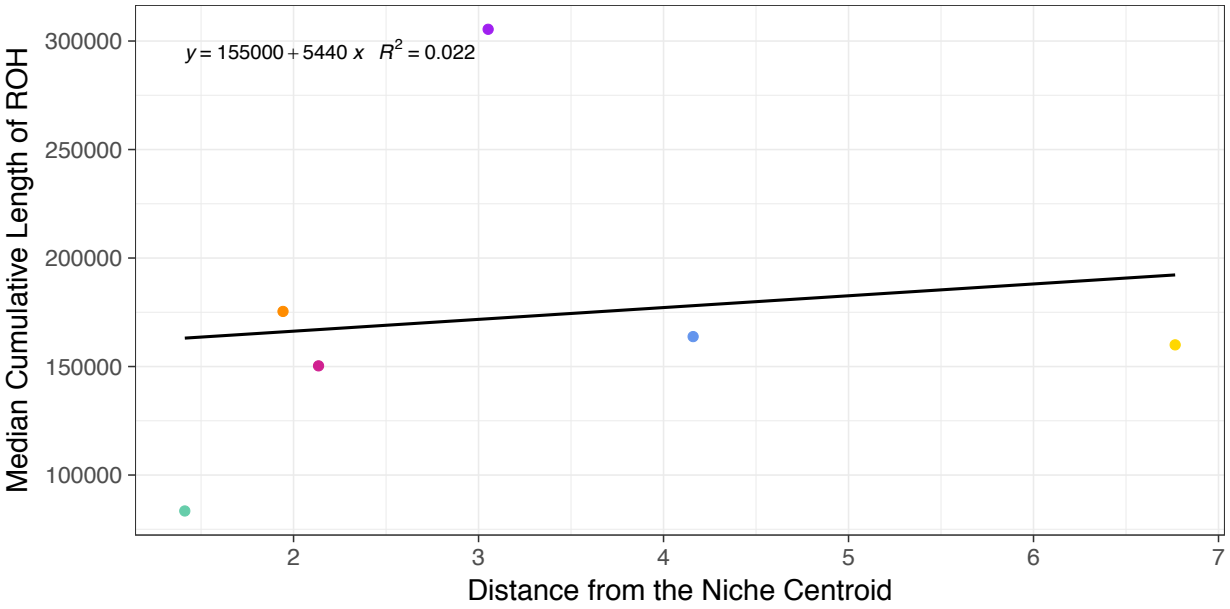


Figure S16: The median cumulative length of ROHs versus geographic distance from the niche centroid . The color of the point indicates the population associated with the value (Green for Los Guajes, yellow for Crucero Lagunitas, orange for El Rodeo, blue for Amatlan de Canas, purple for San Lorenzo and pink for Palmar Chico).

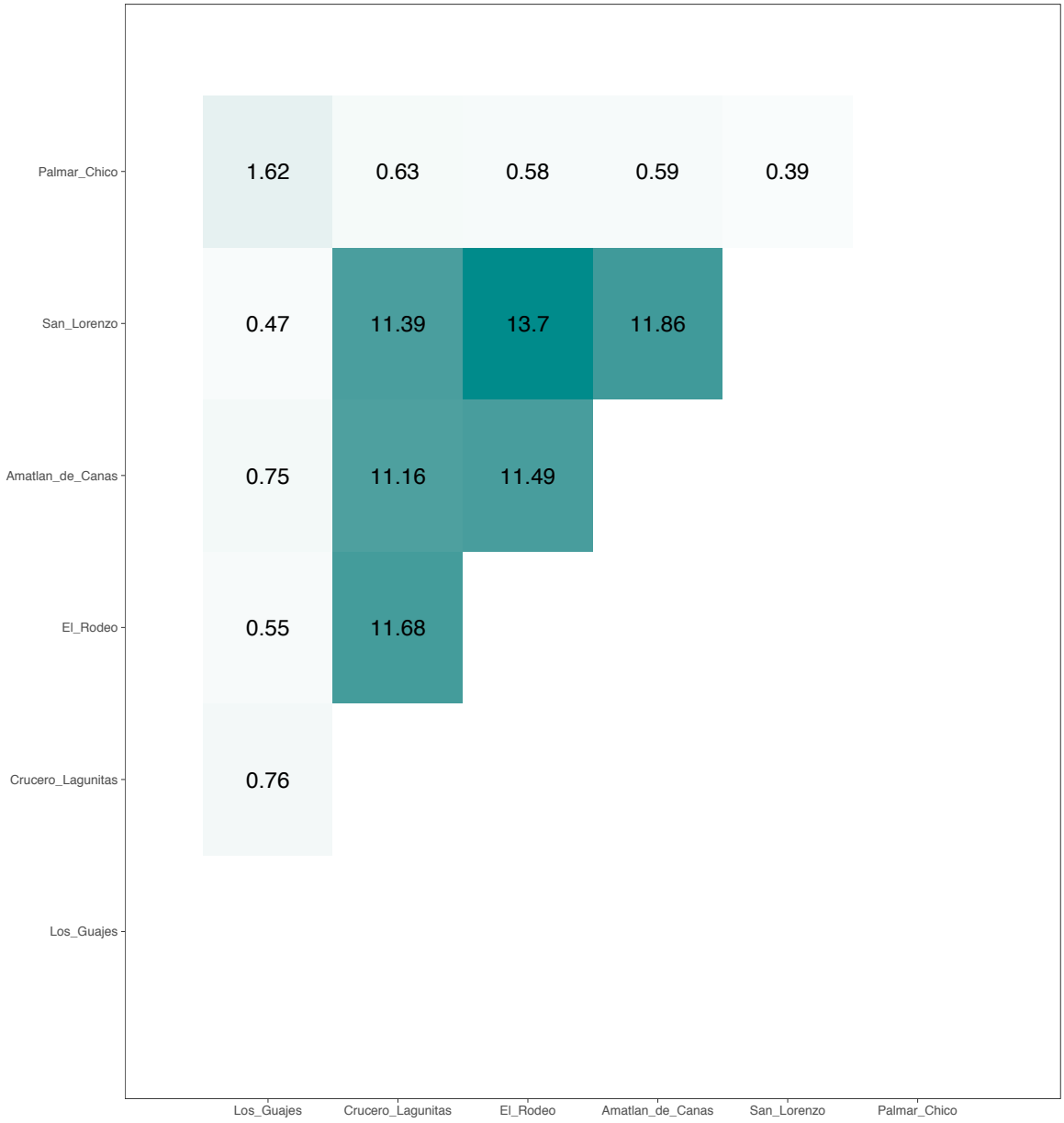


Figure S17: Soft sweeps overlap in percent between all pairs of populations.

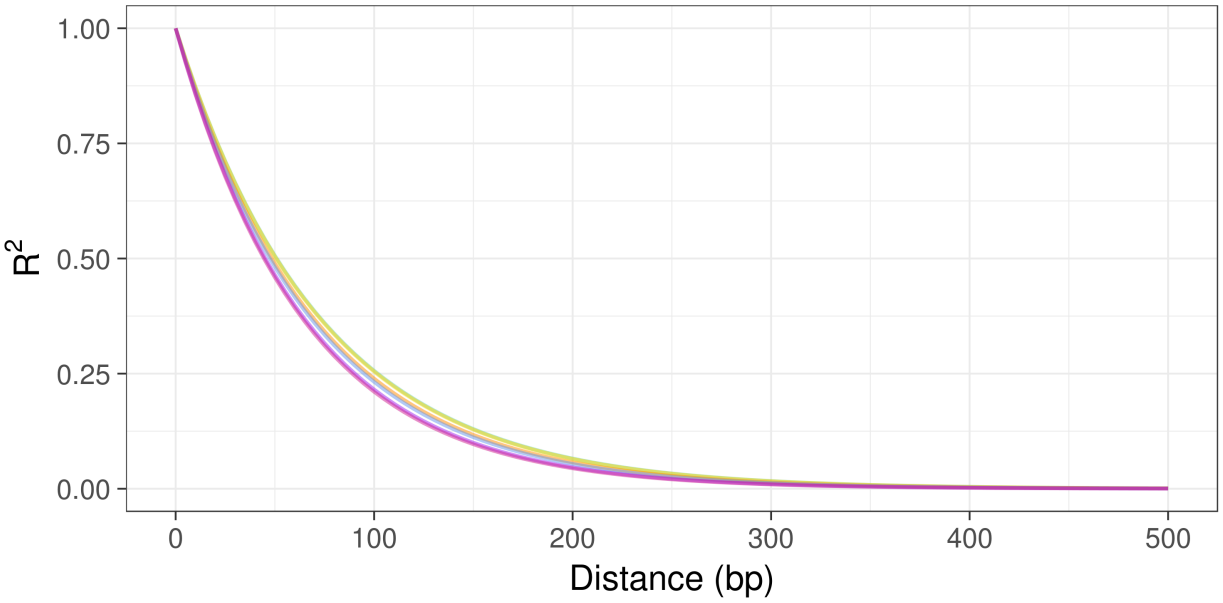


Figure S18: Linkage disequilibrium (r^2) decay along physical distance (bp) for the 6 populations.

VI. DISCUSSION AND PERSPECTIVES

VI.1 Main results

The *Zea* taxa encompass wild and cultivated species, providing a unique opportunity to dissect the determinants of adaptation at both short- (since domestication) and long- time scales (natural selection in wild populations). Moreover, the process of selection during maize domestication has been extensively studied. Maize was domesticated from *Zea mays* subsp. *parviglumis* during a single event, around 9 000 years ago, in the Balsas river Valley of Mexico (Piperno and Flannery 2001; Matsuoka et al. 2002; van Heerwaarden et al. 2011). The domestication of maize resulted in an initial bottleneck (i.e. a reduction of population size) due to human-driven artificial selection, followed by a recovery (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005; Beissinger et al. 2016). These events mildly decreased the genetic diversity of maize compared to *parviglumis* (~20%) (Matsuoka et al. 2002; Hufford, Xu, et al. 2012). However, its demographic history is likely more complex than a single bottleneck followed by an expansion. Indeed, maize rapidly expanded from the center of domestication throughout the Americas, but population growth slowed down as it encountered, and adapted to higher latitudes and altitudes (Da Fonseca et al. 2015; Vigouroux et al. 2008; Swarts et al. 2017). Moreover, gene flow from *Zea mays* subsp. *mexicana* into highland maize, which contributed to its highland adaptation, further complicates these demographic patterns (van Heerwaarden et al. 2011; Hufford et al. 2013).

In the first chapter, we shed light on the effect of post-domestication demography and introgression from teosintes in a sample of maize landraces spanning much of its pre-Columbian

distribution. We found that historical changes in population sizes associated with domestication, as well as founder events during the expansion of maize in the Americas, increased the number of deleterious alleles (i.e. its genetic load). During these bottlenecks, maize experienced inbreeding depression, due to a significant reduction in *Ne*. Such reduction is known to reduce the efficacy of selection at a genome wide scale, therefore allowing slightly deleterious variants to increase in frequency. This genetic load was particularly pronounced in the Andean maize, which underwent a stronger founder effect. Interestingly, gene flow from *mexicana* in highland maizes contributed significantly to restore genetic diversity. Notably, this introgression was not possible in Andean maize, because the range of *mexicana* is limited to Mexico. Together, these observations reinforce the importance of gene flow and demography as key players during the post-domestication adaptation of maize.

Characterizing deleterious alleles is important as those are putative targets for improving maize's breeding germplasms. Indeed, although, highly deleterious variants are quickly removed by purifying selection, weakly deleterious variants can persist in the population. Moreover, recessive deleterious variants can escape purifying selection in the heterozygous stage and reach considerable allele frequencies. Because selection acts on phenotypes which are for the most part quantitative, the cumulative contribution of deleterious allele can reduce fitness. As an example, deleterious alleles have been shown to influence traits tightly related to fitness, such as grain yield in the modern hybrids (Yang et al. 2017). Regions of low recombination rate are prone to a higher load and are associated with an excess of heterozygosity in maize inbreeds. This suggests that these regions are enriched in deleterious alleles which prevents homozygosity (Rodgers-Melnick et al. 2015). The purging process during inbreeding lead to fewer segregating deleterious variants in modern inbred lines. However, these lines also contain a higher number of fixed deleterious

variants due genetic drift (van Heerwaarden, Hufford, and Ross-Ibarra 2012; Yang et al. 2017). The modern cultivation of maize consists exclusively of hybrids, which are the results of a cross between two inbred lines. These hybrids show better performance when compared to their parents, an effect know as hybrid vigor or heterosis. The genetics underlying heterosis is for the most part unknown, but one of the proposed process is the dominance hypothesis. It attributes the superiority of hybrids to the complementation of the deleterious recessive alleles from one parent by the dominant alleles from the second. While hybrids allow for higher yield and crop uniformity, they also maintain deleterious variants in the alleles pool. Therefore, looking at genetic load in landraces and teosintes can reveal target alleles for the improvement of inbred lines.

In the second chapter, we investigated an underappreciated aspect of the evolution of novel traits during domestication, which is plasticity. Plasticity is defined as the capacity of a genotype to express different phenotypes in response to varied environmental conditions. Populations with plastic genotypes generating flexible phenotypes are predicted to better cope with environmental changes, colonize broader niches, and display a greater potential for expansion (Wennersten and Forsman 2012). The phenotypic flexibility allowed by plasticity is a particularly important process in plants, as they are fixed in a specific location and have limited abilities to protect themselves from their environment (Des Marais, Hernandez, and Juenger 2013). In this context, we investigated the plasticity of gene expression of maize and *parviglumis* between Early-Holocene and present conditions. As compared to contemporary levels, low CO_2 and temperature induced a plastic response in multiple *parviglumis* phenotypes, some typical of maize (e.g. changes in inflorescence sexuality and vegetative architecture). Interestingly, these traits were canalized in maize, as no variation in branching or flower sexuality was observed under the same conditions. On top of these changes in morphology, we observed substantial changes in co-expression

networks and over 2,000 genes that exhibited differential expression solely in *parviglumis*. These results suggest that during maize domestication, the crop has lost substantial plasticity.

Moreover, environmentally induced sex determination and reduced branching in *parviglumis* has been observed during other stress conditions, such as drought and nutrient limitation (West-Eberhard 2003). Although the environment could have originally induced these traits, phenotypes were most likely selected and fixed by genetic assimilation during domestication and have become main differentiating characters between maize and teosintes. This was first hypothesized by John Doebley in 1995 based on his observations of plants in their natural environment (reviewed in West-Eberhard 2003). Variation in plastic responses can be described as variation in genotype-by-environment interactions (GxE), an important factor in plant breeding. In a recent study, authors observed a lack of plastic response (GxE) in the US temperate maize germplasm, likely resulting from selection for stable crop performance across varying environments (Gage et al. 2017). Reduced plasticity can impact the potential for future maize adaptation, especially in the context of climate change. Interestingly, another study showed that plasticity in maize is controlled by other loci than those controlling mean trait values (Kusmec et al. 2017). These independent mechanisms suggest that breeders could, at least to some degree, select for stable performances, while simultaneously selecting phenotypic plasticity to increase performances to changing environments.

In the last chapter, we identified genomic signatures of local adaptation in six natural populations of the subspecies *parviglumis*. *Parviglumis* is an annual wild grass native to Mexico that underwent extensive local adaptation (Aguirre-Liguori et al. 2017; Fustier et al. 2017; Pyhäjärvi et al. 2013). We specifically searched for traces of recent directional selection, such as hard and soft sweeps, and inferred demographic histories for these teosinte populations. We

observed that the post-glacial expansion of teosintes engendered a differential reduction in *Ne* among populations, although all expanded their ranges. The fact that populations shared few sweeps confirms strong local adaptation in *parviglumis*. Our observation of population expansion is in agreement with analyses of pollen content in sediment cores from Mexico which suggested that grasses have expanded over the last 10,000 years due to changing environmental conditions during the Holocene and human management of vegetation with fire (Piperno et al. 2007; Correa-Metrio et al. 2012). Moreover, both absolute and relative variation in numbers of hard and soft selective sweeps within these populations indicates a tight interplay between demography and selection, which bears importantly on the adaptive potential of both individual populations and species. Although we looked specifically at directional selection, it is probable that other types of selection are acting on these populations, such as balancing selection. It is worth noting that differential selection can create signatures of balancing selection at the metapopulation level, but balancing selection is most likely also acting at the population level. Selection for pathogen resistance associated with their temporal fluctuations for instance generates patterns of balancing selection (Stahl et al. 1999; Tiffin and Gaut 2001).

In summary, 1) genetic adaptation to a small geographical scale; 2) repeated introgressions between wild and domestic forms; and 3) genetic assimilation that "cements" domesticated phenotypes initially induced by plastic responses to the environment, are all mechanisms involved in *Zea mays* adaptation. Fully characterizing the relationships between demography, gene flow and selection will provide new avenues to understand how history influenced the current evolutionary trajectory of a species. In addition, because of its close relatedness to maize, knowledge gathered from teosinte can be translated to maize to improve its breeding. Introgression from teosintes into maize is not uncommon and helped maize adapt to high altitudes, were the

environment is drier and colder compare to the center of domestication (Takuno et al. 2015). In other words, alleles from teosinte could be used to improved maize germplasm for important processes that impact maize breeding, such as drought or cold tolerance. Introgression of wild alleles has been utilized in other breeding programs, for example to improve yield in tomatoes (Tanksley and McCouch 1997).

VI.2 Methods

We observed some discordance between the demographic histories calculated using two different methods, namely, MSMC and SMCPP. What motivated the choice of SMCPP in the third chapter was its ability to use unphased data for inferences, which was important for us as phasing teosinte is difficult and is known to be inaccurate (personal communication with the Buckler lab, Cornell University, which has been attempting to improve the accuracy of the phasing in teosinte samples). Moreover, it has been shown that even small amounts of phasing errors can lead to incorrect estimates of demographic histories with programs such as MSMC (Terhorst, Kamm, and Song 2017). The discordance between results is not unique to our dataset as, for example, a recent paper on grape reported that MSMC found a bottleneck associated with grape domestication, while SMCPP did not (Zhou et al. 2017). With our dataset, these differences can be striking such as, for example, an opposing demographic history found for Palmar Chico between MSMC and SMCPP. While both programs infer a recent expansion in the population, in the first chapter, the demographic inference of MSMC showed a continuous increase in *Ne* until the recent past (~ 1200−1800 years BP) in Palmar Chico, while in the third chapter, using SMCPP, we found a continuous decrease in *Ne*. We performed simulations with demographic histories obtained from

both chapters for Palmar Chico. Both frameworks resulted in simulations inferring some population genetics parameters correctly, while other inaccurately. For example, we could get the right value of Tajima's D but not π with the simulations utilizing the demographic history from MSMC. As for SMCPP we could get π right but not Tajimas' D. This is likely the result of differences in the underlying models. Indeed, MSMC utilize recombination and LD while SMCPP, SFS and LD. Furthermore, these programs were developed and optimized with human genetic data, but not with plant data. Which can explain why they are both correctly interfering demographic histories in humans but not in teosintes. Indeed, there are differences between humans and plants, such as their shorter haplotypes and their higher heterozygosity. Instead of considering our results within the context of two different demographic histories, we will investigate the proper approach to harmonize these inferences, or at least, select the most likely history given our current knowledge of teosinte populations.

In the second chapter, we decided to study plasticity using transcriptomic data generated from leaf tissue. Although it provided some insight about plasticity in teosinte and maize, this study also suffers from a number of limitations. Indeed, we choose to study differences in plasticity between maize and teosinte in response to low temperature and low CO_2 conditions which unlikely represent all the climatic conditions encounter during the Early-Holocene period. This suggests that the choice of only two conditions will result in an incomplete representation of the phenomenon. Moreover, selection during domestication has most likely not selected uniquely on genes responding to temperature or CO_2, which probably explains the low overlap with domestication candidates. Another issue is that leaf tissue is not the best choice to study changes in inflorescence sexuality and branching, the traits that have been affected the most by domestication. The choice for leaf tissues was motivated by the need for keeping inflorescence

and meristems for phenotyping and saving seeds for further experiments. Another issue of this study is the difficulty to harvest maize and teosintes at equivalent stages. The space in the greenhouses being limited, we chose to grow teosinte and maize during separated growing seasons and analyze both data independently. For *parviglumis*, we collected leaf tissue 60 days after germination. For maize, plants started to flower after 50 days only. This difference in developmental stages was partially solved by removing 15 genes known to be differentially expressed before and after flowering (Alter et al. 2016). Finally, we chose to study a set of genes that are plastic in teosintes but not in maize, although a set of genes equivalent in size exhibited changes in maize but not in teosintes. This was motivated by the supposedly greater plastic responses in teosintes (Piperno et al. 2015) and the selection from teosinte to maize during domestication. The set of genes showing plastic responses in maize are enriched in pathways linked to photosynthesis that were downregulated for the Early Holocene conditions. This suggests that stressful conditions mostly impacted photosynthesis in maize and is most likely the result of human selection on some traits during modern breeding.

While most studies looking at the genetic adaptation of teosintes have used data gathered from SNP arrays, we instead chose to sequence individual genotypes at high depth. This key difference allows us to study adaptation across entire genomic landscapes and to utilize haplotypes. Indeed, SNP arrays only provide information for a subset of the genome enriched in genic regions, while loci involved in local adaptation in teosintes are instead enriched in intergenic regions. Moreover, we observed, using an analysis of LD decay, that complete decay occurs within 300bp in our populations. This demonstrates that haplotypes are short in *parviglumis* and substantiate our claim that most variants in the genome will not be tagged by a variant on a SNP array. Furthermore, the SNP arrays that are used for teosintes are biased as they were originally designed using a maize

reference panel, resulting in variants specific to teosintes being missed. This ascertainment bias particularly impacts methods that utilize outliers for inference, such as F_{ST} methods, which are based on extreme allele frequency differentiation between populations. Indeed, these methods will misclassify common variants as outliers and target of selection, not because of true biological reasons but rather because of the absence of true outliers. More generally, because SNP arrays represent such a small portion of total genomic variants, they are not equipped to identify most loci involved in local adaptation. In addition, as most traits driving local adaptation are polygenic, this will result in an incomplete picture of the complex genetic mechanisms underlying these traits. As an illustration, the data from a maize Illumina 55K arrays, leading to 37K quality SNPs (Pyhäjärvi et al. 2013), will result in ~300 to 400 outliers (at a one 1% tail cutoff) which might not represent a complete picture of adaptive variants.

An alternative to SNP arrays is Pool-sequencing which remains substantially cheaper compared to full sequencing of individual genomes. Although, pooled sequencing can suffer from severely biased allele frequency estimations due to uneven pooling, it has been successfully used in studying population genetics in many organisms (Christian Schlötterer et al. 2014) and, when the pool is balanced gives access to highly accurate allele frequencies. Although, pool-sequencing suffers from the lack of haplotype information at the level of an individual, limiting the study of the mechanisms that led to differentiation in allele frequencies.

On the other hand, whole genome sequencing of numerous individuals is still expensive for studying large genomes, such as maize, of numerous individuals. To reduce cost, we were therefore limited to the analysis of only 6 populations of *parviglumis*, compared for example to a previous study that used 49 populations genotyped using a SNP array (Aguirre-Liguori et al. 2017). In addition to the low number of populations, errors in short read mapping and SNP calling are

frequent. This is due to the reference genome originating from a single modern improved inbred maize line and therefore a reference bias exists with respect to teosinte. Because of this bias, fewer reads will map as the distance from the sample to the reference increases. As a result, misalignment of short reads to the reference occurs. Furthermore, the high diversity found in maize and teosintes, combined with very frequent CNVs at the level of single populations (Ross-Ibarra Lab, unpublished), together suggest that we are missing many uncommon variants due to SNP calling artifacts. A solution to this problem would be to use a reference genome specific to *parviglumis* or a set of reference genomes, known as a pan-genome. This would permit to more accurately assess the diversity in the samples and reduce reference biases. Although the panzea group (www.panzea.org) is currently working on both solutions, this technology is not yet available for *parviglumis*.

Using whole genome sequencing of individuals, one can detect traces left from recent directional selection in the genome and their underlying generative processes. We were able to detect soft and hard sweeps because of our high depth individual sequencing and look at their relative contribution during local adaptation. In this study, we do not require any priors about the number of sites involved in the process, but are encouraged by the ability to further associate these selected variants with the causal agents of selection.

VI.3 Future directions

We are currently investigating the plasticity of maize and teosinte under future projected climate conditions (i.e. higher temperature and higher CO_2), or in other words, the opposite conditions compared to our study reported in the second chapter. We planned on collecting

individual genotypes and phenotypes to measure the extent of this response, as well as identifying responsible genes.

In parallel, we are duplicating the experiment of chapter two with landraces instead of a modern inbred line to better represent the plastic diversity found in maize and identify loci that were fixed during domestication and modern breeding. This has the potential to identify candidate loci to improve modern germplasm in face of the global warming associated with the increase of atmospheric CO_2. Finally, we proposed to grow *parviglumis* in Holocene climate conditions, over many generations, while selecting individuals for plastic traits identified by our previous study and that were likely fixed by human selection during maize domestication.

Regarding the third chapter, we are using a new unpublished version of the software S/HIC (Schrider and Kern 2016), that utilizes supervised Neural Network to detect hard and soft sweeps, in collaboration with the Kern Lab. We used it with demography histories generated using SMCPP and determined that S/HIC detect hard and soft sweeps with a power of ~80% in our samples (private communication with the Kern Lab). However, this machine learning approach requires to train the network on simulations, and the method that provides us with the most likely demographic history needs to be determined first.

Next, we will infer the model of convergence between the different selective sweeps in our populations of *parviglumis* using a recently released method (Lee and Coop 2017). This method utilizes the covariance of allele frequencies in windows around a selected site. This will enable us to compare models in which the convergent allele sweeping in the populations are the result of the selection on standing genetic variation, arise from multiple mutations events in the populations or the same mutation event.

Last, the progeny of our teosinte populations has been phenotyped by the Doebley Lab for 16 traits including height, tillering, seed production, and ear morphology. Progenies were genotyped using Genotyping-By-Sequencing (GBS). This GBS data are used for ongoing genome-wide association analyses that will identify selected phenotypes and provide the basis to apply F_{ST}-Q_{ST} analyses (Leinonen et al. 2013). This analysis allows to identify phenotypes with stronger differentiation among populations compared to neutral genome-wide markers.

VI.4 References

Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. "Connecting Genomic Patterns of Local Adaptation and Niche Suitability in Teosintes." *Molecular Ecology* 26 (16):4226–40. https://doi.org/10.1111/mec.14203.

Alter, P., S. Bircheneder, L-Z. Zhou, U. Schlüter, M. Gahrtz, U. Sonnewald, and T. Dresselhaus. 2016. "Flowering Time-Regulated Genes in Maize Include the Transcription Factor ZmMADS1." *Plant Physiology* 172 (1). Am Soc Plant Biol:389–404.

Beissinger, T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, and J. Ross-Ibarra. 2016. "Recent Demography Drives Changes in Linked Selection across the Maize Genome." *Nature Plants* 2 (7):16084. https://doi.org/10.1038/nplants.2016.84.

Bierne, N., D. Roze, and J. J. Welch. 2013. "Pervasive Selection or Is It.? Why Are FST Outliers Sometimes so Frequent?" *Molecular Ecology* 22 (8):2061–64. https://doi.org/10.1111/mec.12241.

Charlesworth, D.. 2006. "Balancing Selection and Its Effects on Sequences in Nearby Genome Regions." *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.0020064.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. "Using Environmental Correlations to Identify Loci Underlying Local Adaptation." *Genetics* 185 (4):1411–23. https://doi.org/10.1534/genetics.110.114819.

Correa-Metrio, A., S. Lozano-García, S. Xelhuantzi-López, S. Sosa-Nájera, and S. E. Metcalfe. 2012. "Vegetation in Western Central Mexico during the Last 50 000 Years: Modern

Analogs and Climate in the Zacapu Basin." *Journal of Quaternary Science* 27 (5):509–18. https://doi.org/10.1002/jqs.2540.

Eyre-Walker, a, R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. "Investigation of the Bottleneck Leading to the Domestication of Maize." *Proceedings of the National Academy of Sciences of the United States of America* 95 (April):4441–46. https://doi.org/10.1073/pnas.95.8.4441.

Fonseca, R. R. Da, B. D. Smith, N. Wales, E. Cappellini, P. Skoglund, M. Fumagalli, J. Alfredo Samaniego, et al. 2015. "The Origin and Evolution of Maize in the Southwestern United States." *Nature Plants* 1. https://doi.org/10.1038/nplants.2014.3.

Fustier, M. A., J. T. Brandenburg, S. Boitard, J. Lapeyronnie, L. E. Eguiarte, Y. Vigouroux, D. Manicacci, and M. I. Tenaillon. 2017. "Signatures of Local Adaptation in Lowland and Highland Teosintes from Whole-Genome Sequencing of Pooled Samples." *Molecular Ecology* 26 (10):2738–56. https://doi.org/10.1111/mec.14082.

Gage, J. L., D. Jarquin, C. Romay, A. Lorenz, E. S. Buckler, S. Kaeppler, N. Alkhalifah, et al. 2017. "The Effect of Artificial Selection on Phenotypic Plasticity in Maize." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01450-2.

Heerwaarden, J. van, J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. de Jesus Sanchez Gonzalez, and J. Ross-Ibarra. 2011. "Genetic Signals of Origin, Spread, and Introgression in a Large Sample of Maize Landraces." *Proceedings of the National Academy of Sciences* 108 (3):1088–92. https://doi.org/10.1073/pnas.1013011108.

Heerwaarden, J. van, M. B. Hufford, and J. Ross-Ibarra. 2012. "Historical Genomics of North American Maize." *Proceedings of the National Academy of Sciences* 109 (31):12420–25. https://doi.org/10.1073/pnas.1209275109.

Hufford, M. B., P. Lubinksy, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. "The Genomic Signature of Crop-Wild Introgression in Maize." *PLoS Genetics* 9 (5). https://doi.org/10.1371/journal.pgen.1003477.

Hufford, M. B, X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J-M. Chia, R. A. Cartwright, R. J Elshire, et al. 2012. "Comparative Population Genomics of Maize Domestication and Improvement." *Nat Genet* 44 (7):808–11. https://doi.org/10.1038/ng.2309.

Kusmec, A., S. Srinivasan, D. Nettleton, and P. S. Schnable. 2017. "Distinct Genetic Architectures for Phenotype Means and Plasticities in Zea Mays." *Nature Plants* 3 (9):715–23. https://doi.org/10.1038/s41477-017-0007-7.

Lee, K. M., and G. Coop. 2017. "Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data." *Genetics*, genetics.300417.2017. https://doi.org/10.1534/genetics.117.300417.

Leinonen, T., R. J.S. McCairns, R. B. O'Hara, and J. Merilä. 2013. "QST -FST Comparisons: Evolutionary and Ecological Insights from Genomic Heterogeneity." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3395.

Lobell, D. B, W. Schlenker, and J. Costa-Roberts. 2011. "Climate Trends and Global Crop Production since 1980." *Science (New York, N.Y.)* 333 (6042):616–20. https://doi.org/10.1126/science.1204531.

Marais, D. L Des, K. M Hernandez, and T. E Juenger. 2013. "Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment." *Annual Review of Ecology, Evolution, and Systematics* 44. Annual Reviews:5–29.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. S. Buckler, and J. Doebley. 2002. "A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping." *Proc Natl Acad Sci U S A* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

Piperno, D. R., and K. V. Flannery. 2001. "The Earliest Archaeological Maize (Zea Mays L.) from Highland Mexico: New Accelerator Mass Spectrometry Dates and Their Implications." *Proc Natl Acad Sci U S A* 98 (4):2101–3. https://doi.org/10.1073/pnas.98.4.2101.

Piperno, D. R., J. E. Moreno, J. Iriarte, I. Holst, M. Lachniet, J. G. Jones, A. J. Ranere, and R. Castanzo. 2007. "Late Pleistocene and Holocene Environmental History of the Iguala Valley, Central Balsas Watershed of Mexico." *Proceedings of the National Academy of Sciences* 104 (29). National Acad Sciences:11874–81.

Pyhäjärvi, T., M. B Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biol Evol* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler. 2015. "Recombination in Diverse Maize Is Stable, Predictable, and Associated with Genetic Load." *Proceedings of the National Academy of Sciences*, 201413864. https://doi.org/10.1073/pnas.1413864112.

Schlötterer, C., R. Tobler, R.Kofler, and V. Nolte. 2014. "Sequencing Pools of Individuals-Mining Genome-Wide Polymorphism Data without Big Funding." *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3803.

Schrider, D. R., and A. D. Kern. 2016. "S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning." *PLoS Genetics* 12 (3). https://doi.org/10.1371/journal.pgen.1005928.

Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski, J. Holland, M. Kruse-Peeples, et al. 2017. “Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America.” *Science* 357 (6350):512–15. https://doi.org/10.1126/science.aam9425.

Takuno, S., P. Ralph, K. Swart, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. 2015. “Independent Molecular Basis of Convergent Highland Adaptation in Maize.” *Genetics* 200 (4):1297–1312. https://doi.org/10.1534/genetics.115.178327.

Tanksley, S. D., and S. R. McCouch. 1997. “Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild.” *Science* 277 (5329):1063–66. https://doi.org/10.1126/science.277.5329.1063.

Tenaillon, M. I., J. U’Ren, O. Tenaillon, and B. S. Gaut. 2004. “Selection versus Demography: A Multilocus Investigation of the Domestication Process in Maize.” *Molecular Biology and Evolution* 21 (7):1214–25. https://doi.org/10.1093/molbev/msh102.

Vigouroux, Y., J. C. Glaubitz, Yoshihiro Matsuoka, Major M. Goodman, Jesús Sánchez G., and John Doebley. 2008. “Population Structure and Genetic Diversity of New World Maize Races Assessed by DNA Microsatellites.” *American Journal of Botany* 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097.

Wennersten, L., and A. Forsman. 2012. “Population-Level Consequences of Polymorphism, Plasticity and Randomized Phenotype Switching: A Review of Predictions.” *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2012.00231.x.

West-Eberhard, M. J.. 2003. *Developmental Plasticity and Evolution*. Oxford University Press.

Wright, S.I, I. V. Bi, S. G Schroeder, M. Yamasaki, J.ohn F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. “The Effects of Artificial Selection on the Maize Genome.” *Science (New York, N.Y.)* 308 (5726):1310–14. https://doi.org/10.1126/science.1107891.

Yang, J., S. Mezmouk, A. Baumgarten, E.S. Buckler, K. E. Guill, M.l D. McMullen, R. H. Mumm, and J. Ross-Ibarra. 2017. “Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize.” *PLoS Genetics* 13 (9). https://doi.org/10.1371/journal.pgen.1007019.

Zhou, Y., M. Massonnet, J. S. Sanjak, D. Cantu, and B. S. Gaut. 2017. “Evolutionary Genomics of Grape (Vitis Vinifera Ssp. Vinifera) Domestication.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (44):11715–20. https://doi.org/10.1073/pnas.1709257114.

Titre : Plasticité et adaptation génétique comme contributeurs de l'histoire évolutive du maïs cultivé et des formes sauvages apparentées

Mots clés : Maïs, téosinte, domestication, adaptation locale, expression génique

Les complexes d'espèces rassemblant des formes sauvages et cultivées offrent une opportunité unique de disséquer les déterminants de l'adaptation à une échelle de temps courte (depuis la domestication) et à une échelle de temps plus longue (sélection naturelle dans les populations sauvages). Nous avons étudié ici le complexe formé par le maïs cultivé, *Zea mays* ssp. *mays*, et les formes sauvages apparentées, les téosintes des sous-espèces *Zea mays* ssp. *mexicana* et ssp. *parviglumis*. Le maïs a été domestiqué à partir de cette dernière au Mexique, il y a environ 9 000 ans. Nous avons tout d'abord étudié l'histoire démographique du maïs en utilisant des données de re-séquençage (profondeur 24-53x) de 31 variétés locales de maïs couvrant sa distribution précolombienne. Nous avons confirmé l'existence d'un goulot d'étranglement pendant la domestication, entraînant une réduction importante de la taille effective de la population, ainsi que des effets de goulots en série au cours de sa dissémination post-domestication. Les effets indésirables de ces événements démographiques se traduisent par une augmentation des allèles délétères chez le maïs en comparaison des formes sauvages, particulièrement marquée chez les maïs andins. De façon intéressante, nous avons détecté des introgressions de la sous-espèce *mexicana* vers les maïs cultivés des régions montagneuses du Mexique, du Guatemala et du sud-ouest des États-Unis, qui réduisent la prévalence des allèles délétères. Nous avons ensuite étudié les changements plastiques dans l'expression des gènes en comparant les transcriptomes de maïs et de téosintes dans les conditions climatiques actuelles et des conditions proches de celles trouvées au moment de la domestication – basse température et faible teneur en CO_2. Nous avons identifié plus de 2 000 locus qui présentent une expression différentielle entre conditions chez les téosintes, mais dont l'expression ne varie pas chez les maïs. Ces résultats suggèrent une plus grande plasticité de réponse chez les téosintes, appuyée par l'observation de changements plus substantiels dans les réseaux de co-expression chez les téosintes comparativement au maïs. Nous avons ensuite recherché les signatures génomiques de l'adaptation locale dans six populations naturelles de la sous-espèce *parviglumis* (re-séquençage de profondeur 20-25x). Nous avons identifié grâce à scans génomiques, des locus présentant des traces de balayages sélectifs forts et doux. Le faible recouvrement de ces locus entre les populations indique une forte adaptation locale. Ainsi, l'adaptation génétique à une échelle géographique réduite, les introgressions répétées entre formes sauvages et domestiques et l'assimilation génétique qui « cimente » les phénotypes domestiqués initialement induits par des réponses plastiques à l'environnement, sont autant de mécanismes qui ont contribué à l'émergence et la diffusion du maïs cultivé.

université PARIS-SACLAY | ÉCOLE DOCTORALE Sciences du végétal : du gène à l'écosystème

Title : Plasticity and genetic adaptation as contributors to the evolutionary history of cultivated maize and its wild relatives

Keywords : Maize, teosinte, domestication, local adaptation, gene expression

Species complexes combining wild and cultivated forms provide a unique opportunity to dissect the determinants of adaptation at a short time scale (since domestication) and at a long-time scale (natural selection in wild populations). Here, we studied the complex formed by the cultivated maize, *Zea mays* ssp. *mays*, and related wild forms, teosintes subspecies *Zea mays* ssp. *mexicana* and ssp. *parviglumis.* Maize was domesticated from the later in Mexico about 9,000 years ago. We first studied the demographic history of maize using re-sequencing data (24-53x depth) of 31 local maize varieties covering its pre-Colombian distribution. We confirmed a bottleneck during domestication, resulting in a significant reduction in effective population size, as well as additional founder effects during its post-domestication diffusion. The adverse effects of these demographic events have resulted in an increase of deleterious alleles in maize compared to its wild relatives, especially in Andean maize. Interestingly, we detected introgression of *mexicana* subspecies into corn grown in the highlands of Mexico, Guatemala, and the southwestern United States, which reduced the prevalence of deleterious alleles.

We then studied the plastic changes in gene expression by comparing maize and teosinte transcriptomes under current climatic conditions and conditions similar to those encountered at the time of domestication - low temperature and low CO_2 content. We have identified over 2 000 loci that exhibit a differential expression between conditions in teosintes, but whose expression does not vary in maize. These results suggest a greater plastic response in teosintes, supported by observations of more substantial changes in co-expression networks in teosintes compared with maize. Finally, we searched for genomic signatures of local adaptation in six natural populations of the subspecies *parviglumis* (20-25x sequencing depth). Genomic scans have identified loci with hard and soft selective sweeps. The low overlap of these loci between populations indicates a strong local adaptation. Thus, genetic adaptation to a small geographical scale, repeated introgression between wild and domestic forms, and genetic assimilation that "cements" domesticated phenotypes initially induced by plastic responses to the environment, are all mechanisms that contributed the emergence and spread of cultivated maize.