ORIGINS OF TEMPERATE ADAPTATION IN MAIZE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kelly Lynn Swarts

January 2017

ORIGIN OF TEMPERATE ADAPTATION IN MAIZE

Kelly Lynn Swarts, Ph. D.

Cornell University 2017

Four thousand years ago, maize began migrating out of Mexico and into the geographically diverse Southwestern US (*1*, *2*). Maize spread quickly through the lowland deserts, but full maize agriculture in the uplands lagged for another 2,000 years (*1*, *3*). Why did maize agriculture fail to fully develop, despite the persistent presence of maize in the uplands? I tested the hypothesis that early maize in the Southwest US was not adapted to the uplands, specifically testing archaeological maize from the uplands at the beginning of agricultural intensification for early flowering.

To better understand temperate adaptation in maize, I led a project to generate tools for accurate imputation and projection of maize inbred haplotypes onto related individuals. We then projected whole genome sequence onto five diverse modern inbred maize populations to further elucidate the genetic architecture of maize flowering through mapping, machine learning and cross population predictions. We find that the genetic architecture of flowering is tightly linked to the temperate-tropical differentiation in American germplasm for four of the five populations, and suggest an independent temperate adaptation occurred in China after 1500. We also find that populations with individuals across the North American temperate adaptation predict others well. We further test that good cross population prediction extends to modern

Southwestern landraces by developing a mapping population for flowering time, and show that modern inbred populations can predict divergent modern landraces with high prediction accuracy for days to flowering. We extend flowering predictions to archaeological maize samples dating to the early period of agricultural adoption, and find that this population was already adapted. Using SNPs with the greatest population differentiation over the extremes of temperate-tropical adaptation, we find that temperate adaptation happened in situ in the Southwest, and early southwestern peoples selected on primarily standing variation.

1. J. B. Mabry (ed.), *Las Capas: Early Irrigation and Sedentism in a Southwestern Floodplain* (Center for Desert Archaeology, Tucson, 2008), *Anthropological Papers*.
2. E. K. Huber, in *Fence Lake Project* (Statistical Research Inc., Tempe, 2005), vol. 1: Introduction and Site Descriptions of *Technical Series*.
3. R. G. Matson, B. Chisholm, Basketmaker II Subsistence: Carbon Isotopes and Other Dietary Indicators from Cedar Mesa, Utah. *Am. Antiq.* **56**, 444–459 (1991).

BIOGRAPHICAL SKETCH

Kelly Swarts was born and raised in Ann Arbor, Michigan, where she spent much of her free time reading, playing hockey and wandering in the woods. She received a B.S. degree in Biology and Anthropology from the University of Michigan in 2005, completing a senior honors thesis with Dr. Richard Ford focused on land use by archaic peoples in the Northern Rio Grande Valley of New Mexico. Following graduation, she spent a summer at one of the archaeological sites that was the subject of her senior thesis then moved to a trailer on the border with Mexico to survey the Sonoran desert grasslands before finding work as a field archaeologist in Tucson, AZ. After a year in the field, she pursued an M.A. in Archaeology at Northern Arizona University, where she studied the meaning of wood choices used in archaeological structures in southern Arizona, graduating in 2008, and worked as a GIS analyst for the Coconino National Forest. Kelly did a summer internship at Crow Canyon Archaeological Center after graduation where she became fascinated with the question of the spread of maize into the Southwest, and also disillusioned with the ability of macromorphological approaches to trace the movement of maize. Kelly returned to Tucson, working as a field archaeologist in southern Arizona and across the west while she plotted how to get training in genetics. After she wasn't offered graduate admissions into the Department of Plant Breeding and Genetics at Cornell (in the nicest and most helpful way possible), she volunteered for a year in the laboratory of Dr. Vicki Chandler at the University of Arizona, where she focused on confirming regulators of the anthocyanin pathway gene *b1*, learning molecular biology and classical maize genetics in the process. She reapplied to graduate schools, and returned east in 2011 to begin Ph.D studies at Cornell in the Department of Plant Breeding and Genetics.

This dissertation is dedicated to the farmers of the Americas, past and present, whose ingenuity, persistence and vision gave us the diversity of maize we cherish today and to Dick Ford and Karen Adams, who gave me the tools and the framework to appreciate it.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Chip Aquadro, Jean-Luc Jannink and, especially, Ed Buckler, my advisor, from whom I have learned so much. Thank you so much for seeing the potential in an ignorant archaeologist, and I hope it was worth it. Additional thanks are due to Peter Bradbury, my unofficial but not unappreciated committee member, collaborator and advisor. Thanks to everyone at Native Seeds/SEARCH and the community of Patagonia, AZ for germplasm and productive and enriching summers in the field, and to the Holland Lab at NCSU who kindly grew a replicate of the landrace hybrid mapping population. Thanks to Jesus Sanchez for graciously contributing teosinte genotypes, and to Jeff Ross-Ibarra for the landrace panel drawn from across the Americas. Of course thanks to the ancient maize group in Tübingen and Vancouver, especially Hernán Burbano, Detlef Weigel and Rafal Gutaker in Germany and Michael Blake and R.G. Matson in Canada, without whom this project would not have been possible. Much thanks especially to all of the members of the Buckler Lab, whose contributions are too numerous to name, but especially Nick Lepak for managing and caring for my field experiments over the years and to Sara Miller for making sure I didn't die and a million other things. Finally, I couldn't have done this without my fellow graduate students and members of the plant breeding community. Thanks always to Sam Duwe, for tolerating my indecision and always supporting me on this path. Extra thanks to Laura, for all of her help, encouragement and for generally being wonderful. I am additionally grateful for the funding that made my time at Cornell possible. Generous support from a Cornell Fellowship, and NSF Grants #0922493, #1238014, and the USDA-ARS made my studies possible.

TABLE OF CONTENTS

INTRODUCTION

Motivation for study

The objective of my dissertation work addresses this question: why did it take so long for maize (*Z. mays* ssp. *mays*) agriculture to establish in temperate regions of the southwestern United States? Maize archaeological remains first appear just over 4000 years ago across the modern Southwest United States (*1–6*), but full agriculture, defined by increased storage and processing capacity, establishes rapidly in the lowlands of S. Arizona, but not in the uplands of the Colorado Plateau, less than 250 miles away, for another 2000 years. My dissertation addresses this problem by (1) developing tools for increasing power of genome-wide association and genomic prediction, (2) dissecting the biological basis and genetic architecture of early flowering, which is highly correlated with population structure in inbred temperate adapted maize, and (3) predicting flowering time, a complex trait critical for temperate adaptation, in archaeological maize samples from Turkey Pen Shelter, a 2,000-year-old archaeological site in the uplands of southeastern Utah.

Imputation tools for increasing power in association mapping and prediction

Because control of flowering time is highly quantitative, controlled by hundreds of loci, it is desirable to have as many high coverage markers across the genome as possible to detect and differentiate signals in highly recombinant populations. Much of the effort in imputation had been undertaken by the human

genetics community, who ethically cannot simplify populations through the development of controlled mating designs, which require fewer markers to analyze. The most effective of the human designed software with respect to imputation accuracy and computational capacity is Beagle v.4 (*7–11*), but Beagle assumes that the population is outcrossing, as is always the case for humans, and phases haplotypes before imputation. This approach is computationally intensive and does not take advantage of inbreeding, which biologically phases haplotypes.

I led an effort to develop an imputation algorithm, Fast Inbred Line Library ImputatioN (FILLIN), to increase power in genome-wide association studies (GWAS) and genomic prediction (GP). Inbreeding in crop plants generates a source of phased haplotypes not available in human research and FILLIN leverages inbred lines and segments for fast and accurate genotypic imputation in inbred lines and populations ($R^2 = 0.99$ in diverse inbreds). FILLIN identifies and collapses identical segments to generate higher coverage haplotypes, then matches these to low coverage target sequences and models recombination between haplotypes with a Hidden Markov Model. Imputation increases power for GWAS by maximizing the sample size for each test while allowing low-coverage, low-cost genotyping. FILLIN can also use this approach to model expensive whole-genome haplotypes onto phenotyped, lower-coverage genotypes. GWAS and GP on whole-genome data break the reliance on linked markers in populations with sufficient recombination, allowing for more precise identification of causal loci and accurate prediction.

Reanalysis of modern inbred populations for flowering time

Before I could hope to address flowering adaptation in southwestern archaeological samples, 2,000 years diverged from any modern samples, I conducted a large-scale analysis of diverse modern mapping populations to understand the genetic architecture of maize in a global context. Flowering time, typically through the metric of days to flowering, is well studied in maize (*12–18*, *18–24*), Arabidopsis (*15*, *25–30*), and other crop plants and models (*13*, *27*, *31–35*) due to high heritabilities (up to 0.96 in maize (*12*)) and the importance for adaptation and transferring alleles between tropical and temperate germplasm in a breeding context. Only a handful of loci, mostly in the autonomous flowering pathway, have confirmed effects for modulating flowering time either through mutagenesis or positional cloning, (*15*, *18*, *20*, *36–40*), but very few of these have common standing variation and have been mapped in the tens of structured and unstructured populations that have mapped flowering time in maize (*12*, *13*, *19*, *22*, *41*, *42*). These studies have all confirmed that the genetic basis for flowering in maize is quantitative and dominantly additive in inheritance.

Comparative population and synteny studies found moderate overlap between significant loci between maize populations (*19*, *22*), but also between maize and other grasses (*13*, *35*). Although cross-population predictions using genomic prediction models are common in animal breeding (*43–45*), there has been only one instance in maize (*42*). Lehermeier et al looked at the cross-population prediction accuracy between the two European heterotic groups, the Dent and Flint populations in the European NAM, and found negative cross

population prediction accuracy. This is unsurprising, since by definition the two heterotic groups have been bred in complementary and opposite directions. In order to predict the archaeological samples, I needed to establish the validity and parameters necessary for good cross-population prediction for flowering time, which I did in a joint analysis of five previously published populations from across the world.

To detect loci responsible for flowering in maize, I reanalyzed published flowering phenotypes from two mapping populations, a diversity panel and a linked biparental population (Nested Association Mapping, or NAM), focused on maize germplasm from the Americas, two European NAM populations, and a Chinese NAM population using FILLIN projected whole-genome genotypes. Whole genome projected data allows for detection of causal or nearly causal loci when recombination is sufficient, which can be used for stable prediction of unrelated, or archaeological, maize since recombination cannot disassociate a marker when it is the causal variant.

I analyzed these populations using four different methods for detecting association between phenotypes and genotypes; these include two single marker approaches, genome-wide association studies (GWAS) and resampling mean inclusion probabilities (RMIP), and two regional approaches, Regional Heritability Mapping (RHM) (*46*, *47*) , and a novel method developed for this study, Boosted Regional Heritability Mapping (BRHM). None of these methods found substantial overlap in significant regions between populations, indicating on the surface a lack of shared genetic architecture, or lack of segregating variants. However, I also performed cross-population predictions (using RR-BLUP) across the five populations, and many of the

populations were able to predict each other with high accuracy, especially training on the diversity panel; additionally, for those populations predicted with low accuracy, using subsets of the results significant in most of the populations could increase prediction accuracy. This suggested that there was in fact a shared genetic architecture across four of the populations (the Chinese population did not accurately predict, nor was it well predicted by any of the other populations). Machine learning (with a random forest classifier) using the GWAS results as the response and other mapping results or F_{ST} statistics for population structure could accurately cross-predict all of the populations except for the Chinese NAM. F_{ST} estimates between tropical and temperate maize germplasm from the Americas were especially good predictors of flowering time, except in Chinese NAM, which we suggest carries alleles from a distinct temperate adaptation that took place in China after 1500.

These results suggest that the lack of consistency in mapping results can be primarily attributed to differences in the structure of linkage equilibrium and diversity in the parents among the mapping populations, rather than lack of shared genetic architecture. Because distant cross-population prediction worked in inbreds, these results suggested that it should be possible to predict the archaeological maize samples from Turkey Pen, especially since they were located at the juncture between temperate and tropical N. America.

Prediction for temperate adaptation in Turkey Pen

Prediction in ancient DNA has traditionally focused on traits controlled by one or a few genes, such as pigmentation in humans (*48*). Ancient DNA extraction

methodology is improving rapidly (*49*), and now genomic (*50–54*) and increasingly whole-genome (*55*, *56*) DNA is readily available. Adaptation traits such as flowering time are quantitatively inherited, increasing the complexity of prediction. Additionally, prediction in ancient DNA cannot be verified, and the argument for accurate prediction must be built from modern proxies. In this study, I establish cross-population prediction accuracies from the modern inbred populations in the flowering time study and from a population of modern descendant landraces.

Turkey Pen Shelter is a dry cave site in upland southeastern Utah, and the study samples date to the time period when people first began to intensify agricultural production in the upland regions, including at Turkey Pen (*57*). The preservation at Turkey Pen was exceptional, and 14 samples had greater than 80% endogenous maize DNA, allowing for 5-20X coverage whole genome resequencing. Population genetic analysis of Turkey Pen genotypes, in conjunction with over 1000 low-density genotypes from landraces and the wild progenitor, teosinte (*Z. mays* ssp. *parviglumis* and *Z. mays* ssp. *mexicana*), showed that Turkey Pen was most closely related to modern temperate Southwest samples.

Because it is not possible to validate flowering time in archaeological samples, I assembled a modern landrace mapping population that spans the tropical germplasm from Northern Mexico to temperate adapted germplasm from the upland Southwest. These individuals are outcrossing landraces, and only distantly related to the inbred germplasm used in the reanalysis for flowering time. Because they are also closely related to the Turkey Pen samples, prediction accuracies should be comparable to those of the modern mapping population. Turkey Pen flowers comparably early to

modern temperate Southwest germplasm, suggesting that Turkey Pen maize was in fact sufficiently adapted with respect to flowering time. We also find that Turkey Pen shared temperate adapted SNPs with Southwestern populations, but not with those in north Mexico, suggesting an *in situ* temperate adaptation. Additionally, the top SNPs associated with temperate adaptation are segregating in teosinte, suggesting that selection for temperate adaptation relied on standing variation.

We additionally predicted less complex, but culturally informative, traits for yellow kernel color (*y1*) and flour endosperm (*su1*). Maize is central to the culture of indigenous southwestern peoples (*58*), and different varieties are required for both cuisine and, critically, for the proper performance of ritual. We find that Turkey Pen mostly contained the ancestral flint or pop type endosperm at the *su1* locus, but one individual was segregating for the flour type allele, prized by modern southwestern people. Vitamin A is a critical dietary nutrient for human health in the developing world (*59*), and would have also been important for ancient farmers. Vitamin A accumulates in the endosperm of yellow corn (*60*), and we looked at the recently selected *y1* gene in Turkey Pen to determine the presence of yellow kernel color in Turkey Pen (*61*). At *y1*, we find that Turkey Pen segregates for yellow kernel color, and neighbor-joining trees also suggest that yellow color may have originated in the Southwest. This is intriguing, as yellow is one of the six color directions (including zenith and nadir) that organize ritual space and is part of a pan-New World cosmology (*62*); in the Southwest today, indigenous groups require the full complement of maize kernel colors for ritual performance. I was not able to predict red or blue color in Turkey Pen, as described in Appendix A.

Discussion of outcomes and further directions

My dissertation began with the question: Why did it take so long for maize agriculture to establish in the temperate regions of the Southwestern United States? After building the tools to increase power in GWAS and GP with FILLIN, further elucidating the genetic architecture underlying flowering in the largest combined mapping panel, and demonstrating the accuracy of distant cross population prediction for flowering time in maize, the short answer is because it took 2,000 years to select the right combination of standing variants for maize agriculture to succeed in the uplands. While the Turkey Pen results are well supported, additional questions remain. Turkey Pen is only one location; how adapted was maize at contemporaneous sites across the Southwest? How quickly did the temperate germplasm travel, and by what routes? How adapted was maize in the period just before the introduction of agriculture in the uplands? As sequencing becomes increasingly affordable, and ancient DNA methods continue to improve, the answers to these questions are increasingly in reach. More pressing questions refer to the future of maize, and maize adaptation to a rapidly changing climate. The lessons from Turkey Pen are that it happened once, and primarily from standing variation. Maintaining germplasm diversity should enable further maize adaptation going forward.

REFERENCES

1. J. B. (ed. . Mabry, *Archaeological Investigations of Early Village Sites in the Middle Santa Cruz Valley* (Center for Desert Archaeology, Tucson, 1998), *Anthropological Papers*.
2. J. B. (ed. . Mabry, *Las Capas: Early Irrigation and Sedentism in a Southwestern Floodplain* (Center for Desert Archaeology, Tucson, 2008), *Anthropological Papers*.
3. D. A. (ed. . Gregory, *Excavations in the Santa Cruz River Floodplain: The Early Agricultural Period Component at Los Pozos*. (Center for Desert Archaeology, Tucson, 2001), *Anthropological Papers*.
4. D. A. Gregory, N. M. Stevens, F. L. Nials, M. R. Schurr, M. W. (editors) Diehl, *Excavations in the Santa Cruz River Floodplain: Further Investigations at Los Pozos*. (Center for Desert Archaeology, Tucson, 2007), *Anthropological Papers*.
5. E. K. Huber, C. V. . (editors) West, *Fence Lake Project. Archaeological Data Recovery in the New Mexico Transportation Corridor and First Five-Year Permit Area, Fence Lake Coal Mine Project, Catron County, New Mexico* (Statistical Research Inc., Tempe, 2005), *Technical Series*.
6. F. E. Smiley, The Agricultural Transition in the Northern Southwest: Patterns in the Current Chronometric Data. *Kiva*. **60**, 165–189 (1994).
7. B. L. Browning, S. R. Browning, Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics*. **194**, 459–471 (2013).
8. B. L. Browning, S. R. Browning, A Fast, Powerful Method for Detecting Identity by Descent. *Am. J. Hum. Genet*. **88**, 173–182 (2011).
9. S. R. Browning, B. L. Browning, High-Resolution Detection of Identity by Descent in Unrelated Individuals. *Am. J. Hum. Genet*. **86**, 526–539 (2010).
10. S. R. Browning, B. L. Browning, Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am. J. Hum. Genet*. **81**, 1084–1097 (2007).
11. S. R. Browning, Multilocus Association Mapping Using Variable-Length Markov Chains. *Am. J. Hum. Genet*. **78**, 903–913 (2006).
12. E. S. Buckler *et al*., The Genetic Architecture of Maize Flowering Time. *Science*. **325**, 714–718 (2009).
13. F. Chardon *et al*., Genetic Architecture of Flowering Time in Maize As Inferred From Quantitative Trait Loci Meta-analysis and Synteny Conservation With the Rice Genome. *Genetics*. **168**, 2169–2185 (2004).
14. F. Chardon, D. Hourcade, V. Combes, A. Charcosset, Mapping of a spontaneous mutation for early flowering time in maize highlights contrasting allelic series at two-linked QTL on chromosome 8. *Theor. Appl. Genet*. **112**, 1–11 (2005).
15. J. Colasanti *et al*., The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genomics*. **7**, 158 (2006).
16. O. N. Danilevskaya, X. Meng, B. McGonigle, M. G. Muszynski, Beyond flowering time: pleiotropic function of the maize flowering hormone florigen. *Plant Signal. Behav*. **6**, 1267–1270 (2011).

17. E. Durand *et al*., Flowering Time in Maize: Linkage and Epistasis at a Major Effect Locus. *Genetics*. **190**, 1547–62 (2012).
18. S. Ducrocq *et al*., Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information. *Genetics*. **178**, 2433–2437 (2008).
19. Y. Li *et al*., Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J*. **86**, 391–402 (2016).
20. M. G. Muszynski *et al*., delayed flowering1 Encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. *Plant Physiol*. **142**, 1523–1536 (2006).
21. J. Romero-Navarro *et al*., Tapping a world of genetic variation: surveying flowering time diversity in maize landraces. **Nature Genetics (in review)**.
22. S. (1967) Salvi, S. Castelletti, R. Tuberosa, An updated consensus map for flowering time QTLS in maize. *MAYDICA*. **54**, 501–512 (2009).
23. J. Xu *et al*., The Genetic Architecture of Flowering Time and Photoperiod Sensitivity in Maize as Revealed by QTL Review and Meta Analysis. *J. Integr. Plant Biol*. **54**, 358–373 (2012).
24. J. M. Thornsberry *et al*., Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet*. **28**, 286 (2001).
25. H. Guo, H. Yang, T. C. Mockler, C. Lin, Regulation of Flowering Time by Arabidopsis Photoreceptors. *Science*. **279**, 1360–1363 (1998).
26. U. Johanson *et al*., Molecular Analysis of FRIGIDA, a Major Determinant of Natural Variation in Arabidopsis Flowering Time. *Science*. **290**, 344 (2000).
27. G. G. Simpson, C. Dean, Arabidopsis, the Rosetta Stone of Flowering Time? *Science*. **296**, 285–289 (2002).
28. P. Suárez-López *et al*., CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature*. **410**, 1116–1120 (2001).
29. M. J. Aukerman, H. Sakai, Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. *Plant Cell Online*. **15**, 2730–2741 (2003).
30. I. Ausín, C. Alonso-Blanco, J. A. Jarillo, L. Ruiz-García, J. M. Martínez-Zapater, Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat. Genet*. **36**, 162–166 (2004).
31. J. Cockram *et al*., Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *J. Exp. Bot*. **58**, 1231–1244 (2007).
32. P. Q. Craufurd, T. R. Wheeler, Climate change and the flowering time of annual crops. *J. Exp. Bot*. **60**, 2529–2539 (2009).
33. D. A. Laurie, Comparative genetics of flowering time. *Plant Mol. Biol*. **35**, 167–177 (1997).
34. B. Ma *et al*., Characterization and comparison of key genes involved with flowering time regulation from Arabidopsis thaliana, Oryza sativa and Zea mays. *Afr. J. Biotechnol*. **12**, 353–363 (2013).

35. E. S. Mace, C. H. Hunt, D. R. Jordan, Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* **126**, 1377–1395 (2013).
36. S. Salvi *et al.*, Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol.* **48**, 601–613 (2002).
37. X. Meng, M. G. Muszynski, O. N. Danilevskaya, The FT-like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. *Plant Cell.* **23**, 942–960 (2011).
38. T. A. Miller, E. H. Muslin, J. E. Dorweiler, A maize CONSTANS-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta.* **227**, 1377–1388 (2008).
39. K. Bomblies, J. F. Doebley, Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. *Genetics.* **172**, 519–531 (2006).
40. N. D. Coles, M. D. McMullen, P. J. Balint-Kurti, R. C. Pratt, J. B. Holland, Genetic Control of Photoperiod Sensitivity in Maize Revealed by Joint Multiple Population Analysis. *Genetics.* **184**, 799–812 (2010).
41. M. C. Romay *et al.*, Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55 (2013).
42. C. Lehermeier *et al.*, Usefulness of Multiparental Populations of Maize (Zea mays L.) for Genome-Based Prediction. *Genetics.* **198**, 3–16 (2014).
43. M. A. Pérez-Cabal, A. I. Vazquez, D. Gianola, G. J. M. Rosa, K. A. Weigel, Accuracy of Genome-Enabled Prediction in a Dairy Cattle Population using Different Cross-Validation Layouts. *Front. Genet.* **3**, 27 (2012).
44. B. J. Hayes, P. J. Bowman, A. C. Chamberlain, K. Verbyla, M. E. Goddard, *Genet. Sel. Evol.*, in press.
45. R. F. Brøndum *et al.*, Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *J. Dairy Sci.* **94**, 4700–4707 (2011).
46. Y. Nagamine *et al.*, Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. *PLoS ONE.* **7**, e46501 (2012).
47. M. Shirali *et al.*, Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. *Heredity.* **116**, 333–338 (2016).
48. W. Branicki, M. Kayser, in *eLS*, John Wiley & Sons Ltd, Ed. (John Wiley & Sons, Ltd, Chichester, UK, 2015; http://doi.wiley.com/10.1002/9780470015902.a0023851), pp. 1–10.
49. A. W. Briggs *et al.*, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
50. H. N. Poinar *et al.*, Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science.* **311**, 392–394 (2006).
51. M. L. Carpenter *et al.*, Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am. J. Hum. Genet.* **93**, 852–864 (2013).

52. H. A. Burbano *et al*., Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science*. **328**, 723–725 (2010).
53. R. R. da Fonseca *et al*., The origin and evolution of maize in the Southwestern United States. *Nat. Plants*. **1**, 14003 (2015).
54. Q. Fu *et al*., The genetic history of Ice Age Europe. *Nature*. **534**, 200–205 (2016).
55. M. Meyer *et al*., A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*. **338**, 222–226 (2012).
56. M. Meyer *et al*., Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*. **531**, 504–507 (2016).
57. R. G. Matson, B. Chisholm, Basketmaker II Subsistence: Carbon Isotopes and Other Dietary Indicators from Cedar Mesa, Utah. *Am. Antiq*. **56**, 444–459 (1991).
58. R. . Ford, The Color of Survival. *Discovery*, 17–30 (1980).
59. E. J. Johnson, The role of carotenoids in human health. *Nutr. Clin. Care Off. Publ. Tufts Univ*. **5**, 56–65 (2002).
60. R. Vallabhaneni, E. T. Wurtzel, Timing and Biosynthetic Potential for Carotenoid Accumulation in Genetically Diverse Germplasm of Maize. *Plant Physiol*. **150**, 562–572 (2009).
61. K. Palaisa, M. Morgante, S. Tingey, A. Rafalski, Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. U. S. A*. **101**, 9885–9890 (2004).
62. J. H. Hill, The Flower World of Old Uto-Aztecan. *J. Anthropol. Res*. **48**, 117–144 (1992).

CHAPTER 2

FSFHAP (FULL-SIB FAMILY HAPLOTYPE IMPUTATION) AND FILLIN (FAST, INBRED LINE LIBRARY IMPUTATION) OPTIMIZE GENOTYPIC IMPUTATION FOR LOW-COVERAGE, NEXT-GENERATION SEQUENCE DATA IN CROP PLANTS

Abstract

Next-generation sequencing technology such as Genotyping-by-Sequencing (GBS) made low-cost, but often low-coverage, whole-genome sequencing widely available. Extensive inbreeding in crop plants provides an untapped, high quality source of phased haplotypes for imputing missing genotypes. We introduce Full-Sib Family Haplotype (FSFHap), optimized for full-sib populations, and a generalized method, Fast Inbred Line Library ImputatioN (FILLIN), to rapidly and accurately impute missing genotypes in GBS-type data. FSFHap and FILLIN impute missing genotypes with high accuracy in GBS-genotyped maize inbred lines and breeding populations, while Beagle v.4 is still preferable for diverse heterozygous populations. FILLIN and FSFHap are implemented in TASSEL 5.0.

Background

The number of genotyped individuals available to researchers has vastly increased in recent years due to the advent of low-cost, genome-wide genotyping platforms, such as Genotyping-by-Sequencing (*1*) (GBS). GBS provides a reduced representation of the genome by targeting sequences adjacent to restriction enzyme cut sites, enabling parity in read location across samples. By adding barcoded adapter

sequences to the restriction-digested DNA, up to 384 samples can be multiplexed in one flowcell lane. However, the resulting GBS data may have high rates of missingness and heterozygote undercalling, depending on genome size, genome structure, and the number of samples combined. To effectively use GBS sequence data while maintaining low costs, we thus need a mechanism to impute these missing genotypes.

Missing data is often a function of genome size and degree of multiplexing, where some sequences are simply not sampled when the genome size is large or many samples are combined in a flowcell. Heterozygote undercalling is also a function of low coverage sampling; to call a heterozygote for a given genotype, that genotype must be covered by at least two reads, and those reads must be from different chromosomes. In the case of maize, an organism with a genome size of approximately 2.3 Gbp (*2*), in GBS data with an overall coverage of 0.6X we expect only 12% of heterozygous sites to be called *correctly* based on binomial sampling from a Poisson distribution.

Missing data can also reflect true biologically missing sequence due to small insertions or deletions or larger structural variants in the genome. Because these missing data provide a real biological signal, it is desirable to capture this type of missing data in imputation. In maize, for example, not only is allelic diversity high (*3*), but 70% of genes and 90% of the genome shows structural variation in a panel of only 103 diverse inbred maize and teosinte lines (*4*). Maize is not unique in this respect; many agronomically important crop plants show a similar pattern of high structural variation and allelic diversity (*5*–*9*). Accurate imputation of missing sequence data

increases power and can improve for the results of downstream applications such as genetic mapping and genomic prediction (*10*).

Imputation benefits genome-wide association and linkage mapping studies by accurately identifying rare variants, which in turn increases the power to detect statistical associations (*11*). Imputation is expected to provide the greatest benefit for mapping studies when linkage disequilibrium between markers is low. This is often the case in natural populations or primarily outcrossing species, since fewer markers are present on each haplotype to tag a statistical genotype-phenotype association. (*11*) Even crop species, which because of recent breeding efforts often have extended linkage disequilibrium, only share short haplotypes when comparing distantly related individuals. Accurate imputation is thus critical for effectively using the output of low-coverage, low-cost genotyping platforms such as GBS.

While, for the reasons stated above, accurate imputation increases the value of low-cost, low-coverage genotyping, much of the available software for imputation has been tailored for humans (. Humans are a highly heterozygous outcrossing species with no controlled mating designs, little inbreeding, and much less structural variation than that observed in crop plants. Because of this, the available implemented algorithms are not optimized to accurately impute, or leverage unique information from, crop systems. While there are crop specific algorithms that have been been developed for unordered markers, (*15*) for known pedigrees, (*16*) and in the context of genomic prediction (*17*, *18*) most of these are not publically implemented, and here we present a solution.

A number of agronomically important systems use inbred lines extensively, (e.g., maize, rice, wheat, soybean, barley, sorghum), and many of these have structured population resources for mapping traits of interest (*19–21*). Very accurately mapping the recombination break points in these populations is desirable for fine mapping studies and for appropriately assigning effect estimates to the proper parent in association studies. We present here Full Sib Family Haplotype Imputation (FSFHap), a fast and accurate imputation algorithm optimized for full-sib families. FSFHap is intellectually similar to methods published for detecting recombination breakpoints in *Drosophila*. (*22*, *23*)

Researchers and breeders in the public sector have increased access to genotyping at large scale, but most of the resulting sequence is low coverage (e.g., GBS at 0.5X to 2X coverage per site). Since most of the available algorithms are for genomic prediction in crops and do not output genotype estimates (*17*, *18*), or are tailored for human populations (*13*, *14*, *24*), we developed a genotype imputation algorithm that allows for missing data in the imputed file and leverages information from inbred segments. Additionally, populations in crop plants are often large, so computational time should scale linearly to enable fast breeding decisions. The algorithm should also tolerate high amounts of missing data in the unimputed set, as crop plants are often sequenced at lower depth than human populations.

We present FILLIN (Fast, Inbred Line Library ImputatioN) here, a fast and accurate generalized imputation strategy built on the FSFHap algorithm that leverages inbred segments from large but sparse genotypic datasets to identify parental haplotypes and impute missing genotypes. FILLIN was optimized for crop plants,

typically outcrossing species with high structural diversity and rapid LD decay but with widely shared alleles, tolerance of inbreeding, and low to moderate inbreeding coefficients even in heterozygous accessions. (*4*, *25*, *26*) The biology of most crop plants suggests a different model for genotype imputation than those developed for humans, since we can reasonably assume phase for inbred lines and inbred segments to impute rapidly and accurately. Both the FSFHap and FILLIN algorithms are implemented in TASSEL 5.0. (*27*)

We compare the speed and accuracy of our novel algorithms on inbred and outbred maize (*Zea mays*) genotyped using GBS (*1*) with Beagle v.4. (*24*) Beagle v.4 was chosen because in early tests we found it to be the most powerful and comparable published algorithm available: it does not require an external haplotype library, it accepts high levels of missing data, and it has the computational speed to impute whole genomes or chromosomes. We also compare accuracy between FSFHap, FILLIN, and Beagle v.4 in a full-sib family recombinant inbred line (RIL) population.

Materials and Methods

Algorithms

Viterbi Algorithm: Both FSFHap and FILLIN rely on a Hidden Markov Model (HMM) to detect recombination break points between haplotypes. HMMs define genotype as the true, unobserved genotype and the SNP or sequence variant calls made by the sequencing pipeline as the observations. Using this formulation, the problem of imputation can be restated as the problem of determining the unobserved genotype that best explains the observed data. If for a given sample, **y** is a vector of

the observed SNP calls, **g** is a vector of unknown, unobserved genotypes, and **M** is a probability model describing the data and the genotypes, then one approach is to seek to maximize the likelihood L (**y**|**g**, **M**). In general, maximizing L for genotypes of n sites requires evaluating this likelihood for each of 2^n possible genotypes, an impossible task for 100 markers, let alone for the hundreds of thousands of markers in the data described here. Fortunately, because nucleotides are arranged sequentially on a chromosome, the problem can be modeled as a Markov chain. Doing so allows the use of the Viterbi algorithm (*28*) to identify the genotype, **g**, that maximizes the likelihood of the observed data. The Viterbi algorithm only needs to evaluate a small, tractable subset of the potential genotypes in order to maximize the likelihood.

Applying the Viterbi algorithm requires defining two separate probability matrices. Taken together these two probability matrices determine **M**, the probability model. The first is a transition probability matrix, which describes the probability of each possible genotype at a site given the genotype at the previous site for all possible genotypes at the previous site. The second is an emission probability matrix, which describes the probability of observing each possible allele call, given each possible genotypic state. This probability matrix has to capture the both the probability of a genotyping error and the probability that only one of the two possible alleles was observed at a heterozygous site, which results in that site being incorrectly scored as homozygous. In a classic hidden Markov chain, both probability matrices are constant across all sites. In our application, we treat the emission probabilities as constant, but allow the transition probability to vary depending on distance between sites and location in the genome.

Both the transition and emission matrices are estimated from the data set being imputed, which is known to contain errors. An EM method is used to improve that estimate. Because the imputed data provides a better indication of the actual genotypes than the original data, after the initial imputation, the imputed states can be used to make new estimates of the probability matrices. This process is repeated to convergence. Estimating the probability matrices is the expectation step. Applying the Viterbi algorithm constitutes the maximization step.

Initializing the matrices: Many of the DNA samples in our data (as is common with other crop studies) were created by bulking DNA from several plants. Most were presumed to be homozygous but often had residual heterozygosity or heterogeneity. Typically, the bulked plants were progeny of a single self-pollinated plant. In that case, the progeny represented a random sample of 2n gametes from the parent, where n is the number of progeny bulked. As a result, for a single bulked sample, the minor allele frequency at a heterozygous site, instead of being 0.5 as it would have been for a DNA sample of a singleparent plant, ranged from 0 to 0.5 with probabilities equal to 2n draws from a binomial distribution with p=0.5. Within a sample the allele frequencies at adjacent segregating sites will be expected to be the same, since they all represent the same sample. To accommodate chromosome segments with different allele frequencies, we allow for 5 genotype states, representing homozygous A, 3A:1B, 1A:1B, 1A:3B, and homozygous B. The initial emission probabilities were set as shown in Table 1.

Table 1. The Viterbi Algorithm initial emission probability matrix, P (Allele Call|State). Note that all of the rows sum to 1.

	Allele Call		
State	A	H	B
AA	.998	.001	.001
3A:1B	0.6	0.2	0.2
1A:1B	0.4	0.2	0.4
1A:3B	0.2	0.2	0.6
BB	.001	.001	.998

The transition probabilities between states at adjacent sites are calculated differently for the two algorithms we present here. For FILLIN, the transition matrix is fixed (Table 2). For FSFHap, the transition matrix is dependent on the expected rate of recombination and the distance between the sites. The transition probability between the different states was estimated using all intervals between non-missing markers, then adjusted based on the ratio of the actual interval to the average interval length. An initial estimate was based on expected recombination rates. Convergence of the EM algorithm is not dependent on the initial estimate as long as it is reasonable.

Table 2. The Viterbi Algorithm transition probability matrix for FILLIN.

State	AA	3A:1B	1A:1B	1A:3B	BB
AA	0.999	0.0001	0.0003	0.0001	0.0005
3A:1B	0.0002	0.999	0.00005	0.00005	0.0002
1A:1B	0.0002	0.00005	0.999	0.00005	0.0002
1A:3B	0.0002	0.00005	0.00005	0.999	0.0002
BB	0.0005	0.0001	0.0003	0.0001	0.999

FSFHap: As with most imputation methods, FSFHap begins by identifying haplotypes. In the case of a bi-parental population when the objective is to find recombination breakpoints, the interest is in determining which chromosome segments are identical by descent (IBD) from which parent. Alternatively one could also look for segments IBD from the grandparents. If the population descends from homozygous

inbred parents, as is the case for many crop populations, it makes sense to restrict our attention to segments IBD from the parents. The algorithm thus attempts to identify two parental haplotypes, ignoring any sites that happen to be heterozygous in either of the parents. The algorithm as written has not been tested for families derived from a cross between outbred, heterozygous parents but handles segments that are heterozygous in one parent by using only the homozygous sites in those segments.

One assumption of the Viterbi algorithm is that the probability of an error at one site is independent of other sites, which is not necessarily the case for sites in the same GBS sequence read. Consequently, before identifying parental haplotypes, if any SNP pair came from the same tag, one of the pair is deleted from the dataset. Next, within each bi-parental family, the algorithm clusters lines in a window of 50 variant sites at the beginning of a chromosome using a custom clustering method (described below). Large clusters identify parental haplotypes while small clusters are generally heterozygous or contain individuals with genotyping errors. If the first window tested has more than two large clusters, the next adjacent window is checked until a window with only two large haplotype clusters is found. Once a window is found meeting that criterion, it serves as an anchor for determining the next haplotype block. Starting immediately after the anchor window, the allele calls for subsequent sites are evaluated one at a time. Labeling the two anchor haplotypes A and B, if a site's allele calls for RILs in haplotype A are mostly the same and the allele calls for the RILs in haplotype B are mostly different from the haplotype A majority allele, the site is assigned to the correct haplotype. Otherwise the site is removed from the dataset. The majority allele within the respective haplotypes is recorded and the next

adjacent site evaluated. Once alleles have been assigned to A and B for 50 additional sites, these 50 sites become the new anchor window. Two clusters are formed of lines that are within a minimum distance of each of the two new haplotype sequences. The entire process is repeated to extend each haplotype another 50 sites until the end of the chromosome is reached.

Once the parental haplotypes are identified, the progeny are scored as parent A, parent B, or heterozygous at each site. Then for each of the progeny individually the Viterbi algorithm is applied to the non-missing sites to determine the most likely genotype given the observations. The missing sites are then imputed based on the flanking non-missing markers. If the flanking markers match (both are A, B, or H), then the missing site is imputed to the same value as the flanking markers; otherwise it is left missing. Finally, the sites are converted back to nucleotides by examining the original nucleotide calls of all the individuals in the A and B classes at each site.

Custom clustering method: Because of data scarcity, standard hierarchical clustering methods perform inadequately for classifying haplotypes formed from a limited number of sites. Of these, the "complete" method, which calculates distance between clusters as the maximum pairwise distance gives the most useful results. The algorithm described here modifies that method by defining distance between two haplotypes as the sum of the differences across sites, where a site difference is 2 for different homozygotes, 1 if one site is homozygous and the other heterozygous, or 0 if either haplotype had a missing value. Further, each cluster contains all of the individuals less than a given distance from all the other individuals in the cluster. Because of missing data, this means some individuals could belong to more than one

cluster. Each cluster is interpreted as all the individuals that could have the same genotype for the entire window. Because heterozygous sites are often called as one of the possible homozygotes at random, heterozygous individuals do not form large clusters. A useful measure of cluster size is the sum of 1/ (the number of clusters to which an individual belonged) over the individuals in a cluster.

FILLIN: FILLIN (Fast, Inbred Line Library ImputatioN) is optimized to leverage inbred segments for fast imputation in very large, sparse datasets. Like other algorithms, we separate haplotype generation and imputation. (*13*, *14*) FILLIN first generates high coverage haplotypes from inbred lines and inbred segments by dividing the genome into non-overlapping windows. Within each window, the Hamming distance is used to cluster sequences that share highly similar genotypes, and these clusters are then collapsed to generate higher-coverage haplotypes. To calculate distance, sites with a missing genotype call in either taxon are ignored and the distance between a heterozygous and homozygous genotype is considered to be half the distance of one homozygous genotype to the alternate homozygote. To best represent high levels of structural variation, we do not require complete coverage for the resulting haplotypes. Additionally, a small amount of residual heterozygosity propagates to the resulting haplotype donor files, as the algorithm makes no effort to phase residual heterozygous genotypes. This approach results in very fast haplotype generation (Figure 8), but is less sensitive than other algorithms if the samples are highly heterozygous, since we make no effort to phase. (*24*, *29*, *30*) The haplotype generation step should always be performed with all of the samples available, as (1) small-scale haplotype windows may be replicated across even genetically distant

individuals, and (2) the algorithm requires at least two samples to generate a haplotype.

To impute these higher coverage haplotypes back to the target samples, FILLIN takes an iterative approach to imputation. First it selects possible donors based on shared minor alleles within each window. Shared minor alleles are particularly informative, since most of the minor allele states derive from more recent mutation and when two taxa share these alleles it suggests recent common ancestry. FILLIN then ranks haplotype donors by genetic distance to the taxon being imputed (again, looking only within the current window); if the distance falls below a user-specified threshold, it then imputes one haplotype to the entire window (Figure 1-1a). If this fails, the algorithm looks for two donors that can together adequately explain the minor alleles in the entire window (Figure 1-1b). This is functionally like assuming that this represents a recombination break point between two known haplotypes, and it uses the above Viterbi Hidden Markov algorithm to decide where to switch. The Viterbi is run in both directions, with disagreements defaulting to the genotype with the longest path length (i.e., highest likelihood).

If one or two donors cannot be found to explain the entire window, the algorithm repeats this process for smaller, 64-site windows within the larger window. Each 64-site window serves as a focus, and the algorithm extends out right and left until this window (the "focus block") contains a minimum number of minor alleles to calculate Hamming distance. FILLIN then attempts to impute based on single haplotype (Figure 1-2a) and the two-haplotype (Figure 1-2b) Viterbi imputation, if distance between the donor and target falls below a threshold. If these attempts fail to

explain sufficient minor alleles, the algorithm will then find two haplotypes that explain the minor alleles at a higher error threshold, combine these two haplotypes, and impute using this combined haplotype sequence, modeling the region as heterozygous (Figure 1-2c). If *this* search fails, that 64-site window will not be imputed. Because low-coverage sequence data often results in undercalling heterozygotes, an option to resolve homozygotes predicted to be heterozygous is available for all imputation types except 2c.

The maximum genetic distance thresholds for the focus block are customizable by the user, but by default are set more stringently than those for the entire window since the focus blocks are shorter and are expected to contain fewer sequencing errors if the haplotype is truly IBD to the target. These thresholds are also different for outbred versus inbred taxa, since when two haplotypes explain the minor alleles of a target sequence in an outbred taxon, it is more probable that the target sequence is heterozygous rather than a segment containing a recombination between two inbred haplotypes. For a taxon that falls above a user-defined per taxon heterozygosity threshold (is outbred), the threshold for using Viterbi (2b) is set to 0. If a taxon is considered generally inbred, any discrepancy between the two combined haplotypes that generates a heterozygous genotype (2c), is set to missing.

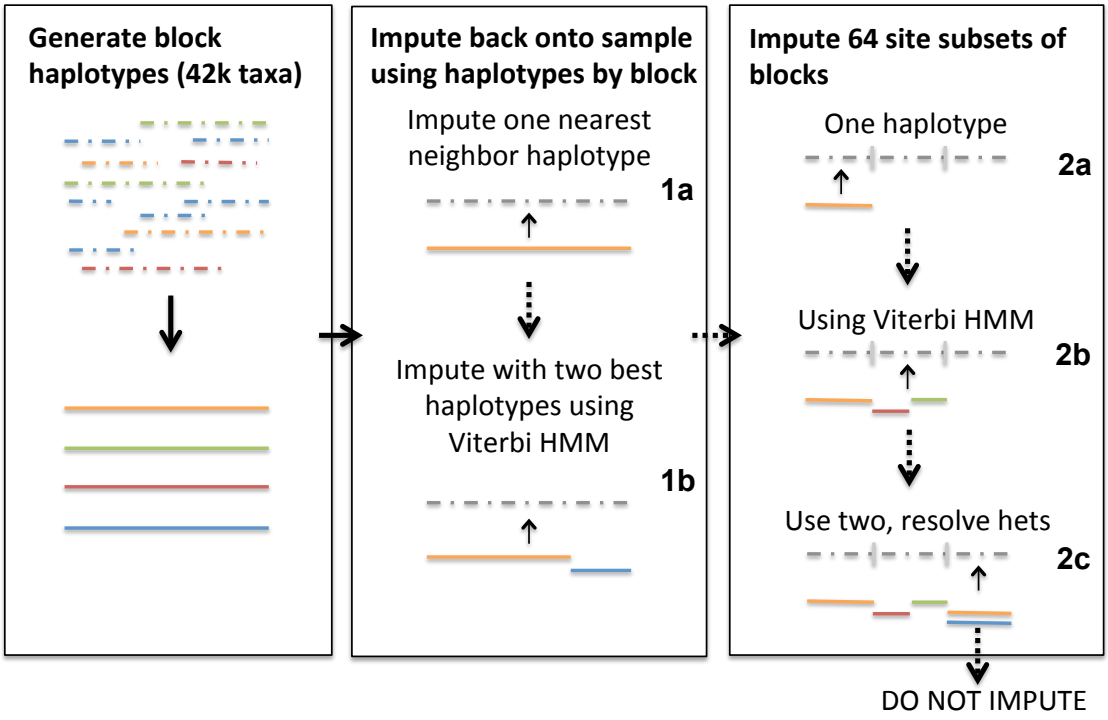


Figure 1. Algorithm overview

Comparison with existing algorithms: FILLIN differs from other available algorithms, most of which have been designed for human-derived sequence data, primarily in its approach to haplotype generation, phasing, and inbreeding assumptions. (*12–14*, *24*) Imputation algorithms either generate haplotypes *de novo* or rely on a densely genotyped reference panel, such as 1000 Genomes in humans, which are not available for most species. The public algorithms that generate *de novo* haplotypes implicitly assume that the unimputed individuals have significant heterozygosity and must be phased. (*12*, *24*) In the case of Beagle v.4, this increases runtime exponentially by the number of samples (Figure 8). However, if haplotypes really do only exist in the heterozygous state, Beagle's refinedIBD (*24*) algorithm should find these segments better than FILLIN and thus make the extra computation

worthwhile. In contrast, FILLIN allows for significant inbreeding in the target population and saves computation by first checking for high similarity between one of the haplotypes and the target sequence. FILLIN and Beagle v.4 both differ from many other algorithms (*13*, *14*, *31*),in that they can impute whole genomes or chromosomes in one run.

Test datasets and analysis optimizations

We tested the FILLIN and FSFHap algorithms against Beagle v.4 (*24*), which we found in preliminary tests to be the most comparable available algorithm: it can generate haplotypes, tolerate high levels of missing data, and impute entire chromosomes with tens of thousands of markers in one run. We also compare against a naive imputation method, which imputes missing genotypes based solely on allele frequencies in the unimputed data. We compare results from three distinct maize datasets genotyped using maize GBS build version 2.7 (*32*): 1) 429 replicate samples representing 287 temperate inbred Ex-PVP and Iowa breeding lines (Temperate Inbreds) (*33*), 2) 467 replicate samples from a panel of well-studied 282 diverse inbred lines from around the globe (*34*) (Diverse Inbreds), and 3) 366 outbred (highly heterozygous) landraces where one half originate from the American Southwest, one quarter from the rest of the Americas (*35*), and one quarter from Spain (*36*) (Diverse Landraces).

A fourth dataset, a recombinant inbred line (RIL) population of full-sib families from the maize Nested Association Mapping (NAM) panel, (*19*) is used to compared FSFHap to FILLIN and Beagle v.4. Each dataset is genome-wide and

filtered so that only polymorphic sites with 10% minimum coverage and taxa with 10% sites present are retained. Because each of the individual NAM full-sib families was derived from three distinct F1 ears and because some parents had residual heterozygosity, any given site might be polymorphic in one sub-family and monomorphic in another. To deal with this after the parental haplotypes were identified for each family, the individual subpopulations were checked to make sure each site was more likely to be segregating 1:1 than to be monomorphic. Any site determined to be monomorphic in a subpopulation was set to missing within that subpopulation. At the same time each site was checked to make sure it was in LD with all its neighbors within a 30-site window. Because of occasional contamination by foreign pollen during the inbreeding process, a few individual RILs carry substantial amounts of non-parental DNA. To find individuals containing significant amounts of non-parental DNA, after an initial imputation, individuals that were more than 30% heterozygous were removed from the data set and the families re-imputed from the original data.

Haplotype generation: FILLIN generates haplotypes using a GBS-derived dataset of 40,992 samples, one-eighth of which are outbred landrace accessions, and the rest are made up of diverse inbred maize lines, inbred teosinte, and biparental mapping and breeding populations. Beagle v.4 could not generate haplotypes from such a large dataset, and for all Beagle runs, Beagle generates haplotypes internally from the target samples. To make a more fair comparison, Beagle is given all replicate samples for lines in the Temperate Inbred and Diverse Inbred panels.

Masking and calculating accuracy

To calculate accuracy, we masked a subset of known genotypes with high read-depth (exactly 7 reads per site) and with a physical position divisible by 7. If GBS can be expected to sample either diploid chromosome equally, the probability that a heterozygous genotype with a read depth of seven is called as a homozygote for either the major or minor allele is Pr (AA|Het)+Pr (BB|Het) = 0.5^7+0.5^7 = 0.0157. Additionally, only heterozygote calls supported by at least two reads for both alleles were masked to exclude calls based on potential sequencing errors.

We calculate accuracy between known masked sites and imputed sites using the coefficient of determination, R^2, as an overall statistic for accuracy. The minor allele is coded as 0, the major allele as 1, and heterozygotes as .5. This implies that a major or minor allele imputed to a heterozygote is half correct, and vice versa for known heterozygotes. We chose this statistic because, by definition, the major allele is the most commonly masked allele in a dataset and the easiest to impute, and also the least useful in downstream applications that associate genotype to phenotype. By using R^2 to calculate accuracy, we reduce the major allele bias present if one takes only the absolute accuracy.

Computational time and algorithm parameters

FILLIN and Beagle v.4 were run on two 6-core Intel Xeon E5 2620 with 2 GHz CPU, 4TB SATA HD, 1TB SSD HD, and 128GB RAM. Beagle v.4 was run using the default parameters, with no external pedigree information or reference panel. FILLIN was run with a window size of 8,000 sites. Haplotypes were required to have

a minimum site presence of 0.6, and the maximum genetic divergence between samples to generate haplotypes was set to 0.01. For imputing haplotypes to the target sequences, 20 informative minor sites were required within a search window and up to 20 haplotype donor hypotheses were explored for a given window. The maximum genetic distance between the haplotype donor and target taxon to impute one haplotype for the entire sites window (Figure1-1a) was set to 0.01, and the maximum distance to impute two haplotypes was set to 0.003 (Figure1-1b). To impute donors to the smaller focus windows (64-site focus, but extended so that the focus window covers 20 informative sites) when the whole-block imputation thresholds were breached, the settings for imputing two haplotypes to inbred lines (with heterozygosity below .02) with Viterbi (Figure1-2b), one haplotype (Figure1-2a), or the combined hybrid haplotype (Figure1-2c) were set to .001, .003, and .01. For heterozygous genotypes, these thresholds were set to 0, .001, and .01. Genotypes were not imputed if these thresholds were not met.

Results

The NAM RILs consist of 25 biparental families with around 200 F6 progeny each. (Table 3) They have an average of .3X coverage per site, are polymorphic in at least one family at 556,000 sites across the genome, and are highly inbred. Average heterozygosity per line is about 0.001, which is lower than reality due to heterozygote undercalling. While for each family we expect minor allele frequencies of .5, across the whole population minor alleles frequencies are very low (Figure 2). To test FILLIN versus Beagle we test three maize datasets differing in degree of inbreeding

and haplotype diversity. All have approximately .5X coverage per site and range in average heterozygosity from 0.001-0.029 (Table 3). The number of polymorphic sites across the genome ranges from 433,000-600,000. As the datasets become more diverse, there is an increased skew towards rare alleles (Figure 2).

Table 3. Raw datasets used for analysis

Dataset	N Filtered Taxa	N Filtered Sites	Avg Prop Present	Avg Prop Heterozygous (± std error)
Temperate Inbreds	429	443036	0.431	.003 ± .004
Diverse Inbreds	467	545154	0.462	.003 ± .005
Diverse Landraces	366	600724	0.509	.052 ± .017
NAM RILs	4776	556001	0.301	.003 ± .002

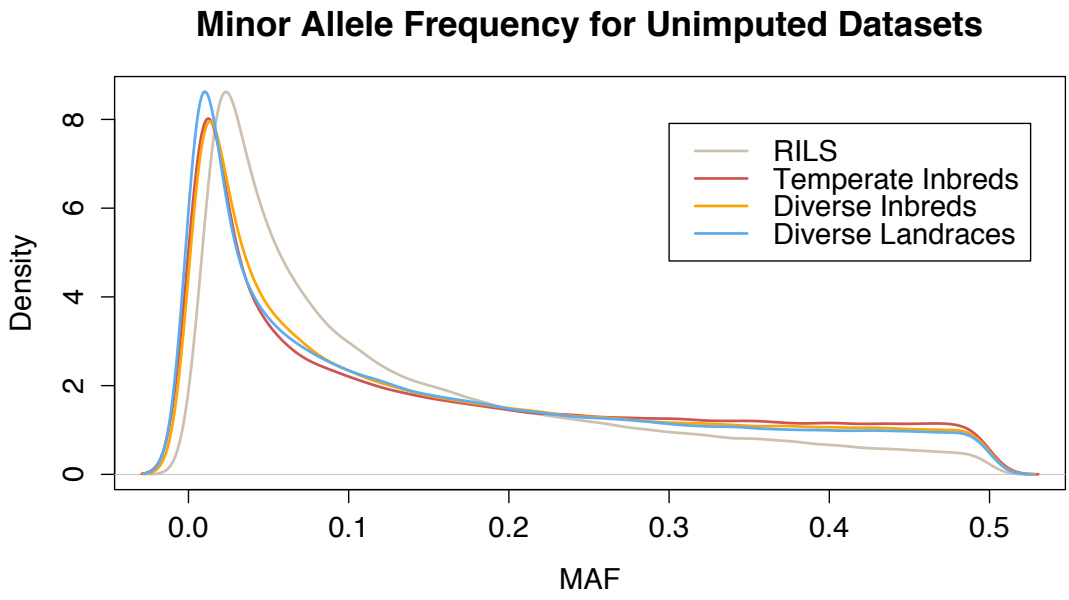


Figure 2. Minor allele frequency densities for the unimputed datasets used for this study.

Because we only masked a subset of genotypes with a read depth of seven, we can compare the distribution of read depths at sites where seven-read depth sites are masked versus sites where they are not. Figure 3 shows that the sites that contain masked genotypes have the same read depth distribution relative to sites without masked genotypes. For all of the datasets tested, we chose to quantify accuracy using the coefficient of determination, R^2, versus a more simplistic measure, such as total percent accuracy where known genotypes are coded as categorical variables. We did this because the great majority of genotypes masked are of the major allele, skewing the accuracy calculation towards imputation accuracies for this genotypic class (Table 4). Because minor alleles are actually of most interest for imputing correctly in downstream applications and the harder allele to predict, we chose to use R^2 to better represent the capabilities of the different methods.

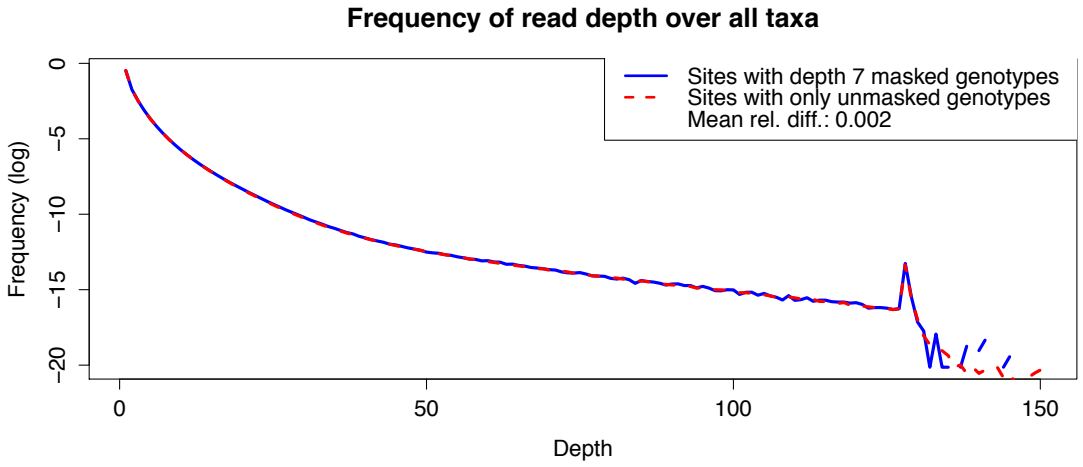


Figure 3. Sites that contain masked genotypes show the same read depth distribution as sites without masked genotypes.

Table 4. R^2, absolute proportion correct, and accuracies by known genotype. Note that the absolute proportion correct mirrors the accuracy for the major allele.

		R^2	Absolute	Minor	Het	Major
Temperate Inbreds						
	Naive	0.046	0.642	0.116	0.348	0.748
	Beagle v.4	0.942	0.984	0.956	0.452	0.993
	FILLIN	0.991	0.997	0.996	0.149	0.999
Diverse Inbreds						
	Naive	0.04	0.641	0.105	0.334	0.75
	Beagle v.4	0.883	0.97	0.905	0.484	0.986
	FILLIN	0.992	0.996	0.995	0.228	0.999
Diverse Landraces						
	Naive	0.064	0.643	0.116	0.358	0.762
	Beagle v.4	0.662	0.892	0.698	0.656	0.957
	FILLIN	0.522	0.827	0.518	0.674	0.889
NAM RILs						
	FSFHap	0.974	0.99	0.968	0.846	0.995
	FILLIN	0.988	0.996	0.988	0.675	0.999
	Beagle v.4	0.987	0.996	0.988	0.68	0.998

For the full-sib NAM RILs, FSFHap and FILLIN performed very similarly, but FSFHap imputed more heterozygous sites. Both algorithms outperformed Beagle v.4 (Figure 4). All algorithms performed far better than the naive allele frequency imputation for all datasets tested with FILLIN and Beagle v.4 (Figure 5). FILLIN outperformed Beagle v.4 for closely related and diverse inbred lines (Figure 5-A,B), but Beagle v.4 outperforms FILLIN for heterozygous landraces (Figure 5C), as well as the few residual heterozygous sites in inbred lines (Figure 6).

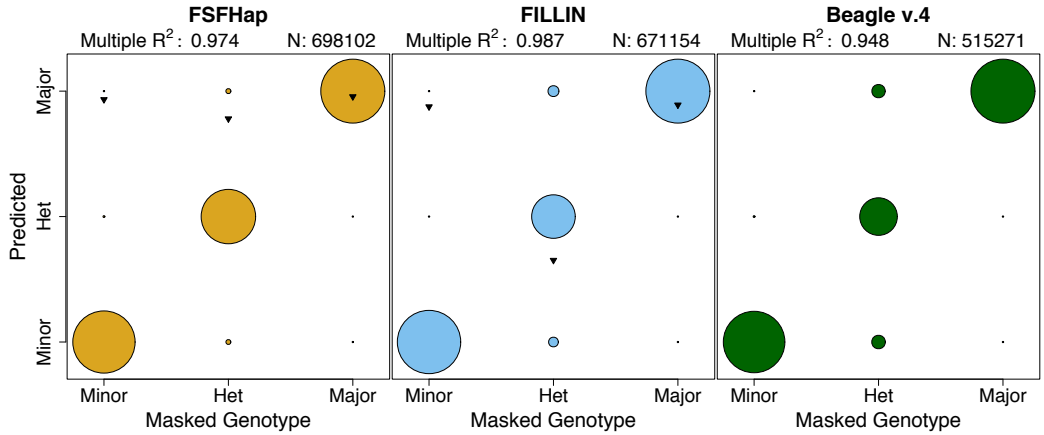


Figure 4. Accuracy comparison between FSFHap, FILLIN, and Beagle v.4 for full-sib NAM RILs. The diameter of each circle represents the proportion within each known genotype class. Triangles mark proportion imputed by each known class; Beagle imputes 100% of missing genotypes

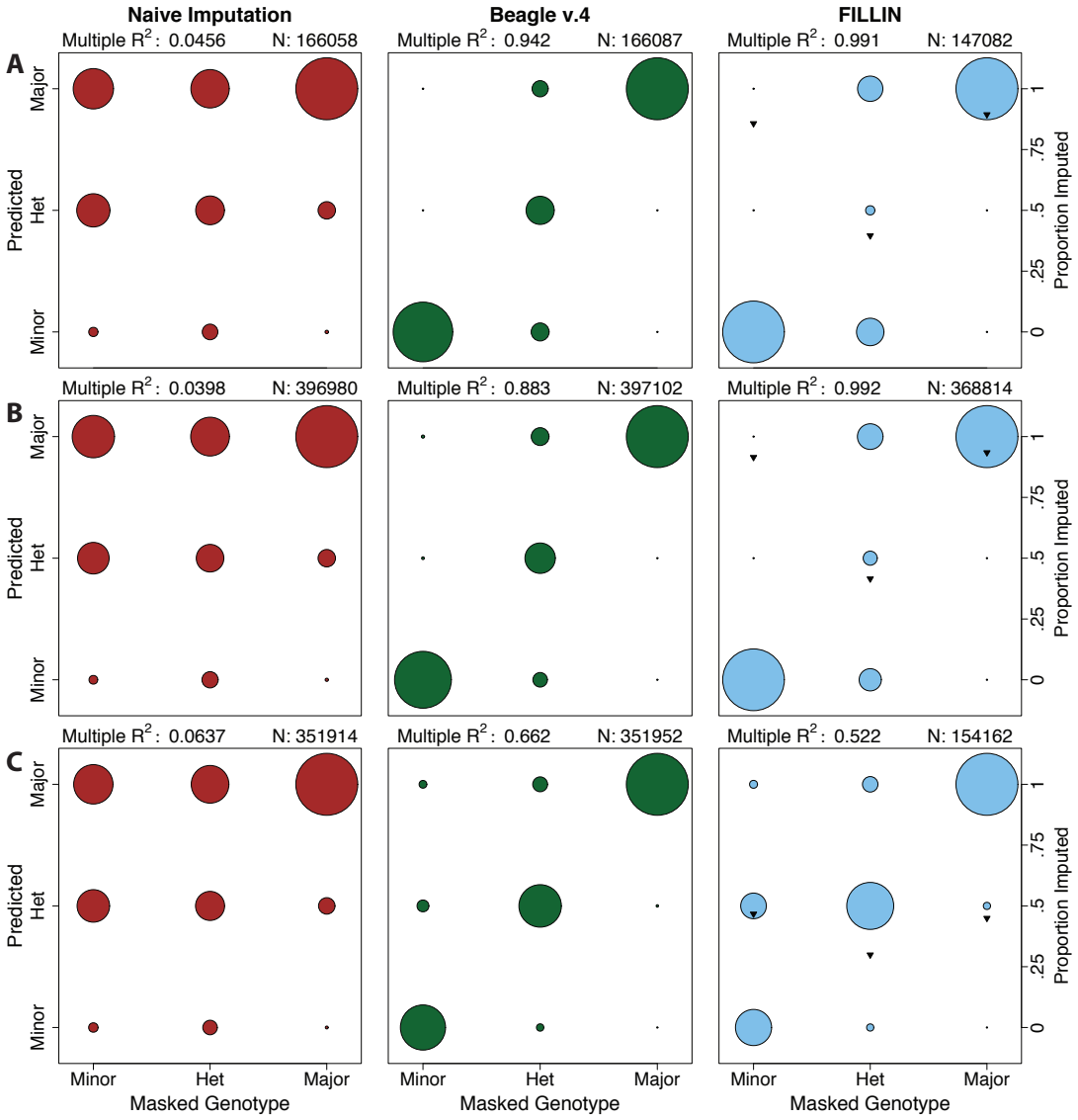


Figure 5. Accuracy for temperate inbreds. The diameter of each circle represents the proportion within each known genotype class. Triangles mark proportion imputed by each known class; Beagle and the naïve allele frequency imputation impute 100% of missing genotypes.

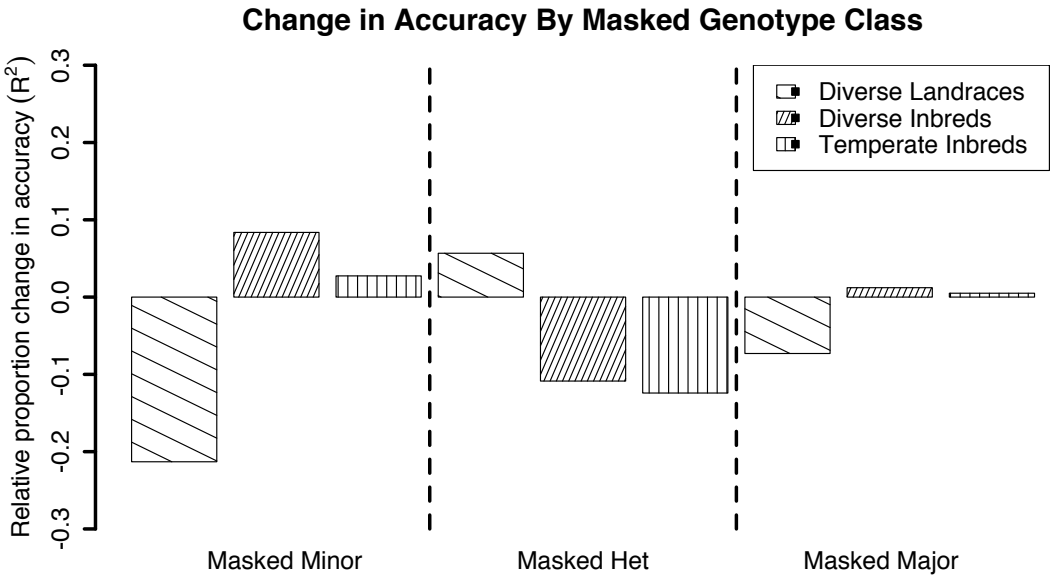


Figure 6. Accuracy for diverse inbreds. The diameter of each circle represents the proportion within each known genotype class. Triangles mark proportion imputed by each known class; Beagle and the naïve allele frequency imputation impute 100% of missing genotypes.

For all of the inbred datasets (diverse, temperate, and RILs) FILLIN most often imputed the whole site window with one haplotype or two, using the Viterbi algorithm to model recombination breakpoints (Table 5). The temperate inbreds, although they are more closely related, use the focus block imputation more often than the diverse inbreds and this may explain their slightly decreased accuracy and suggests more residual heterozygosity in these lines. The landraces almost never impute using the whole site window, which is expected given their high heterozygosity and increased historical recombination. The landraces also use the two combination haplotype mode more often, and set more focus blocks to missing, reflecting a lack of accurate haplotypes.

Table 5. Method of imputation for windows averaged across taxa (± one standard deviation) and proportion of each method for those site windows that go into focus block imputation. Number of windows (N) change per dataset based on number of segregating sites.

	Mode for window		Proportion of focus block			
	Site Window	Focus Block	Inbred	Viterbi	Combo	Missing
Temperate Inbreds (N=55)	38.85 ± 13.97	16.14 ± 13.96	0.48 ± 0.17	0.09 ± 0.07	0.4 ± 0.15	0.03 ± 0.04
Diverse Inbreds (N= 75)	61.56 ± 12.77	6.44 ± 12.77	0.34 ± 0.24	0.14 ± 0.13	0.49 ± 0.26	0.03 ± 0.04
NAM RILs (N= 70)	59.7 ± 5.66	10.18 ± 5.55	0.36 ± 0.12	0.12 ± 0.08	0.41 ± 0.14	0.11 ± 0.06
Diverse Landraces (N= 68)	10.52 ± 22.92	63.17 ± 25.52	0.28 ± 0.09	0.02 ± 0.06	0.43 ± 0.14	0.28 ± 0.13

The stringency settings chosen for FILLIN, which are also the defaults for the algorithm, were decided empirically based on these data, and were optimized for accuracy in inbred and breeding populations. Changing these thresholds leads to an increased number of genotypes imputed, but at the cost of accuracy. For the landrace populations especially, loosening the requirements rapidly leads to decreased accuracies while never imputing more than 60% of the minor alleles.

The gain in accuracy for FILLIN derived from more accurate imputation of minor alleles (Figure 6). Figure 7 suggests that the increase in accuracy for minor alleles derives from FILLIN's insensitivity to the minor allele frequency (MAF). Gain in accuracy from accurate imputation of minor alleles is especially true for inbred lines (Figure 6) and suggests that MAF insensitivity results from imputing one haplotype onto the inbred regions of the target taxon. For Beagle v. 4 and FILLIN in heterozygous populations, imputation accuracy is otherwise a function of the MAF, where lower frequency variants are imputed less accurately (Figure 7).

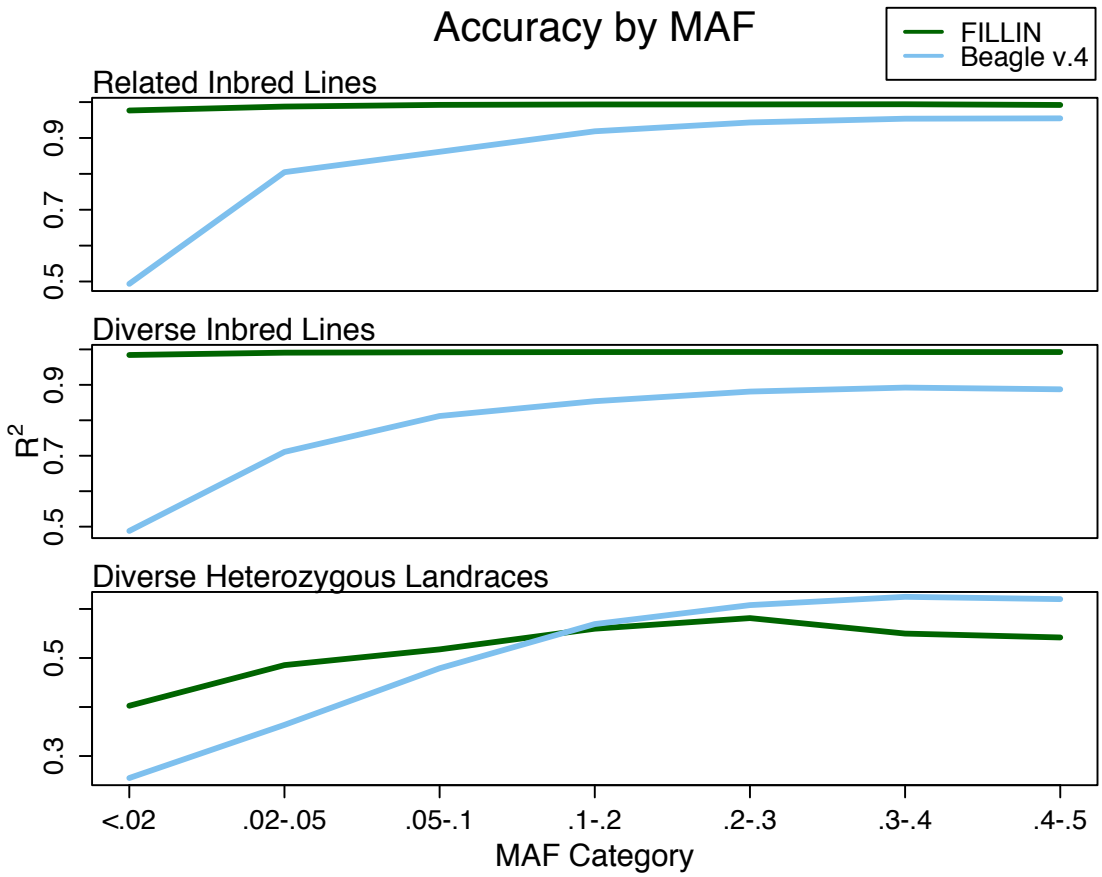


Figure 7. Accuracy for diverse landraces. The diameter of each circle represents the proportion within each known genotype class. Triangles mark proportion imputed by each known class; Beagle and the naïve allele frequency imputation impute 100% of missing genotypes.

These tests suggest that Beagle's advantage in heterozygous populations lies in their haplotype generation and phasing RefinedIBD (*24*) algorithm. FILLIN only draws haplotypes from inbred segments, implicitly assuming that the haplotypes present in heterozygous lines are present as inbred regions in the dataset. This is not necessarily true, either because the inbred line or segment containing that haplotype may not have been sampled, or because it may not exist in a homozygous state

because it contains a fatal deleterious allele. To test whether FILLIN was actually generating haplotypes from inbred lines, we trained haplotypes on only the data submitted for imputation and found that, when given no external information, FILLIN could not generate any haplotypes from the landraces until 3000 samples were input and with these resulting R^2 accuracies are less than 0.1 (Figure 8). FILLIN achieves accuracies of around 0.5 when given information from inbred lines and segments in the 40,000 sample dataset (Figure 6). This suggests that only half of the haplotypes present in the landraces are present as inbred segments in GBS genotyped maize samples, and highlights FILLIN's inability to phase heterozygotes. However, we found that FILLIN performed very well for inbred lines and biparental families, albeit not as well as when provided additional information. For inbred lines and families, accuracy as well as proportion of minor alleles imputed increased with sample size (Figure 8).

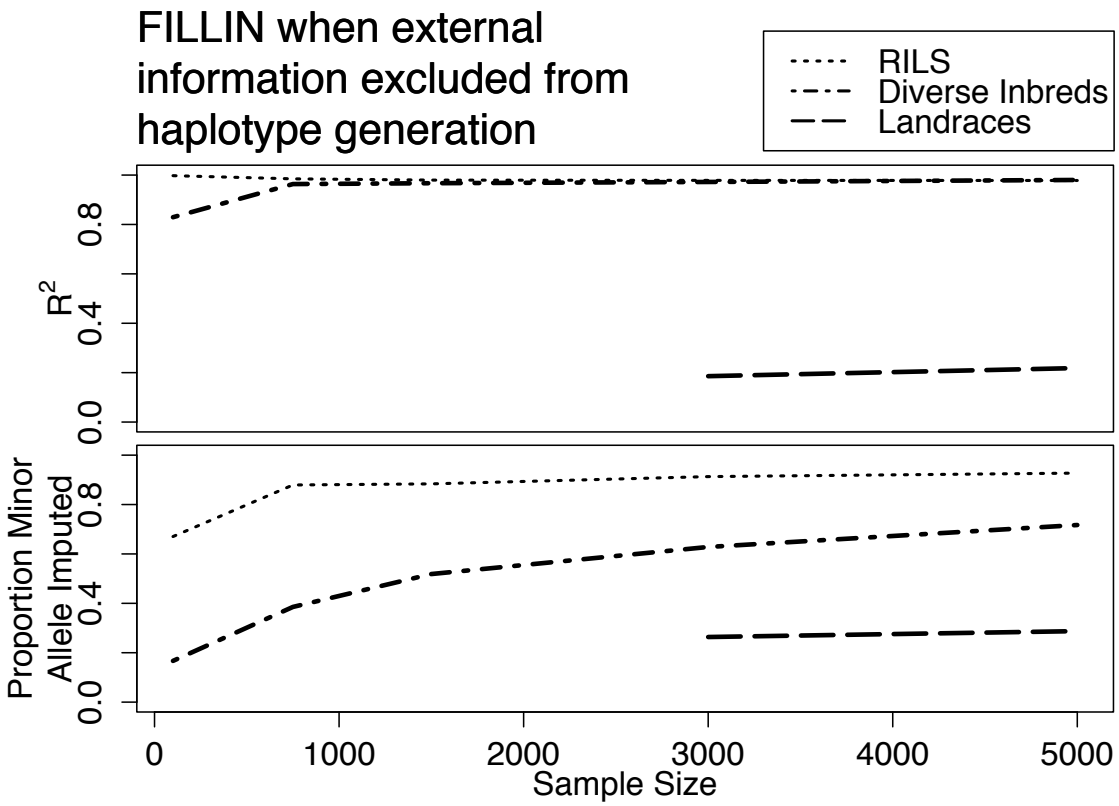


Figure 8. FILLIN at different sample sizes for NAM RILs, diverse inbred lines, and heterozygous, diverse landraces using no external data to generate haplotypes. Each data set is a random sub-sample of the larger data set.

Another feature of FILLIN is that it requires at least two taxa in a given block to share a haplotype to generate a donor. This is done to increase haplotype coverage, and to increase the robustness of the donor haplotypes, but means that diverse taxon at low coverage may not be represented in the donor file well (5.68% of masked, unimputed polymorphic sites are monomorphic in the donor file for the diverse landraces). Figure 8 shows that the more diverse the dataset, the more samples are required to adequately generate haplotypes.

For all tested datasets, imputation accuracy improved when only the consensus genotypes from both imputation methods were accepted (Table 6). A consensus

approach gains from the strengths in each imputation method: FILLIN's sensitivity to inbred segments and Beagle's to highly heterozygous regions. While it is very difficult to accurately identify and mask structurally missing variation in GBS data, the lower R^2 for Beagle at sites that FILLIN chooses not to impute suggest that FILLIN does provide sensitivity to structural variation by allowing for residual missing data in the haplotypes. The consensus approach also reduces the potential for overimputation, which is important in species with high structural variation such as crop plants. (*4*)

Table 6. Accuracy (R^2) for both imputations separately, the consensus imputation, and the accuracy for Beagle when the inbred imputation chooses not to impute.

	R2		
	Diverse Landraces	Diverse Inbreds	Temperate Inbreds
FILLIN	0.522 (154162)	0.992 (368814)	0.991 (147082)
Beagle	0.737 (352033)	0.891 (397163)	0.949 (166097)
Both agree	0.772 (128028)	0.996 (359905)	0.996 (145599)
Beagle, when inbred does not impute	0.607 (199572)	0.776 (28391)	0.845 (19049)

Tests with different sized datasets suggested that the gain in computational time by FILLIN relative to Beagle increased with sample number: where FILLIN scales linearly with sample size, while Beagle runtime increases exponentially (Figure 9). Here again, Beagle performs better on heterozygous taxa than inbreds, and vice versa for FILLIN, as shown by the change in rank between heterozygous and inbred datasets by method.

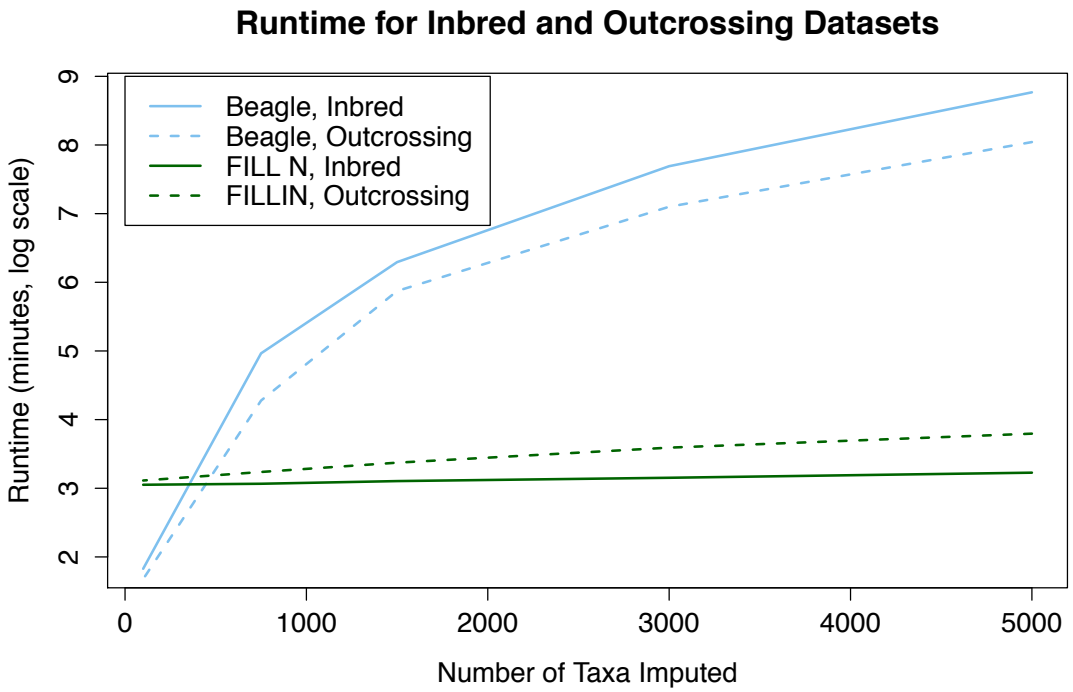


Figure 9. Computational time for FILLIN and Beagle v.4. Only one chromosome (chromosome 10) was compared, since computing the whole genome with >1000 samples using Beagle v.4 is intractable unless parallelized. Each subset is a random sample of the larger taxa set.

Discussion

FILLIN and FSFHap produce highly accurate imputed genotypes, especially for closely related populations with replicated inbred segments. Cross contamination is difficult to completely exclude in even controlled crosses and a number of maize GBS genotyped lines contain errors in pedigree. Because FILLIN and FSFHap do not require known parental genotypes, these algorithms provide a pedigree independent imputation method. If imputing full-sib families, FSFHap is optimized for modeling recombination, which allows it to more accurately impute heterozygotes. Beagle v.4 provides more accurate imputation for highly heterozygous populations.

Accurate and complete haplotype generation is critical to high accuracies in both Beagle v.4 and FILLIN imputed datasets. Results in Figure 7 suggest that the relative gain in accuracies for the two algorithms for different types of datasets directly reflects the strengths of the two algorithms in haplotype generation. It is impossible in both algorithms to impute a minor allele correctly if that variant does not exist in the haplotypes. Beagle is better able to phase and extract haplotypes from heterozygous taxa, and consequently imputes heterozygous datasets better for both heterozygous genotypes and genotypes homozygous for the minor allele. FILLIN focuses on extracting phased haplotypes from inbred segments, and imputes minor alleles better for homozygous datasets.

Overall, FILLIN provides rapid imputation of large N, low-coverage, whole-genome sequence data from predominantly inbred or breeding populations with high overall accuracy. For full-sib families where the objective is to find recombination breakpoints or to do linkage analysis, which requires IBD information, FSFHap provides sensitive and accurate imputation. For highly heterozygous samples with unknown segregating parental haplotypes, we recommend at this time that researchers use Beagle v.4. (*24*) If highly accurate imputation is required, taking a consensus imputation will provide the most accurate results.

Together, these three algorithms, Beagle v.4, FILLIN and FSFHap, provide robust imputation of low coverage GBS data from diverse populations. High quality haplotypes are required for accurate imputation by any of the algorithms presented here. Thus, if genotyping unrelated inbred lines or heterozygous populations in a species without available haplotype panels, resources should be expended to genotype

a subset of individuals covering the diversity of haplotypes at higher coverage to ensure accurate haplotype generation and subsequent imputation. These results suggest that even one generation of selfing, in species where that is possible, can aid in accurate imputation of low coverage genotyped populations. For breeders or researchers desiring to use GBS for genotyping highly related breeding populations, these results suggest that very low coverage genotyping, combined with FILLIN or FSFHap imputation, will provide highly accurate results at low cost. This is true even if the parents have not been sampled elsewhere, since numerous low coverage replicates of each haplotype are expected within the population. Well thought out experimental design can help keep genotyping costs low (<$20 per sample), which enables efficient breeding and conservation biology decisions to be made.

REFERENCES

1. R. J. Elshire *et al.*, A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*. **6**, e19379 (2011).
2. P. S. Schnable *et al.*, The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science*. **326**, 1112–1115 (2009).
3. Y. Vigouroux *et al.*, Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* **95**, 1240–1253 (2008).
4. J.-M. Chia *et al.*, Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* **44**, 803–807 (2012).
5. B. Das *et al.*, Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. *BMC Genet.* **14**, 71 (2013).
6. B. L. Hurwitz *et al.*, Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus Oryza. *Plant J. Cell Mol. Biol.* **63**, 990–1003 (2010).
7. N. K. Bhullar, M. Mackay, B. Keller, Genetic Diversity of the Pm3 Powdery Mildew Resistance Alleles in Wheat Gene Bank Accessions as Assessed by Molecular Markers. *Diversity*. **2**, 768–786 (2010).
8. P. Ramu *et al.*, Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **126**, 2051–2064 (2013).
9. J. C. Schnable, N. M. Springer, M. Freeling, Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci.* **108**, 4069–4074 (2011).
10. M. A. Cleveland, J. M. Hickey, B. P. Kinghorn, Genotype imputation for the prediction of genomic breeding values in non-genotyped and low-density genotyped individuals. *BMC Proc.* **5**, 1–5 (2011).
11. C. C. A. Spencer, Z. Su, P. Donnelly, J. Marchini, Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLoS Genet*. **5**, e1000477 (2009).
12. B. L. Browning, Z. Yu, Simultaneous Genotype Calling and Haplotype Phasing Improves Genotype Accuracy and Reduces False-Positive Associations for Genome-wide Association Studies. *Am. J. Hum. Genet.* **85**, 847–861 (2009).
13. E. Y. Liu, M. Li, W. Wang, Y. Li, MaCH-admix: genotype imputation for admixed populations. *Genet. Epidemiol.* **37**, 25–37 (2013).
14. B. N. Howie, P. Donnelly, J. Marchini, A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet*. **5**, e1000529 (2009).
15. J. E. Rutkoski, J. Poland, J.-L. Jannink, M. E. Sorrells, Imputation of Unordered Markers and the Impact on Genomic Selection Accuracy. *G3 GenesGenomesGenetics*. **3**, 427–439 (2013).
16. T. Meuwissen, M. Goddard, The Use of Family Relationships and Linkage Disequilibrium to Impute Phase and Missing Genotypes in Up to Whole-Genome Sequence Density Genotypic Data. *Genetics*. **185**, 1441–1449 (2010).

17. J. M. Hickey, B. P. Kinghorn, B. Tier, J. H. van der Werf, M. A. Cleveland, A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genet. Sel. Evol.* **44**, 9 (2012).
18. H. D. Daetwyler, G. R. Wiggans, B. J. Hayes, J. A. Woolliams, M. E. Goddard, Imputation of Missing Genotypes From Sparse to High Density Using Long-Range Phasing. *Genetics*. **189**, 317–327 (2011).
19. M. D. McMullen *et al.*, Genetic Properties of the Maize Nested Association Mapping Population. *Science*. **325**, 737–740 (2009).
20. E. S. Mace, C. H. Hunt, D. R. Jordan, Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* **126**, 1377–1395 (2013).
21. B. Diers, J. Specht, D. Hyten, R. Nelson, B. Beavis, (St. Louis, MO, 2013).
22. P. Andolfatto *et al.*, Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* **21**, 610–617 (2011).
23. E. G. King, S. J. Macdonald, A. D. Long, Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. *Genetics*. **191**, 935–949 (2012).
24. B. L. Browning, S. R. Browning, Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics*. **194**, 459–471 (2013).
25. E. S. Buckler *et al.*, The Genetic Architecture of Maize Flowering Time. *Science*. **325**, 714–718 (2009).
26. J. Ross-Ibarra, M. Tenaillon, B. S. Gaut, Historical Divergence and Gene Flow in the Genus Zea. *Genetics*. **181**, 1399–1413 (2009).
27. P. J. Bradbury *et al.*, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**, 2633–2635 (2007).
28. L. . Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*. **77**, 257–285 (1989).
29. A. Gusev *et al.*, Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* **19**, 318–326 (2009).
30. B. L. Browning, S. R. Browning, A Fast, Powerful Method for Detecting Identity by Descent. *Am. J. Hum. Genet.* **88**, 173–182 (2011).
31. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
32. J. C. Glaubitz *et al.*, TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*. **9**, e90346 (2014).
33. M. C. Romay *et al.*, Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55 (2013).
34. S. Flint-Garcia *et al.*, Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* **44**, 1054–1064 (2005).
35. J. van Heerwaarden *et al.*, Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci.* **108**, 1088–1092 (2011).
36. P. Revilla, P. Soengas, M. E. Cartea, R. A. Malvar, A. Ordas, Isozyme variability among European maize populations and the introduction of maize in Europe. *Maydica*. **48**, 141–152 (2003).

CHAPTER 3

A LARGE SCALE JOINT ANALYSIS OF FLOWERING TIME REVEALS INDEPENDENT TEMPERATE ADAPTATIONS IN MAIZE

Author List

Kelly Swarts (1), Eva Bauer (2), Robert Bukowski (3), Jeff Glaubitz (4), Tiffany Ho (1), Lynn Johnson (5), Chunhui Li (6), Yongxiang Li (6), Yu Li (7), Xiaolei Liu (7), Zachary Miller (5), Cinta Romay (4), Jeffrey Ross-Ibarra (8), Chris-Carolin Schön (2), Tianyu Wang (6), Zhiwu Zhang (9), Peter Bradbury (5), Edward S. Buckler (1,5)

1 Department of Plant Breeding and Genetics, 175 Biotechnology, Cornell University, Ithaca, NY 14853
2 Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, D-85354 Freising, Germany
3 Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, NY, 14853
4 Genomic Diversity Facility, Institute of Biotechnology, Cornell University, Ithaca, NY, 14853
5 USDA-ARS, 538 Tower Road, Ithaca, NY 14853
6 Institute of Crop Science, Chinese Academy of Agricultural Sciences, Zhongguancun South Street, Haidian, Beijing, China, 100081
7 Huazhong (Central China) Agricultural University, Hubei, China
8 Dept. of Plant Sciences, 262 Robbins Hall, Mail Stop 4, University of California, One Shields Ave, Davis, CA 95616
9 Department of Crop and Soil Sciences, Washington State University, 105 Johnson Hall, Pullman, WA 99164, USA

Kelly Swarts wrote the manuscript and led all analyses and projection of Hapmap 3.21 SNPs onto the GBS samples. Peter Bradbury contributed substantially to the development of BRHM, ran the RMIP analysis, projected the EUNAM populations and contributed intellectually to the discussion and edited the manuscript. Robert Bukowski called SNPs in Hapmap 3.21, and performed KNNi imputation. Jeffrey Ross-Ibarra provided the maize landraces from across the Americas. Jeff Glaubitz supplied database information for the samples and some of the projected datasets. Tiffany Ho ran preliminary machine learning analyses and Zachary Miller ran the Random Forest Classifier analyses and Lynn Johnson set up and curated the database for the machine learning predictors. Chunhui Li, Yongxiang Li, Yu Li, Tianyu Wang and Zhiwu Zhang contributed the CNNAM datasets and phenotypes. Eva Bauer and Chris-Carolin Schön contributed the EUNAM datasets. Cinta Romay and Edward Buckler contributed the Ames dataset. Cinta Romay additionally added the ZCN genes

to the Dong et al flowering candidate list information and Edward Buckler contributed the NAM dataset and a substantial intellectual input to the BRHM method, machine learning, interpretations, and manuscript edits.

Abstract

Modulating days to flowering is a key mechanism in plants for adapting to new environments, and variation in days to flowering drives population structure by limiting mating. To elucidate the genetic architecture of flowering across maize, a quantitative trait, we mapped flowering in five global populations, a diversity panel (Ames) and four half-sib mapping designs, Chinese (CNNAM), US (USNAM), and European Dent (EUNAM-Dent) and Flint (EUNAM-Flint), together totaling over ten thousand individuals, using whole-genome projected SNPs to increase the resolution of associations. We tested for joint association using GWAS, resampling GWAS and two regional approaches; Regional Heritability Mapping (RHM) (*1, 2*) and a novel method, Boosted Regional Heritability Mapping (BRHM). Direct overlap in significant regions detected between populations and previously published flowering candidate gene lists was limited, but whole-genome cross-population prediction accuracies were up to 0.78. Poor prediction accuracy was correlated with increased population differentiation ($r = 0.41$), unless the parents were broadly sampled from across the North American temperate-tropical germplasm gradient; uncorrected GWAS results from populations with broadly sampled parents were well predicted by temperate-tropical population differentiation in machine learning. Machine learning of the diversity panel GWAS results also suggested shared architecture between the American panels and, more distantly, the European panels, but not the Chinese panel. Machine learning approaches provide a way to reconcile non-linear relationships

between populations and other predictors, but the combined predictive ability of all of the populations did not significantly enhance prediction of candidate genes. While the North American-European temperate adaption is well studied, this study suggest substantially independent temperate adaptation evolved in the Chinese panel, most likely in China after 1500, also supported by differential gene ontology term enrichment for top genes in each population.

Background

This study sought to refine our understanding of the genetic architecture underlying maize flowering by combining diverse populations genotyped using whole genome resequencing data. One of the easiest ways for a plant to adapt to a new environment is to modulate the time until flowering, avoiding obstacles to reproductive success. However, changes in flowering lead to reproductive isolation, structuring populations and generating strong pleiotropic associations with other traits (*3*, *4*). This makes days to flowering an excellent trait to study the biological mechanism of adaptation, but confounds association mapping approaches.

Flowering traits and population structure

Flowering time traits have received much attention in maize due to the quantitative nature of inheritance, high heritabilities (up to 0.96 for days to silking in the USNAM (*5*) population) and easy scoring (*5–18*). Only a handful of loci, mostly in the autonomous flowering pathway, have confirmed effects for modulating days to flowering either through mutagenesis or positional cloning, including Rap2.7 and the

MITE insertion at the Vgt1 locus (*7, 19*), ZCN8 (*20*), ID1 (*21, 22*), conz1 (*23*), df1 (*24*), zfl1 and zfl2 (*25*) and the major photoperiod locus, ZmCCT (*10*). However, independent days to flowering QTL have been mapped in tens of biparental, RIL, and other structured population designs (*8, 12*), including four NAM populations (*5, 26, 27*) and in diversity populations (the largest being the Ames Diversity Panel (*11*)); and of the list above, only Vgt1, ZCN8, and ZmCCT have common standing variation and are routinely detected in mapping populations (*5, 8, 13*). Flowering time is also of interest for breeding, as shifting days to flowering is one of the easiest ways to combat the effects of climate change (*28–31*), and is important for moving traits such as disease resistance or yield between tropical and temperate germplasm.

Previous studies (*5*, *8*, *12*, *15*, *26*, *32*, *33*) have confirmed that days to flowering is quantitative and highly additive in maize, and that many of the effects target common loci, with allelic series present at many loci; a meta-analysis of 22 linkage mapping and ANOVA studies record a total of 313 QTL that collapsed in 62 consensus regions of the genome (*12*). Synteny mapping to rice and Arabidopsis resulted in 19 overlapping associations (*12*), and synteny mapping with sorghum revealed that 92.5% of QTL found in sorghum were less than 10Mbp from a corresponding QTL in USNAM (*33*). A later meta-analysis including 29 studies (*8*), found 441 significant QTL could be collapsed into 59 genomic regions. Similarly, days to flowering in the first NAM population, USNAM, found over 50 genomic regions with significance, and additionally confirmed the existence of allelic series at common loci. A recent, large multi-environment evaluation of maize flowering in the USNAM, Ames diversity panel and a newly developed Chinese NAM (CNNAM)

(*27*) population using low-density GBS markers identified 130 QTL from linkage mapping in USNAM and CNNAM, of which 40 overlapped between the two populations (*32*). Cross heterotic pool predictive accuracy in the two EUNAM panels was typically close to zero or even negative in most families (*26*). This suggests that genetic control for flowering is at least partially differentiated across populations.

The quantitative basis for days to flowering in maize results from the long and complex history of adaptation to new environments. Adaptation to climate was under selection pre-domestication, as the wild progenitor teosinte contended with restricted range during the last glacial period before subsequent expansion as temperatures started to rise 12,000 years ago (*34, 35*). Altitudinal variation also preadapted maize to range expansion; the direct progenitor, *Z. mays* ssp. *parviglumis*, ranges from 400-1,500m elevation in central Mexico, and the closely related upland teosinte, *Z. mays* ssp. *mexicana* specializes in the higher and drier uplands, 1,600-2,700m (*36*), and has contributed substantial variation to domesticated maize(*37, 38*).

Domestication of maize took place early during the mid-Holocene maximum (9,000-5,000 BP) (*39, 40*), where annual temperatures were higher than any time before the industrial era (*35*). After domestication, humans moved maize across the Americas, reaching the southern Andes by at least 3,600 BP (*41*) and moving northward to southern Ontario, Canada by 1,500 BP (*42*). After Spanish contact with the Americas in the late 1,400s, maize was quickly established across the world (*43–47*). Modern breeding in the last 100 years has further complicated these histories with the development of inbred lines, heterotic groups, and the intentional introgression of exotic germplasm into global maize, particularly the historical US Dent germplasm

(*48*). The resulting global modern maize germplasm is highly diverse, and days to flowering in inbred lines varies from 35-120 days (*49*).

Meta-analysis and population structure in human and plant literature

Meta-analysis of GWAS results, or combining p-values from independent studies to increase the significance of common, marginally significant loci, has been routinely applied in human studies since the 1980s, namely because of the problems associated with direct access to human disease data (*50*). Numerous methods exist for combining results (*50*), both Bayesian and Frequentist, but all balance generalizability against power to control for between dataset heterogeneity with respect to study size, phenotypic measurements, or population structure (e.g., MANTRA (*51*)). Human studies tend to be underpowered compared with crop plants due to the inability to replicate phenotypes, and meta-analysis approaches are very effective at increasing power to detect variants in human studies (*1, 52, 53*). One of the reasons for these successes is the reasonably high resolution in human studies (*54*), due to the outcrossing nature of human populations and researchers ability to capitalize on this advantage due to the widespread availability of (imputed) high-density markers (*55*). Additionally, the recent and exponential expansion of human populations has generated an overabundance of rare alleles (*56*), which mapping studies have low power to detect alone, but can be boosted to significance in aggregate (*50*).

Historically in crop plants, researchers have focused more on biparental (*14, 57, 58*) or more recently Nested Association Mapping (NAM) populations (*59*). This is because, unlike in humans, inbreeding is tolerated and ethics allow for structured

mating designs and highly replicated phenotypic estimates. Structured populations require only sparse markers because recombination events are limited, decreasing genotyping costs but also limiting resolution. In return for limited resolution, structured populations control for historical population structure, allowing for accurate effect estimates of even rare variants, which can be a source of novel breeding variation. NAM designs link multiple biparental populations with a common parent, allowing for the estimation of allelic effect estimates across the parents. Most meta-analyses to date in structured populations combine linkage mapping studies rather than GWAS (*8, 12*). In this analysis, we combined the high resolution of diversity panels, whole-genome SNP coverage, and decreased effects of population structure and accurate effect estimates of structured designs to further resolve the genetic basis for flowering.

The motivating purpose of this study is whether combining populations at whole-genome coverage can enable us to detect more associations, with more resolution, in a complex trait like days to flowering. We reanalyzed five mapping populations for association with flowering traits, days to silking and days to anthesis, using both single marker tests and regional approaches. We introduce a novel method, Boosted Regional Heritability Mapping or BRHM, for regional mapping that controls for population structure and extended linkage disequlibrium in mapping populations and that easily integrates results from populations. Because days to flowering is confounded with population structure and adaptation to new environments, we also tested these populations for cross-validation based on population structure captured in similarity matrices, and used a Random Forest Classifier machine learning framework

to understand the basis for predictions. We test for differences in the mechanisms for adaptation in different environments by comparing GO term enrichment across populations, and testing mapping results against genome annotations and F_{ST} estimates in a machine learning framework.

Methods

Datasets

Five publically available maize datasets were reanalyzed for this analysis, four Nested Association Mapping (NAM) designs (US (*59*), Chinese (*27*), and two European panels based on the major European heterotic groups, Flint and Dent (*26*)) and one diversity panel (Ames) (*11*) (Table 1). Phenotypes used were field-corrected BLUPs from the original studies. All of the populations were genotyped at low density in their original study, which we used to anchor whole genome projection for all individuals. The EUNAM panels were genotyped using an Illumina MaizeSNP50 BeadChip (*60*), and the other panels were genotyped with Genotyping-By-Sequencing (GBS) (*61*). Whole genome genotypes from maize Hapmap3.2.1 (*62*) were imputed using K-Nearest Neighbor imputation (KNNi) (*63*) with an overall accuracy of 0.988 and a minor allele imputation accuracy of 0.94 for imputed genotypes, then haplotypes projected onto all populations. Hapmap 3.21(*62*) was called on 1,268 inbred genotypes from across the world, with highly variable depth of coverage, and paralogous sites were retained, as they provide signal in GWAS. Because paralogous sites were retained, we used KNNi (*63*) to impute, which was robust to high error rates in genotype calling, but KNNi over-imputes missing data to the major allele.

GBS-genotyped populations were projected using FILLIN (*64*), and EUNAM was projected using a custom implementation of FSFHap (*64*). KNNi, FILLIN, and FSFHap all use implementations in TASSEL (*65*). For GBS populations, projection was anchored by 465,085 consensus sites between Hapmap3 and GBS, where the physical positions match and the major/minor alleles are shared – projection accuracy was r = 0.99 overall between masked and subsequently imputed genotypes (0.96 for minor alleles). For EUNAM, the parental haplotype breakpoints were imputed for each of the progeny using FSFHap. TASSEL used those breakpoints to project the Hapmap 3.21 genotypes of the parents onto the progeny. While most of the parents of these NAM populations were completely inbred, there were a minority that had residual heterozygosity, which can produce families with three haplotypes segregating. Projected datasets were then filtered using appropriate parameters for the family structure of each population, to ensure a minimum of at least 10 minor alleles for any given site in a population; NAM populations were filtered so that one family must have a minimum minor allele frequency of 0.1 (which controls for the parental residual heterozygosity), giving a minimum minor allele frequency of 0.02 for the population as a whole, and Ames was filtered for minimum MAF of 0.015. Any sites with a maximum heterozygosity above 0.02 or coverage below 0.3 were removed. Before calculating kinships, any residual missing genotypes were assigned a homozygous genotype randomly drawn from the genotypic frequency distribution at each site, by family if appropriate.

Table 1. Population statistics for KNN imputed-projected and filtered genotypes

Population	Pop Type (N families)	N Filt KNNi Sites	Ntaxa DTA	Ntaxa DTS
USNAM	NAM (25)	39,302,149	4,758	4,758
Ames	Diversity	34,967,601	2,312	2,279
CNNAM	NAM (11)	27,047,718	1,928	1,928
EUNAM-Dent	NAM (10)	23,766,799	841	841
EUNAM-Flint	NAM (11)	25,388,592	811	811

Multidimensional Scaling (MDS) of American landraces and NAM parents

It is often unclear how inbred lines fit within an adaptive evolutionary context, since breeding programs cross and select on progeny from unrelated individuals. We performed joint MDS analysis with American landraces from Takuno et al (*66*) spanning the two American temperate-tropical gradients with the parents of the four NAM populations to understand the distributions of the parents across American temperate adaptation. Each landrace individual was included ten times in the IBS distance matrix used to calculate the MDS coordinates so that the first two coordinates reflect the relatedness within the landrace populations. Landraces and parents for CNNAM were natively genotyped using GBS, and EUNAM and USNAM projected in Hapmap3.21 coordinates. MDS was based on 465,085 consensus sites between Hapmap3 and GBS (cmdscale () in R, using an IBS distance matrix generated in TASSEL)

Cross population prediction

Cross population prediction were performed to better understand how populations were related to each other with respect to shared genetic architecture for days to flowering. Cross population prediction was performed using RR-BLUP as

implemented in TASSEL (*65*). The training and test populations were combined in a single kinship (similarity) matrix, calculated using Centered-IBS (after Van Raden (*67*), Endelman and Jannink (*68*)), and phenotypes for the test population masked so that the model was trained solely from the training population phenotypes. Predictions from the resulting model for the test phenotypes were then correlated with the true phenotypes (the "prediction accuracy") in R using the Pearson method in cor.test() in the stats package. Genomic subsets were calculated following Rodgers-Melnick et al (*69*).

Genome-wide Association Study (GWAS)

A GLM was conducted in TASSEL on two sets of phenotypes. One was the unmodified phenotypes, uncontrolled for population structure, because flowering is not only correlated with population structure, but it acts to differentiate populations. Thus, many of the regions that differentiate populations may also control regions important for temperate adaptation and we captured these in the uncorrected model. We also controlled for population structure by fitting 5 MDS coordinates calculated from an IBS distance matrix in TASSEL for the Ames panel, or fitting a family term for the structured populations in a TASSEL GLM for the NAM populations. We additionally tested a mixed linear model framework, calculating residuals in the TASSEL MLM from models incorporating only a kinship matrix as a random effect, or both a kinship matrix and MDS coordinates as fixed, before testing SNPs against the residuals in the TASSEL GLM context for computational efficiency. We do not include these results because we found that, due to the high correlation between

population structure and flowering time, fully controlling for population structure reduced power to detect even well-known flowering loci such as Vgt1 or ZmCCT.

Resample Marker Inclusion Probability (RMIP)

The RMIP (*70*) approach is a resampling model selection procedure that can help identify the most informative SNPs. RMIP was first applied to maize to analyze leaf architecture traits by calculating residuals for each chromosome from a model that included terms from a joint linkage model for all other chromosomes; this was the method used to calculate residuals for Ames and USNAM (*3*). For CNNAM and the EUNAMs, instead of using joint linkage, the residuals were calculated from a mixed model fit with a leave-one-out relationship matrix for each chromosome, which was based on the markers from all other chromosomes except the target chromosome. The residuals were then used to fit a stepwise model to 100 random subsamples of 80% of the data. The RMIP for a SNP was set to the number of times it appeared in any model divided by 100. Tests found that RMIP is insensitive to rare alleles, and only works well when an allele is present in at least three families; because of the limited number of parents in the EUNAM and CNNAM populations, we found that this approach had limited power to identify SNPs and did not focus on the RMIP results in subsequent analyses (but see Figures S9-S18 for RMIP results in context of the other mapping methods).

Regional Heritability Mapping (RHM)

Regional mapping approaches should be better at detecting globally rare alleles present in parents of mapping populations, as those alleles are brought up in frequency in the progeny. Because of the variability in population structure and consequently MAF and SNP overlap between the populations, we first used a regional variance partitioning approach to identify regions that could be jointly analyzed between the populations. We follow the approach in Nagamine et al (*2*), testing each region of the genome using a model with two kinships, one derived from the region of interest and the other from the rest of the genome, fit as random effects Additionally, family is fit as a fixed effect for the NAM populations (Figure 1). The same model is also tested without the regional kinship, and the likelihood of the reduced model is divided by the likelihood of the full model to produce a likelihood ratio test statistic for the region. Following Nagamine et al, the test statistic is compared against a mixed chi-squared (0,1) distribution for significance. For this study, we used 10,024-site kinships, because preliminary tests indicated that this was the smallest number of SNPs to generate reasonably stable variance estimates in this dataset. Before metaanalysis, the resulting p-values are adjusted for multiple testing using the Benjamini-Hochberg method in the p.adjust method in the R package "stats".

Prelimary tests found the RHM mapping approach was only successful in identifying associations in the larger populations (Figures S1-2). While this method has been successful in humans, and located numerous regions of the genome in the Ames diversity panel, and in the large USNAM panel, very few regions were identified in the smaller NAM designs, and meta-analysis did not identify new regions of the genome. The lack of power in the RHM is a symptom of reduced recombination

inherent to especially the smaller NAM designs (*71*), and stimulated development of the BRHM method. In RHM, if LD extends beyond the target region, even if the causal signal is located in the region, including the region may not increase the likelihood of the model since the signal is also represented in the "rest of the genome" kinship.

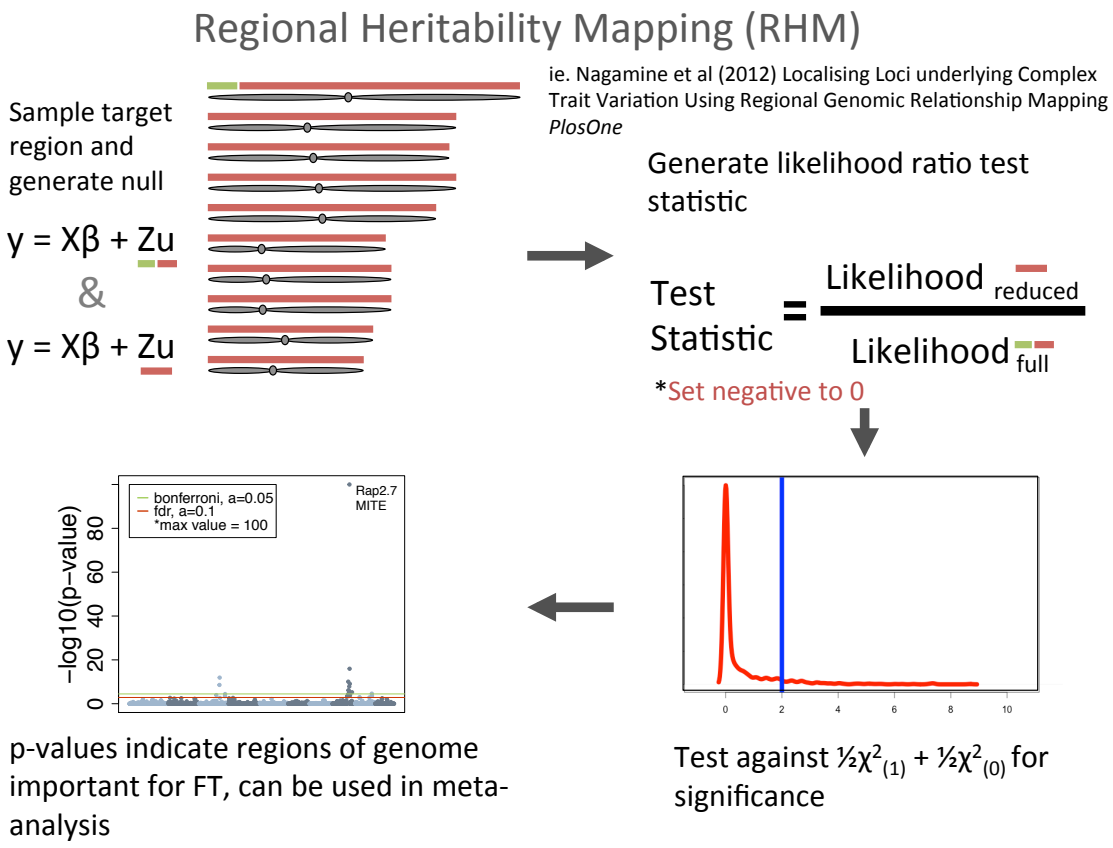


Figure 1. Regional Heritability Mapping (RHM)

Boosted Regional Heritability Mapping (BRHM)

Because the rest of the genome kinship can capture signal from the target region when LD is high, as is common in NAM designs, we developed an extension of RHM, BRHM, to increase power to detect associations, especially in the smaller NAM panels. The BRHM algorithm iteratively samples randomly placed regions on the target chromosome as random effects, controlling for the rest of the genome, in a mixed linear model framework using the REML solver implemented in LDAK against response phenotypes (Figure 2) (*71*). Each iteration randomizes the distance between each of the kinships in a model, and thus the degree of LD shared by any two kinships, generating a distribution of estimated heritabilities by any given region in the genome. We used 30,000 site kinships computed using the "Centered-IBS"(*67, 68*) method for structured populations or "Normalized-IBS"(*72*) for the Ames diversity panel as implemented in TASSEL(*61*). The value thirty thousand was determined empirically to maximize resolution while retaining stable relationships between kinships. The number of chromosomal kinships was scaled for one kinship per 50Mbp, to ensure a consistent probability that any two kinships would be in close physical proximity. For each iteration of the model, kinships were established at random start sites (in increments of 512 site blocks for computational efficiency in the final models), using a uniform random generator, with the caveat that they cannot overlap and any kinship established less than 30,000 sites from the end of the chromosome was adjusted to 30,000 sites before the end.

BRHM determines significance for regions by comparing the resulting distribution to an empirical null distribution of estimated heritabilities, representing

the signal for the region resulting from global population structure. The null variances are estimated using the same genomic kinships and models, but replace the true phenotypes with a null phenotype as the response. The null phenotypes randomly scramble the residuals from the model $y = X\beta + Z\mu$, where the random effects include the kinship calculated from the untested chromosomes and the error term, by adding the error estimate from another randomly chosen individual to the true BLUP estimate for the kinship for each taxon. This generates a null to test the hypothesis that the heritability estimated by the target kinships from the chromosome tested are identical to that expected from population structure as detected from the rest of the genome. Models were run with random starts for all kinships for the number of iterations expected to cover the genome to 40X coverage; after these iterations, the start site of the first kinship assigned to each model was chosen to ensure that a minimum of 40 models covered each SNP.

Paired real and null results were aggregated into regions based on unique model coverage. Models that did not converge (ran more than 25 iterations, or had a difference in the model likelihoods greater than 0.1 between the last 2 iterations) were excluded, rendering a small number of regions with less than 40X coverage. Because the resulting distributions of null and real results were not always normally distributed, significance for each region was evaluated using the non-parametric Wilcoxon signed rank paired two-sample one-sided test implemented in the R method "wilcox.test()". The null hypothesis for this test is that the median of the real population is not greater than the median of the null population. To control for differences in power across regions, model results were randomly downsampled to 30 models per test, and the

average p-value for thirty tests reported. If the distributions were identical – usually in the case that the null and real values were all zero – the p-value was set to 1. Before meta-analysis, the resulting p-values are adjusted for multiple testing using the Bonferroni method in the p.adjust() method in the R package "stats".

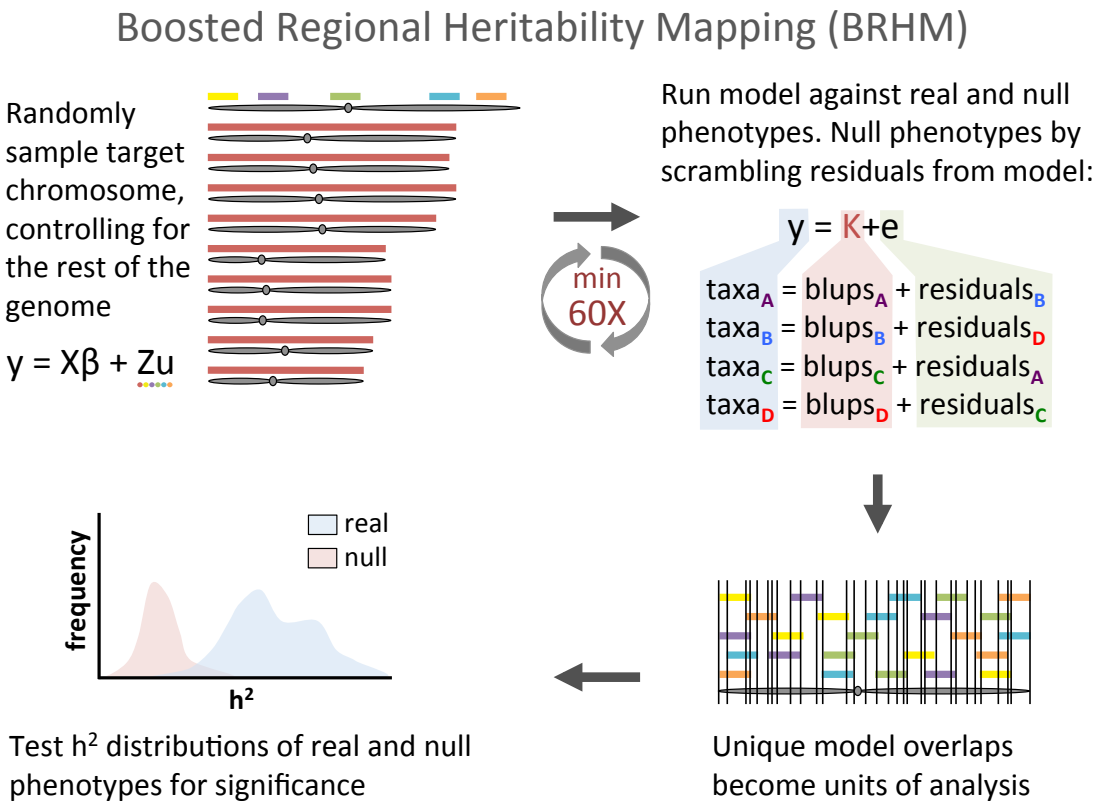


Figure 2. Boosted Regional Heritability Mapping (BRHM)

Meta-analysis for variance component methods

We combine p-values for regions of the genome with unique population-specific region overlap using Fisher's method for combining p-values (*73*), using the formula,

$$\chi^2_{2k} \sim -2\sum_{i=1}^{k} \ln\ (p_i)$$

where p_i is the p-value for *i*th hypothesis in 1…*k* populations. The resulting test statistics are $\chi^2_{(2k)}$ distributed, and the mean false discovery rate is calculated as *a* (k+1)/2k. While this method is not appropriate for combining GWAS results because it does not account for the direction of effects or heterogeneity between populations, this is not a constraint in variance mapping approaches because p-values are generated against internal controls and heritability is unidirectional. Meta-analysis for resampling GWAS entails directly combining results from each population; a meta-analysis for GWAS results was not conducted because of limited shared SNPs and because initial tests indicated that the structure of the populations and extended LD in the NAMs generated no positive associations for the *Vgt1* locus.

Machine learning analyses with GWAS results as the response

We employ the RandomForestClassifier method in Spark (*74*) in a Databricks environment, classifying additive p-values such that the top 1% of results are classified as 1, or true positives, the next 4% of results are skipped, and 2 million random SNPs from the remaining 95% are classed as 0, or true negatives. We employ a classifier rather than a regression based approach, because we are only interested in how the top regions rank relative to all the others, and a regression approach gives more weight to the insignificant results because they vastly outnumber the top hits. P-values for SNPs not reaching filtering criteria in a given dataset were assigned to 1.

We cross-validated by using the other nine chromosomes to predict the tenth. Accuracy is reported as area under the curve (AUC), the integration of the momentary false positive rate against the true positive rate at each unique observed raw predicted value. Overall AUC is simply the average AUC for the 10 chromosomal tests. Top predictors were reported as the average mean ranking across all 10 tests. Within a given tests, predictors are ranked by which splits they participate in for a given tree, averaged across all trees. Mapping predictors for the within population test were the Bonferroni corrected p-values and median heritabilities for BRHM results, the Benjamini-Hochberg corrected p-values and heritabilities for RHM results, and the proportion of models selected and a smoothing spline (smooth.spline () in R, with a spar value of 0.05) for RMIP results.

Combined GWAS

Genotypes and phenotypes were directly combined together for combined GWAS analysis to test the efficacy of directly testing phenotypes derived from non-overlapping research designs together. Genotypes were only minimally filtered for polymorphic sites, leaving 70 million genotypes in the dataset. Phenotypic variances for each population were shrunk using a mixed linear model (the lmer package in R) where population was fit as a random effect. To control for differences in heat units across different studies, phenotypes were then centered by shifting the mean of each individual of the population by the difference between the value predicted for B73 for the population and the value for B73 observed in the Ames diversity panel. All populations were run as a single analysis in a GLM framework using the

FixedEffectsLMPlugin in TASSEL, controlling for population structure using a dataset term and 5 MDS coordinates calculated from an IBS distance matrix in TASSEL(*65*).

Candidate gene lists

We extend the time to flowering candidate gene list from Dong et al(*75*), adding all of the genes from the CENTRORADIALIS (ZCN) gene family, after Danilevskya et al(*76*) (Table S1), as the 79 candidate gene list. We additionally use unmodified the 918 candidate gene list from Li et al (*32*). We also generate five new lists of candidate genes, one from the combined GWAS DTA results and two each from Ames GWAS results, with and without population structure correction for DTS and DTA, based on adding unique genes from a 100kbp window around top GWAS results sorted by p-value (Table S2). The Ames candidate lists, while confounded with true flowering pathway genes, should be more enriched for temperate adaptation alleles than the physiological candidate lists.

Machine learning analyses for candidate genes

We also use Random Forest classifiers to provide insight into the attributes of variants that explain top gene regions based on candidate genes for days to flowering and GWAS results from this study. SNPs from within candidate gene regions and 100kbp on either side are set as class 1, and set the rest to class 0. Tests are done the same way as with GWAS as the response. Mapping results, per-site F_{ST} estimates and power considerations such as MAF and coverage for Hapmap 3.21 SNPs provide

machine learning variables (the total list of predictors used in models is provided in Table S3). F_{ST} was calculated on a per-site basis between populations in vcftools (*77*), missing and negative values set to 0. Additional F_{ST} calculations were derived from sets of thirty Hapmap 3.21 taxa representing temperate US germplasm, tropical germplasm and Northern Flint germplasm (Table S4).

GO term comparison between populations

To better understand which classes of gene function are enriched in the BRHM mapping results for the five populations, we calculated enrichment across the top 1,000 significant genes. The top genes were chosen by iterating through the BHRM results ordered by the downsampled Wilcoxon-signed rank p-values for each population, adding genes from these regions until 1,000 genes were returned so that the genes surveyed for each population were balanced. The top 1,000 gene list was submitted to Agrigo (*78*) for singular enrichment analysis (SEA) against the complete GO database using Fisher's test for significance, no FDR correction, and a minimum of 5 mapping entries, using the Zea mays AGPv3.30 reference. The resulting GO terms were aggregated by the authors based on GO definitions (Table S5). A chi-square goodness of fit was calculated in Excel (CHITEST) from the sum for each category.

Results

This study combines five populations, four NAM designs and a diversity panel, totaling over ten thousand individuals (Table 1). These populations were

previously genotyped using GBS or an Illumina array and imputed with whole genome resequencing using FILLIN and FSFHap, respectively. This resulted in 70 million segregating markers across all populations (Table 1). The populations have variable genetic overlap based on the first two MDS coordinates, where Ames, as expected of a diversity panel, shows the greatest genetic diversity and CNNAM is the most isolated (Figure 3). Only half of the SNPs that survive filtering are shared between study populations (Figure S3), and the allele frequencies at those loci are highly variable (Figure S4).

Evaluation of relatedness between populations by MDS analysis and phenotypic comparison

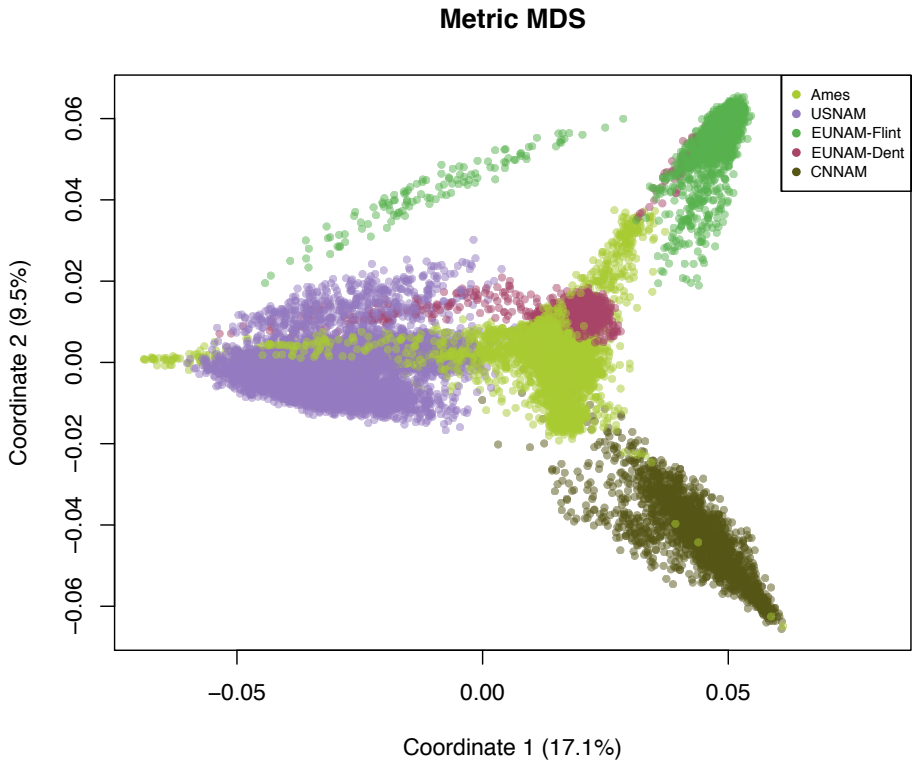


Figure 3. MDS of study populations from 70 million projected Hapmap 3.21 segregating SNPs (cmdscale () in R, using an IBS distance matrix generated in TASSEL).

To better understand the global genetic relationships between the founders of the four NAM populations, Figure 4 shows the genetic relatedness of the NAM-type population parents to a panel of GBS-genotyped landraces from across the Americas (Ames includes all of the USNAM parents). In this context, the NAM founders overwhelmingly cluster on the North American temperate-tropical gradient, rather than the South American-Andean gradient. CNNAM and EUNAM-Dent panel show a

more restricted geographic origin for the founder germplasm, while the other populations contain parents with a greater mix of American tropical and temperate origins.

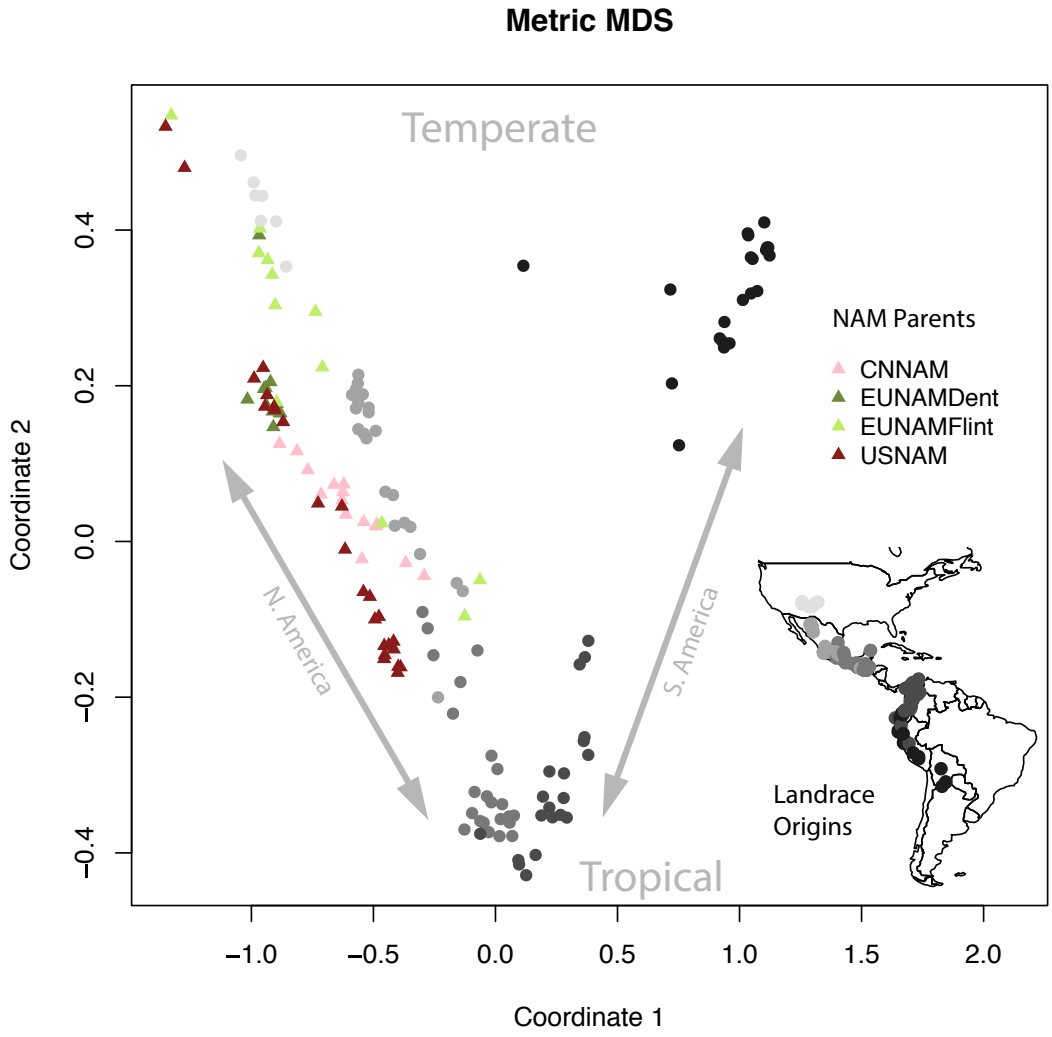


Figure 4. MDS of GBS genotypes American landraces from Takuno et al(*66*), replicated 10X each to drive the first two coordinates, and NAM population parents. The closest US-NAM parents to EUNAM-Dent are the temperate US Dents, including B73, B97, Oh7B, Ky21, and MS71, while the nearest US inbreds to CNNAM parents are South African inbreds M162W, M37W, and the Texas US line, Tx303. All of these are more proximal to the tropical Mexican landraces.

The five populations vary with respect to both the mean and the total variance for flowering (Figure 5). This is the result of both genetic variance within the populations, and environmental effects of the growing regions. Shifting the populations by the difference between the predicted value of B73 in the population relative to the measured value of B73 in the Ames population allows for comparison of the days to flowering of the different populations in the same environmental context. The EUNAM populations, which were grown in exclusively European locations, shift to the short flowering end of the distribution, and CNNAM shifts to a later flowering average relative to Ames. Problematically, the EUNAM-Dent population shifts to the earliest flowering position, earlier than EUNAM-Flint which flowered slightly earlier in Europe. This is likely due the increased relatedness between the EUNAM-Dent and B73, relative to EUNAM-Flint, giving a better estimate of the true flowering.

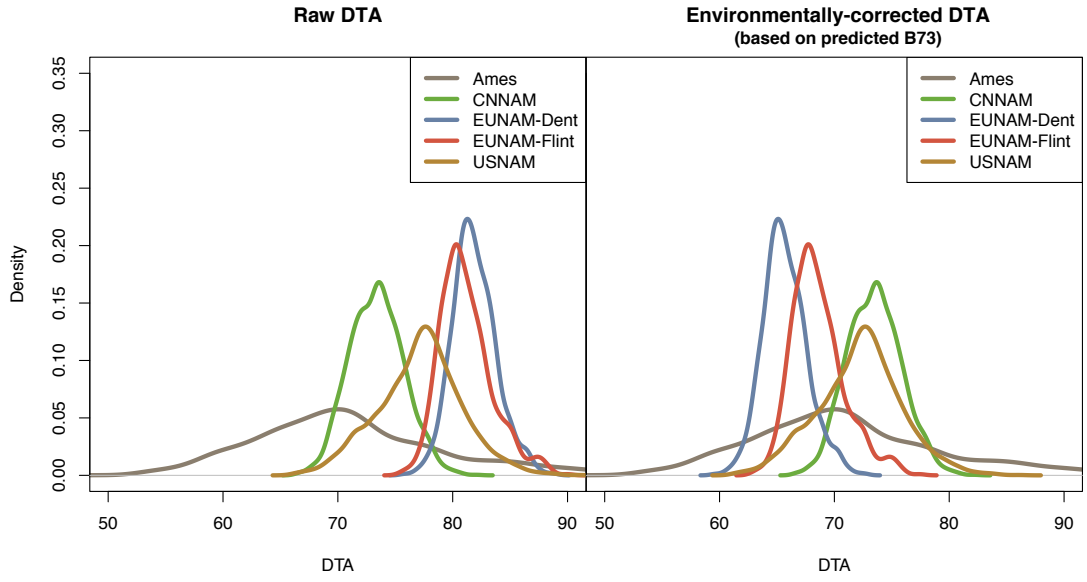


Figure 5. Distribution of reported field-corrected phenotypes for days to anthesis.

Cross population prediction of days to flowering

We performed cross-population prediction to better understand how the genetic basis for flowering is structured across these populations. Cross-population predictions were generated pairwise with RR-BLUP in TASSEL across all genomic SNPs (Figure 6). Ames has the best cross population prediction accuracies (given throughout as the Pearson correlation between observed and predicted, conservatively) overall, and unsurprisingly given the close relationship between Ames and USNAM, the best prediction accuracy is when the model is trained on Ames and predicting USNAM (r = 0.78 for DTA; 0.67 for the reverse). Ames predicted all of the populations better than the others, with the exceptions of CNNAM, which might be expected since Ames is a diversity panel and suggests that Ames is a superset to everything but CNNAM (Chinese lines are present, but poorly represented in Ames (*11*)). The EUNAM-Dent population has the worst overall cross population prediction with all of the populations in the study, and is predicted poorly by everything but Ames. We also looked at predictions using only SNPs significant in mapping, and using SNPs from the functionally annotated regions of the genome, and found that subsets can sometimes improve predictions if genomic prediction accuracy was low, but never if the accuracy was already high (Figures S5 and S6). An exception where a well predicted population is further improved by a subset is when the EUNAM-Flint population is predicting Ames or USNAM; in both of these cases prediction accuracy is already high, and genic or open chromatin subsets further increased accuracy.

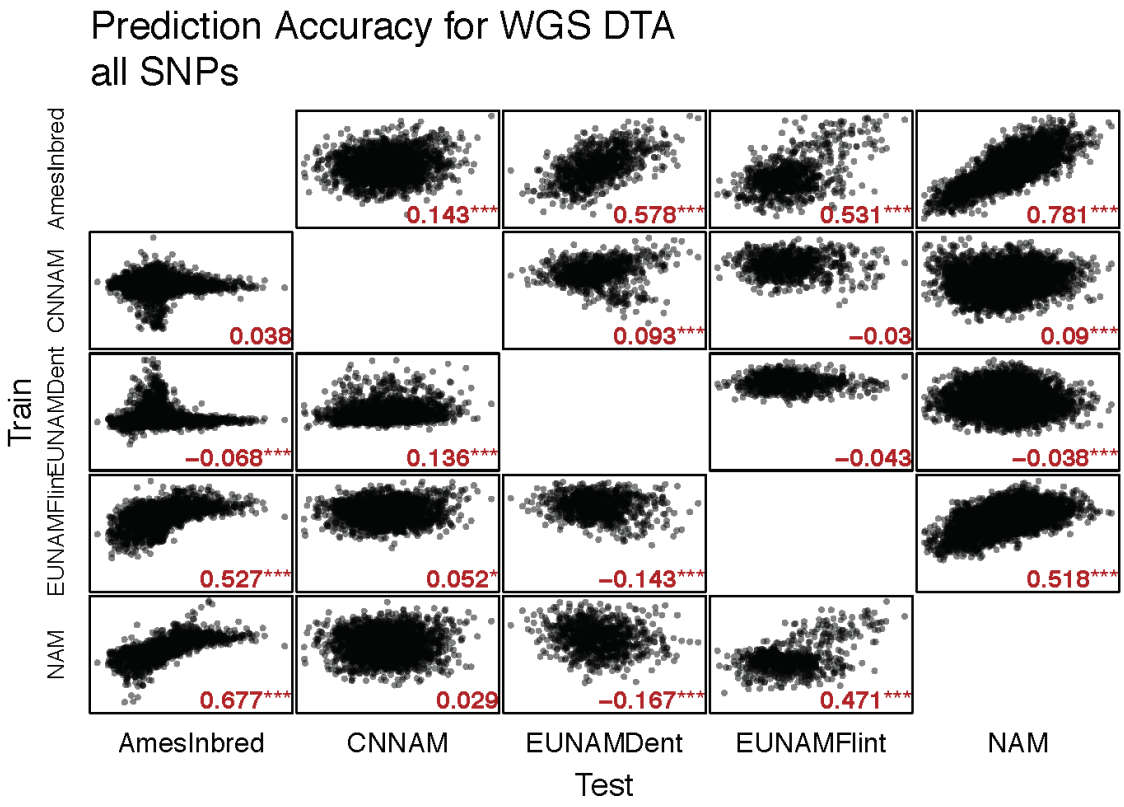


Figure 6. RR-BLUP cross population prediction accuracies for DTA using all 70 million segregating SNPs in Hapmap3.21. Genetic similarity matrix generated using the Centered-IBS method in TASSEL.

Because prediction accuracy is a function of relatedness, we calculated F_{ST} statistics for all pairs of populations. Figure 7 shows a correlation of -0.41 for F_{ST} and prediction accuracy for DTA across all markers, confirming a relationship between close relatedness and high prediction accuracy, but there are outliers that do not fit the expected pattern. Figure 7 shows that CNNAM and EUNAM-Dent generally have lower cross-population predictions than expected by population relatedness, with the exception that EUNAM-Dent is predicted similar to expectation by Ames. In contrast, EUNAM-Flint has higher than expected prediction accuracy with USNAM.

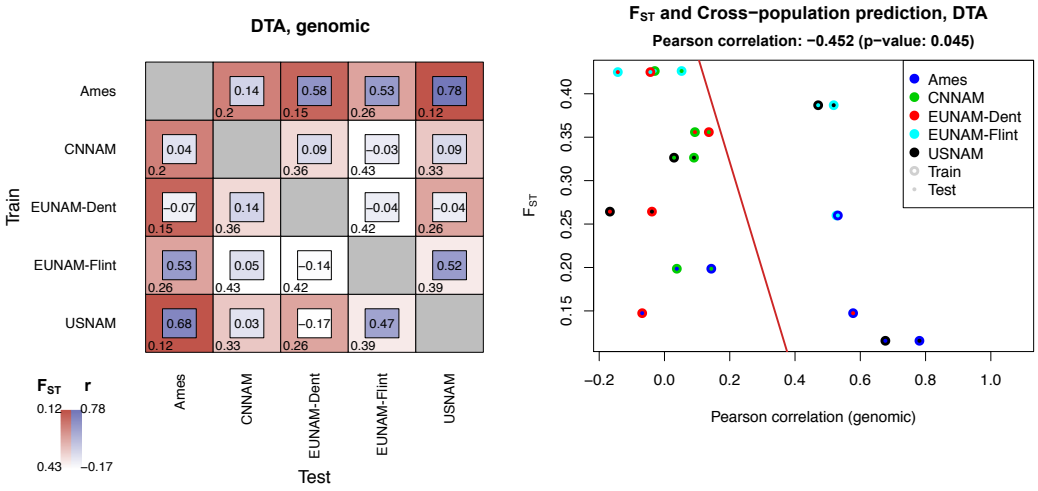


Figure 7. F_{ST} and cross-population predictions (Pearson correlation), DTA.

<u>Mapping days to flowering loci across the genome by population, and combining results in meta-analysis</u>

We mapped SNPs and regions of the genome for populations individually, comparing overlap of significant mapping results between the populations, and with published candidate gene lists. We performed GWAS and resampling GWAS (RMIP) to identify important SNPs, and two regional variance component approaches, RHM and the novel BRHM extension, so that populations with minimal SNP overlap can be compared on a regional basis. We found the RMIP and RHM approaches to be inappropriate for the smaller NAM populations, and did not pursue further analyses.

GWAS results were run for each population, and because changes in flowering time generate population structure, we performed GWAS with no population structure correction and minimal population structure correction (a family term for NAM populations, and 5 MDS coordinates for Ames). GWAS results generally tracked with major findings from other methods (Figure S7, S9-S18), with the NAM populations

tracking large regions of the genome, consistent with lower rates of recombination reducing resolution. Family control in the NAM populations exacerbated this pattern, removing peaks from all but the highest variance regions by BRHM or RHM, especially in the EUNAMs, which have less recombination (Figure S8). In Ames, the only population with resolution for GWAS, population structure control changes the significance of regions, but still identifies regions as significant that BRHM or RHM do not.

GWAS results were not comparable across populations in meta-analysis due to lack of shared segregating SNPs (Figures S3 and S4). As expected, the NAM populations did not gain resolution on their own from marker-level testing, due to high local LD resulting from minimal recombination inherent to the NAM designs (Figure S7). To assess overlapping significant loci, despite lack of overlap of SNPs between filtered sets in each population, we used a random forest classifier machine learning approach, where additive p-values for excluded SNPs for a given population were imputed to 1 (Figures 8, S19). Without population structure control the results are similar to the cross-population predictions, Ames and USNAM are well predicted, mostly by each other. EUNAM-Flint is predicted to a lesser extent by Ames, followed by NAM, and EUNAM-Dent is predicted equally well, but by EUNAM-Flint. CNNAM is predicted less well than by chance. After accounting for population structure, no population is well predicted by the others, and only Ames have a positive predictability.

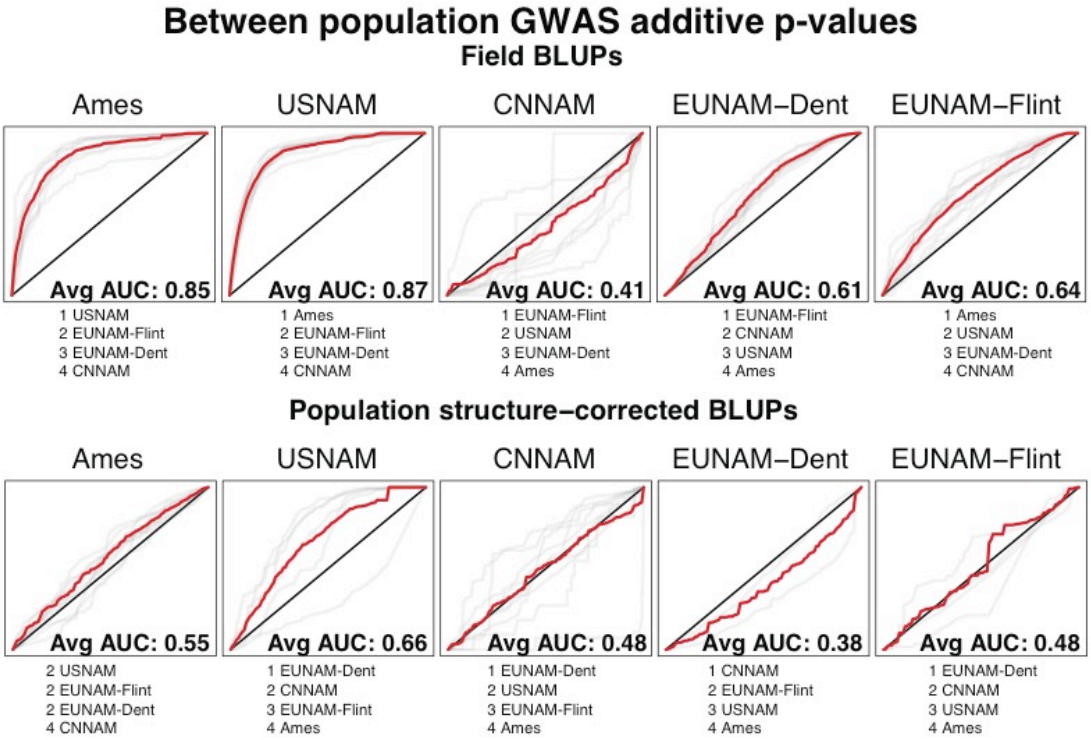


Figure 8. Average AUC (in bold) and predictor rankings across all chromosomes from random forest classifier for GWAS results between populations to evaluate overlap in GWAS results between populations. The red line represents the mean of the chromosomal ROC curves, sampled at 0.01 intervals. Predictors are equivalent GWAS (with or without population structure correction) for the other populations. Trained on nine chromosomes and tested on the 10th. Not all chromosomes for the population structure corrected results have top predictors, and are excluded from overall AUC calculations.

Because the GWAS results contain signal from population structure as well as from underlying flowering loci, we also tested the GWAS results against measures of population differentiation between three different N. American germplasm pools, tropical, temperate, and Northern Flint; the temperate germplasm results from the admixture between the southern Dent and N. Flint pools (Figures 9, S20). Confirming the inference from the cross population predictions, Ames is predicted very well (overall AUC of 0.89) by the tropical contrasts, followed closely by USNAM, with EUNAM-Flint well predicted at 0.67, EUNAM-Dent rather poorly predicted at 0.59

and CNNAM slightly negatively predicted with an AUC score of 0.47. Population structure control with a family term for the NAMs and 5 MDS coordinates for Ames reduced prediction accuracy by population differentiation, but did not eliminate the effects, confirming that flowering time is closely tied to population structure.

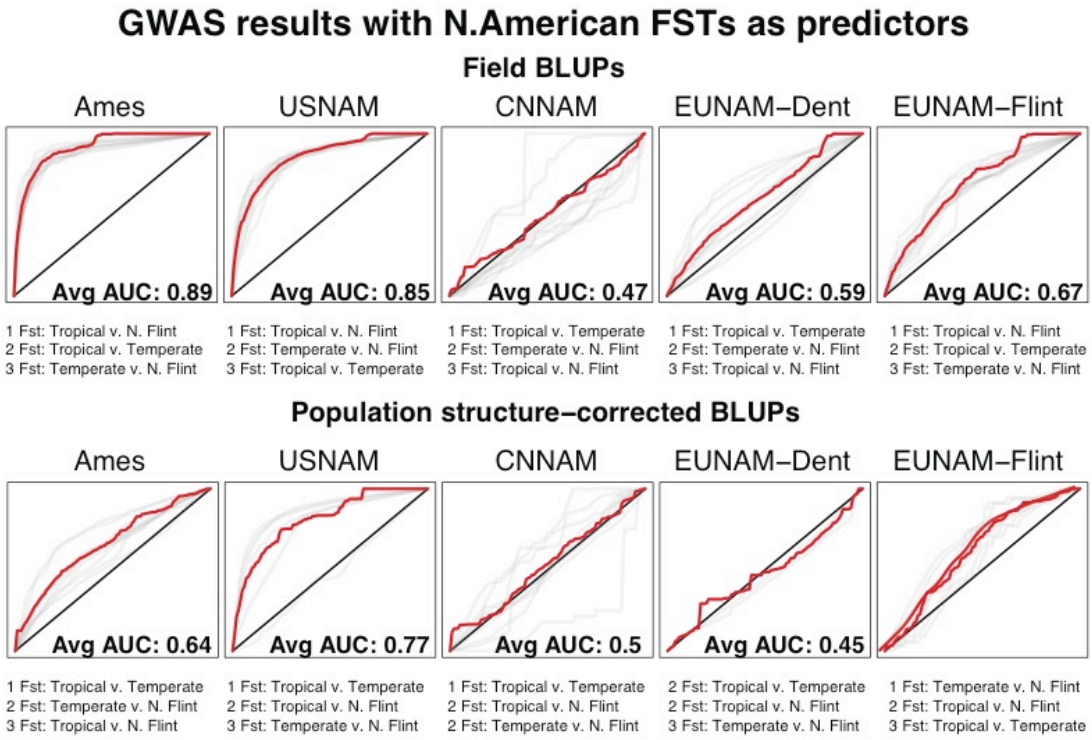


Figure 9. Average AUC and predictor rankings across all chromosomes from random forest classifier with GWAS additive p-values as the response, and N. American F_{ST}s as the predictors. Trained on nine chromosomes and tested on the 10^{th}. The red line represents the mean of the chromosomal ROC curves, sampled at 0.01 intervals. Not all chromosomes for the population structure corrected results have top predictors, and are excluded from overall AUC calculations.

BRHM succeeds in detecting associations, even when LD is high, while still controlling for false associations. Figure 10 shows extended LD around ZmCCT, as indicated by a bimodal distribution of estimated heritabilities for a linked region upstream, where heritability is high when no kinships are more proximal to ZmCCT, and near zero when they are. Both the RHM and BRHM methods successfully control

for this downstream linked region, as it is not significant in either (although it is possible there is real signal from this region, as a candidate gene is nearby, but it is not detectable due to genome structure near ZmCCT). Although false association is well controlled in both, the RHM method is more conservative than BRHM, with lower resolution; BRHM is significant for a smaller region than RHM around ZmCCT, and identifies additional regions across the chromosome, many nearby flowering candidates. BRHM also controls for false association due to population structure, which is a problem for association mapping in diversity panels when the trait is highly correlated with population structure (Figure 3).

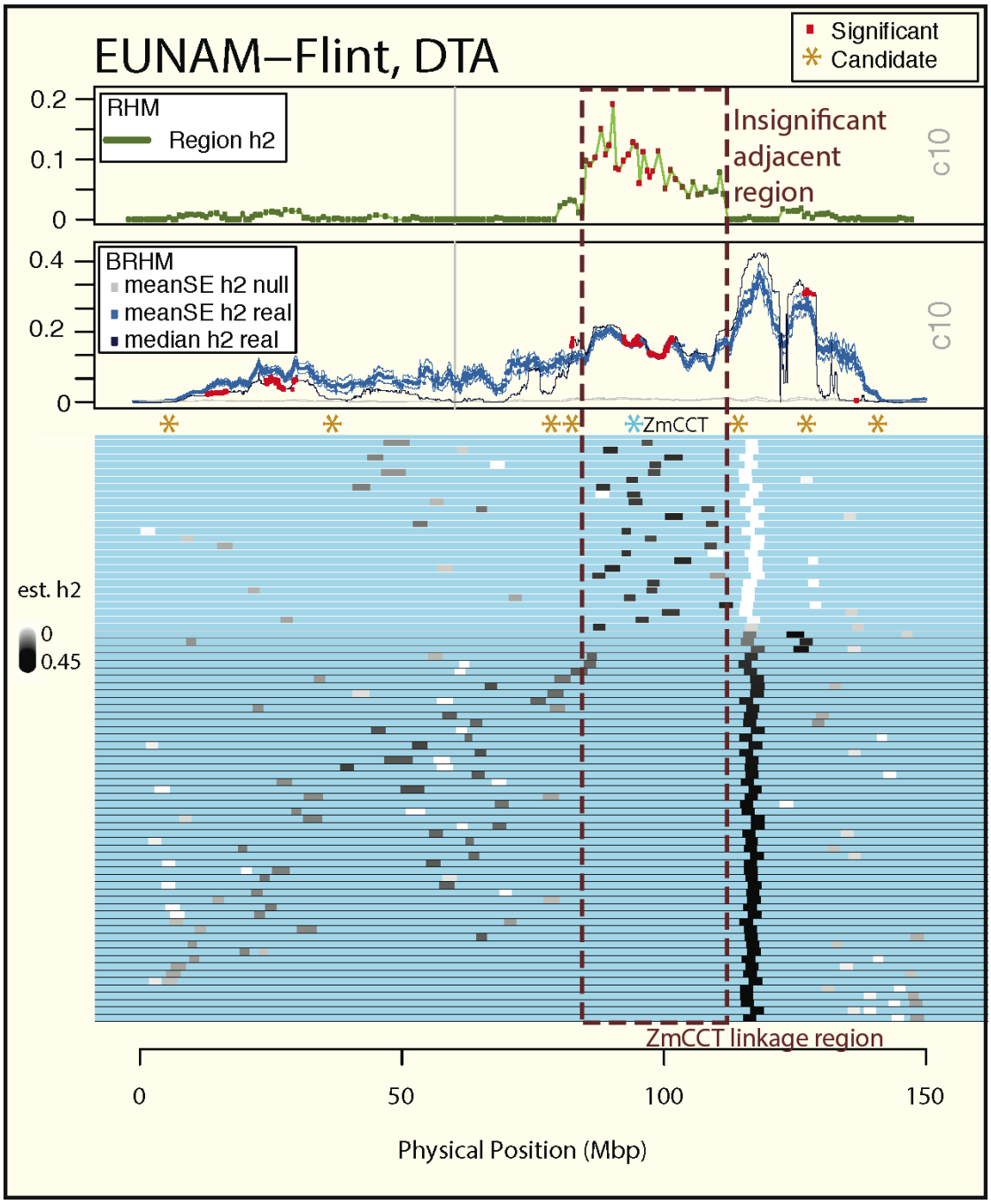


Figure 10. RHM and BRHM results for chromosome 10, and the variances explained for all models in BRHM that contain a kinship that falls on the region downstream of ZmCCT, and has the greatest difference between the median and mean variance estimated. This figure illustrates how large haplotype blocks can generate high variances explained based on variance in sampling, and also how the BHRM model correctly determines significance for such regions, with more power than RHM to detect candidates.

Overlap in significant regions between populations is low in BRHM. The greatest overlap within significant regions for both populations is between the EUNAM-Flint and USNAM populations, followed by Ames and USNAM (Table 4, Figures S6-7).

Table 4. Overlap between significant BRHM results at Bonferroni corrected alpha = 0.01 between populations. Calculated as the shared overlap/total number significant in either population. The top diagonal is for DTS, and the bottom is for DTA.

	Ames	CNNAM	EUNAM-Dent	EUNAM-Flint	USNAM
Ames		0.07	0.038	0.085	0.116
CNNAM	0.062		0.064	0.079	0.096
EUNAM-Dent	0.048	0.061		0.079	0.089
EUNAM-Flint	0.083	0.101	0.04		0.134
USNAM	0.109	0.091	0.104	0.131	

Combined GWAS

Directly combining populations for joint GWAS analysis, bypassing the need for meta-analysis, is desirable to avoid the loss of power inherent in meta-analysis. Combining populations was challenging, however, because the phenotypic variances between populations are due to both the different environments in which the samples were grown, but also due to differences in genetic diversity of the datasets (Figure 5). We shrunk population variances using dataset as a random effect, then predicted the value of the reference line, B73, in each population and shifted each value for the population by the difference between B73 in Ames and the predicted B73 for the population. This was done because the heat units in the different studies vary, so this method centers the data on genotypes rather than a grand mean. This shift better standardizes taxa genetically similar to B73 than those more distant, creating another

bias reflected in the rank change between EUNAM-Flint and EUNAM-Dent before and after correction, a known error because the populations were grown in the same environments (Figure 5). The biases from combining datasets add to known biases for GWAS conducted for flowering time, and GWAS for populations with extended linkage such as found in the NAMs, resulting in extreme deviation from uniform for p-values (Figure 11). Despite this, the well-known Vgt1 locus and ZmCCT, which is significant in 4 of the 5 populations (Figure 10), shows clear and locally dominant signals (Figure 11, inset). However, genome-wide these signals are not especially notable, inconsistent with previous studies (the highest signal is on chromosome 1, in a region not noted for high effect days to flowering loci, but that is a high-linkage region of high effect in many of the populations (Figure 11)).

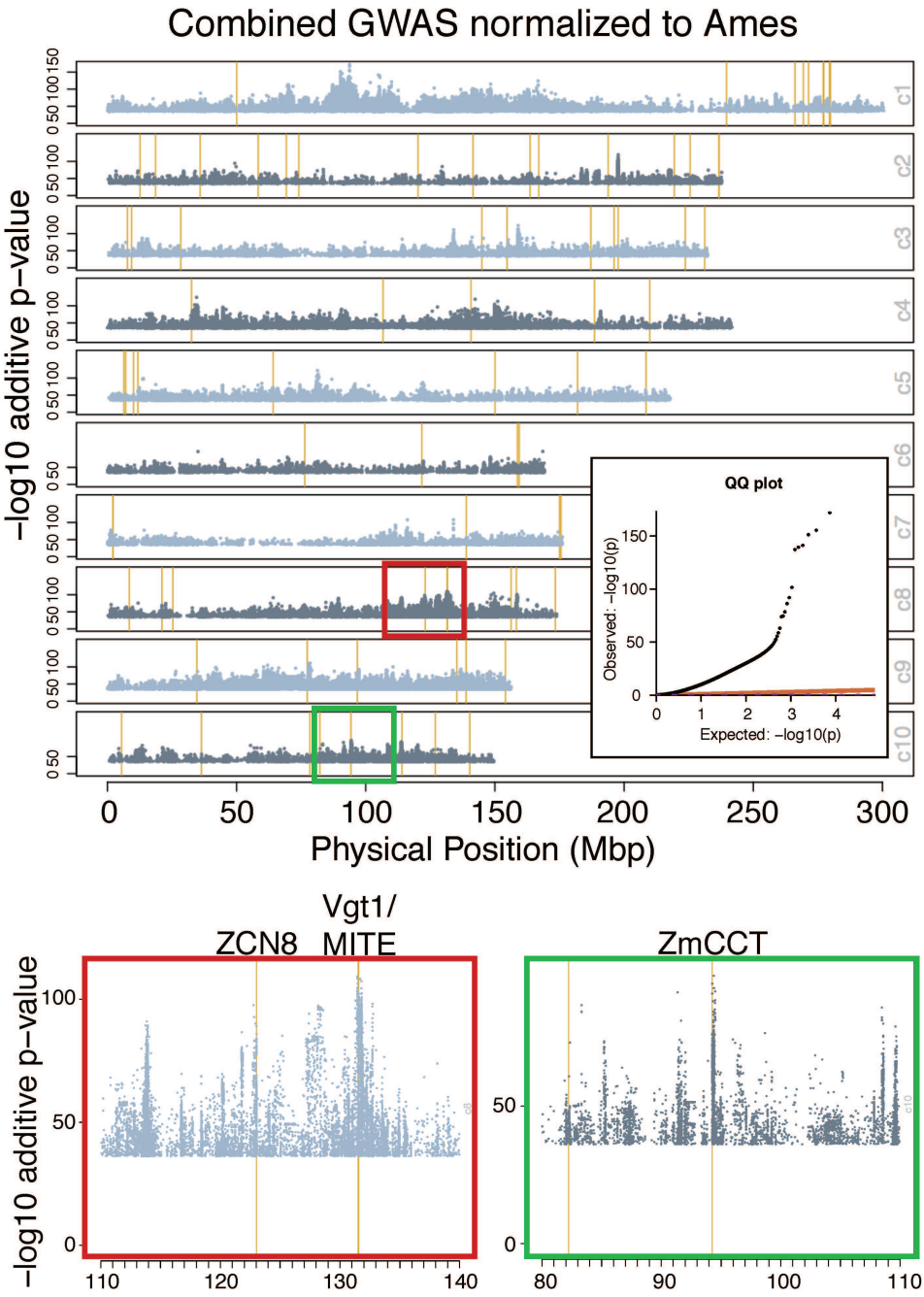


Figure 11. Combined GWAS analysis where traits are shifted by the differences between the predicted value of B73 in the population and the value of B73 in Ames, and the populations shrunken to account for variable GDD in the different study ennvironments. Population structure is controlled by 5PCs and a dataset term. Top 300,000 points by additive p-value plotted. Y-axis tick marks every 50. Vertical yellow lines represent Dong et al/Danilevskya et al candidate genes (Table S2). QQ-plot for combined GWAS results downsampled to 0.0001 (but including the top 10 values). The top flowering time loci in Ames and USNAM are highlighted.

Candidate gene enrichment

We looked at enrichment for the Dong et al days to flowering candidate genes focusing on the BRHM method. Most of the populations show slight enrichment for flowering candidate genes, between 1-2X (Table 5, Table S7). Candidate overlap between populations is also concentrated among the top few, previously identified, flowering loci, namely ZCN8 and the Vgt1 locus on chromosome 8 and ZmCCT on chromosome 10 (Table S1).

Table 5. Proportion of Dong et al candidate genes in significant regions relative to random genes within significant regions. Significance based on Bonferroni corrected significance for BRHM method, Benjamini-Hochberg for RHM, and for SNPs selected in at least a certain number of models for the RMIP method. Random genes in significant regions are the average of 100 samples.

	DTA			DTS		
BRHM	0.001	0.01	0.05	0.001	0.01	0.05
Ames	0.968 (6)	1.316 (10)	1.22 (11)	1.297 (8)	1.282 (9)	1.424 (13)
CNNAM	1.579 (6)	1.777 (11)	1.31 (11)	2.174 (8)	2.609 (15)	1.724 (16)
EUNAM-Dent	1.019 (7)	1.046 (8)	1.012 (12)	1.372 (9)	1.171 (10)	1.033 (14)
EUNAM-Flint	1.156 (8)	0.937 (8)	0.915 (13)	1.441 (10)	1.202 (12)	0.904 (13)
NAM	1.146 (28)	1.095 (32)	1.044 (38)	1.139 (29)	1.16 (36)	1.164 (42)

*Candidates in significant regions/Average over 100 samples of non-candidate genes from the maize v3 gene models in significant regions (Number of candidates in significant regions)

Comparison of GWAS results for SNPs around physiological candidate genes to random matched subsets shows only enrichment for EUNAM-Dent and, marginally, USNAM without population structure correction in the short candidate list, but in the long list is enriched for Ames, EUNAM-Flint and USNAM (Table 6). With population structure correction, only Ames is significantly enriched for p-values in the short list, but all populations are enriched for genes in the long list. Candidate lists were also

generated for the top Ames results and combined GWAS results, based on the sizes of the flowering time candidate lists both with and without population structure correction. Ames is the population with the highest resolution and for which top hits should capture temperate adaptation loci that may or may not be related to flowering explicitly. P-values from around the short 79 Ames gene list is enriched for all populations, but more so for EUNAM-Flint and USNAM without population structure control; with the population term only EUNAM-Flint and CNNAM (DTS) are enriched. Based on the top 918 genes for Ames, only EUNAM-Flint is enriched without population correction, and EUNAM-Flint and CNNAM and USNAM for for DTS only. The p-values from SNPs associated with candidates culled from the combined analysis (Ames-centered DTA) were enriched in all populations, which suggests that the combined analysis does tag regions with enriched signal shared across populations.

Table 6. Enrichment for low p-values within 2kbp of candidate gene regions derived from the 79 and 918 flowering candidate gene lists, and matched sized lists from the Ames and combined GWAS candidate lists (Table S2). P-values from within the candidate gene region are compared with a matched set of p-values from outside the regions using a one-sided Wilcoxen signed rank test in R. Only results from the Ames candidate list matching the GWAS source file by trait and population structure correction are presented.

GWAS source	Candidate source	Trait	Ames	CNNAM	EUNAM-Dent	EUNAM-Flint	USNAM
Based on Dong et al + Danilevskya et al gene list (79)							
uncorr.	Ames uncorr.	DTA	<1.0E-324	1.95E-03	0.965	6.93E-156	1.38E-135
		DTS	<1.0E-324	1.08E-29	1	1.61E-176	4.30E-56
	Combined	DTA	7.90E-173	3.67E-21	5.36E-12	9.14E-123	3.53E-45
		DTS	3.39E-162	1.24E-43	2.49E-13	4.73E-86	8.13E-41
	Phys. cand	DTA	1	1	1.86E-13	1	7.10E-12
		DTS	1	1	8.66E-05	1	1.69E-04
Q-corr.	Ames corr.	DTA	4.06E-83	1	1	<1.0E-324	1
		DTS	4.77E-66	3.90E-166	1	1.69E-170	8.65E-01
	Combined	DTA	2.08E-50	3.97E-279	8.93E-31	7.34E-77	<1.0E-324
		DTS	1.09E-55	4.73E-142	7.16E-10	5.75E-78	<1.0E-324
	Phys. cand	DTA	0.063	1	1	1	9.67E-02
		DTS	2.83E-06	0.153	0.818	0.78	1
Based on Li et al gene list (918)							
uncorr.	Ames uncorr.	DTA	<1.0E-324	5.31E-09	7.90E-08	3.72E-307	<1.0E-324
		DTS	<1.0E-324	2.71E-237	1.26E-14	<1.0E-324	<1.0E-324
	Combined	DTA	<1.0E-324	4.52E-30	6.30E-18	1.58E-284	<1.0E-324
		DTS	<1.0E-324	1.41E-26	2.81E-03	1.61E-242	2.80E-269
	Phys. cand	DTA	0.031	0.012	5.71E-06	2.53E-02	6.60E-21
		DTS	0.149	1.73E-04	5.15E-03	1.91E-04	7.12E-31
Q-corr.	Ames corr.	DTA	<1.0E-324	1	1	<1.0E-324	1
		DTS	<1.0E-324	7.73E-164	1	<1.0E-324	1.44E-25
	Combined	DTA	<1.0E-324	<1.0E-324	1.12E-163	<1.0E-324	<1.0E-324
		DTS	<1.0E-324	1.02E-171	3.79E-08	<1.0E-324	<1.0E-324
	Phys. cand	DTA	0.063	1	1	2.19E-06	0.924
		DTS	1.28E-03	1	1	2.83E-15	1

We used machine learning at candidate genes (Li et al, Dong et al plus Danilevskya et al, and Ames without population structure correction) to determine if non-linear combinations of BRHM and GWAS results, power considerations such as

MAF and coverage, and measures of population differentiation between populations and on the temperate-tropical gradient could better explain results (Figures 12, S23). We expected this might be the case if differential segregation at candidate genes within different populations generated the observed pattern of minimal overlap within a given population. The ability to predict physiological candidate genes is only slightly enriched above baseline (an AUC of 0.5) when BRHM, GWAS, power and population differentiation are taken into account, suggesting that even non-linear combinations of these predictors cannot explain the candidate lists. Interestingly, in addition to mapping results, coverage statistics also rank high as predictors, suggesting that structural variation may be diagnostic for candidate genes. Prediction of the Ames candidates used Ames BRHM results as the top predictors, unsurprisingly, and were followed by mapping results for other populations, with EUNAM-Dent and CNNAM as the least important predictors among the top BRHM results. That BRHM results, rather than population differentiation scores or even GWAS results, are the best predictors for Ames suggests that the top hits in Ames are generally shared across populations, with EUNAM-Flint and NAM as the most important and EUNAM-Dent and CNNAM sharing less.

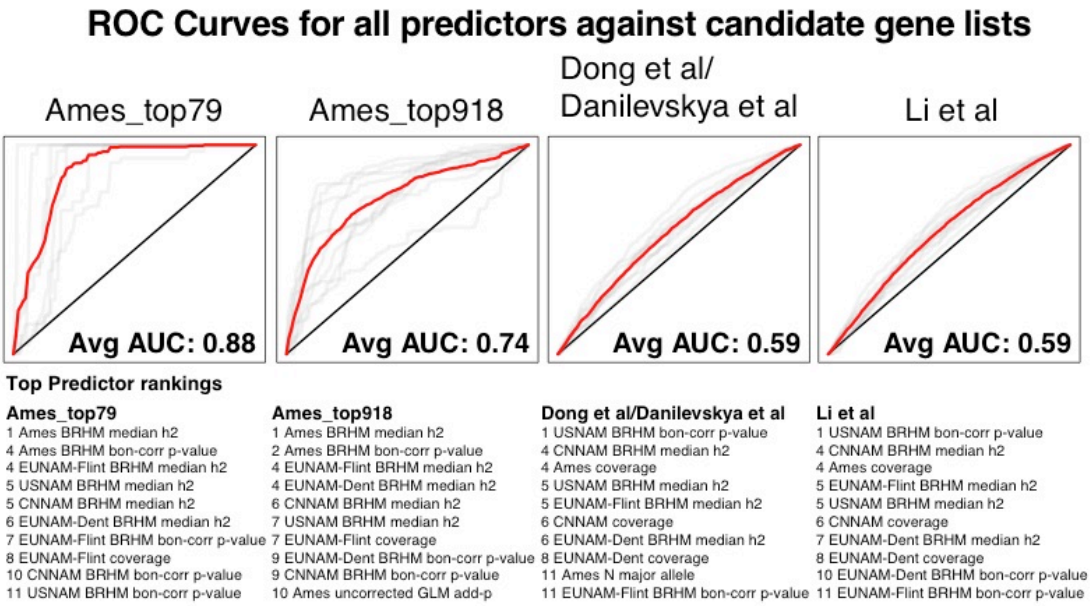


Figure 12. Machine learning using candidate genes as the response, and GWAS and BRHM mapping, power estimators, and F_{ST}s as the predictors. The red line represents the mean of the chromosomal ROC curves, sampled at 0.01 intervals. Candidate genes were classified as 1, with a 20kbp buffer around the gene start and end position (maize AGP v3). AUCs of 0 result when no candidates are found on the chromosome tested. Not all chromosomes for the Ames top 79 candidates have top predictors, and are excluded from overall AUC calculations.

Go term enrichment from top genes from BRHM analysis

The limited overlap in significant regions and candidate lists across all of the methods is not consistent with previous findings (*8, 27*) using lower density SNP sets, and not supported by the generally high cross-population prediction accuracies. The discrepancy could be due to truly different underlying genetic architecture that had previously been collapsed due to limited marker coverage, which should generate low cross-population prediction accuracies, differential allelic series at common genes across populations as was found in the USNAM population (*5*), or it could be due to differences in population structure, both historical and recent, shifting the signal from or differentially segregating for the same causal variant within populations. We looked

at GO term enrichment for top BRHM results to see if association regions share a functional enrichment to help reconcile the disparities between the limited overlap observed in the mapping results, and the broadly shared genetic basis for flowering in most populations implicated in the cross population prediction results.

We focus on the top 1,000 genes from the BRHM analysis for each of the populations to see if differences in mapping results are also related to differences in function (Table 7). A chi-square goodness of fit test for equal representation within biological functional categories reports a highly significant deviation from equal representation across populations, with a p-value of 7.54e-17 for DTA. All of the populations show similar levels of enrichment for gene models related to housekeeping, biotic response and macromolecules, a catchall category for macromolecule biosynthesis, metabolism, and catabolism (Table S5). Reproduction, signaling, development and biotic response are the classes with the largest deviations from expected. Ames and USNAM, which have the greatest genetic overlap, have distinctly different enrichment profiles, likely reflecting differences in resolution between the populations, and absolute variance in days to flowering.

Table 7. GO term enrichment against expectation of even representation for each population, based on the top 1,000 genes from the most significant regions of the genome for DTA, as determined by BRHM analysis. Chi-square goodness of fit p-value for test of deviation from the null hypothesis that there are no differences between populations for annotated function is 7.54E-17

	Ames	CNNAM	EUNAM-Dent	EUNAM-Flint	USNAM
AbioticResponse	1.026	0.982	0.765	1.002	1.224
BioticResponse	0.993	1.034	0.938	1.072	0.969
Development	1.016	0.894	0.866	0.935	1.28
Housekeeping	1.014	0.994	1.02	1.005	0.97
MacroMolecules	0.974	1.017	1.042	0.994	0.97
Regulation	0.958	1.05	1.107	0.978	0.903
Reproduction	1.159	0.937	0.83	1.156	0.952
Signalling	1.267	0.885	0.869	0.946	1.059

Discussion

Complex traits, such as yield, are well known to aggregate fitness effects across all pathways and systems of a plant, but this study highlights that this is also true for days to flowering (see also Li et al(*32*)). When a plant is induced to flower depends on the alleles present at ZmCCT, ZCN8, and Vgt1, but it also depends on the pleiotropic response of hundreds of other loci responding to signals for heat, circadian signaling, light quality, moisture availability, starch accumulation, and others (*3*, *4*). Functional biological constraints ensure some overlap across populations for high effect genes in the autonomous flowering pathway; this is supported by the increase in genome-wide prediction accuracy for poorly predicted populations when the top significant SNPs from any population are used. Candidate gene enrichment suggested around ten genes may be universally important and segregating. However, limited overlap in mapping results and differential GO term enrichment suggest that, as maize spread across the world during improvement and modern hybrid breeding, complex

population dynamics may have led to differential selection for secondary pathways implicated in temperate adaptation.

That some populations, namely EUNAM-Flint and CNNAM, do not predict Ames as expected based on population differentiation suggests more complicated population dynamics than simple relatedness (Figure 7). What better explains the patterning in prediction accuracy is the spread of the population founders on North American temperate / tropical gradient (Figure 4). The two populations with a localized geographic origin on this gradient, CNNAM and EUNAM-Dent, are the two populations with more limited cross-population prediction accuracy, and low prediction accuracy for GWAS results in machine learning. In contrast, EUNAM-Flint, which has higher than expected cross-population prediction accuracy when predicting USNAM, has parents spanning the tropical / temperate North American gradient, and F_{ST}s between Northern Flints and tropical American germplasm are top predictors in machine learning against uncorrected GWAS results. Lower than expected prediction accuracy relative to population differentiation could be interpreted in two ways; a narrow germplasm base or, if the population has high variance for flowering time, that these populations contain novel temperate adaptation not captured on either American temperate / tropical axis.

The history of germplasm introduction can shed light on the discrepancies between prediction accuracy and population differentiation especially for the EUNAM-Dent and CNNAM. The earliest germplasm in Europe was Caribbean in origin, and subjected to selection for early flowering upon entry into Spain, but it was only after the introduction of the Northern Flint varieties in the mid-1600s from the

northeast of the modern US that maize agriculture spread to European climates north of the Pyranees (*43, 46, 47, 80*). Genetic evidence from European maize suggests that most of the early flowering adaptation was acquired from the North American Northern Flint germplasm introduced in the 1600s (*43, 46*), but also that there are unique rare alleles in the Southern European (Spanish and Portuguese) germplasm (*47*). Finally, in the past century, the development of the European heterotic groups introduced primarily American Iodent, Stiff Stalk, Lancaster, and Minnesota germplasm into Europe, captured primarily in the Dent panel (*46, 81, 82*). Although Chinese maize now contains both early and late flowering varieties, the most likely first point of entry for maize to China was by Spanish and Portuguese traders through the Port of Macau, and these are reported to be humid, tropical adapted varieties(*44, 45*). A regular trading route between Acapulco and Manila, which started in CE 1565, could have introduced western lowland Mexican maize to Asia (*83*).. This panel of Chinese lines is dominated by lines whose origins likely predate modern hybrids of the 20th Century, suggesting an indigenous temperate adaptation to China based on a very different germplasm source.

The recurrent parent of the EUNAM-Dent population is representative of the agronomically important Iodent germplasm, and the additional lines in the dent panel derive from the US Stiff Stalk, US Lancaster, and Hohenheim Dent populations (*84, 85*). The Hohenheim Dents were bred from US temperate germplasm, and especially the early flowering Minnesota lines (*82*), which result from crosses between the US Southern Dents and Northern Flints, followed by selection for extreme temperate adaptation, with early flowering contributed by the Northern Flints (*81*). EUNAM-

Dent is well predicted by Ames, but not NAM; Ames is enriched for temperate US germplasm, including Iodent and Minnesota lines (*11*), while NAM is not (*59*). It is not surprising that Ames then predicts the Dent panel, as it is a superset, but it is perhaps surprising that the Ames GWAS results predict the EUNAM-Dent germplasm so poorly in machine learning, given that the pedigree suggests that temperate adaptation in the EUNAM-Dent panel is Northern Flint in origin. The lack of overlap in significant regions and poor cross-population prediction suggests that the Dent germplasm base is so narrow that there is little variance for flowering time or temperate adaption. This is possible, given the good prediction of Dent germplasm by Ames, and moderately supported by the reasonable, but lower than Ames or USNAM, narrow-sense heritability for the EUNAM-Dent panel, 0.7 (DTS) and 0.61 (DTA). Alternately, poor overlap and machine learning prediction could be a result of recent selection for temperate adaptation in the development of the Minnesota germplasm and subsequent selection in Europe.

Almost all of the CNNAM parents derive from Chinese sources across heterotic groups, and the recurrent parent is a derivative of the temperate Chinese landrace TangPiSingTou (*86*). While most of the CNNAM parents are temperate (*32*), they are significant at the major photoperiod locus ZmCCT confirming that the population contains tropical alleles. Additionally, broad sense heritabilities for CNNAM were high (0.91 and 0.90 for DTA and DTS respectively) so the genetic basis for days to flowering is not narrow. High heritability and lack of overlap in significant regions and poor cross-population prediction across all populations suggests that the CNNAM population does not suffer from a narrow germplasm base

but rather contains novel alleles for temperate adaptation, suggesting an independent origin relative to the other four populations.

The minimal enrichment for physiological candidate genes, even in a machine learning framework that can incorporate information from all of the predictors in a non-linear framework, suggest that days to flowering is primarily mapping temperate adaptation. If temperate adaptation is a suite of traits that incorporates more loci than those associated with the autonomous flowering time pathway, and independent adaptations for these traits from China and the US/Europe are relevant in NAM germplasm, genes in significant regions should be enriched for different non-autonomous flowering loci. The CNNAM and Dent panels show different patterns of enrichment in the top 1,000 genes from the other panels, tentatively supporting this hypothesis. In contrast, Ames and the EUNAM-Flint population are the most enriched of all the populations for reproductive GO terms, perhaps helping explain how these populations have good cross-population prediction, despite high population differentiation.

However, questions remain as to why Ames, USNAM, and EUNAM-Flint, which all share parents across the temperate tropical gradient, have high cross-population prediction accuracy, and can predict each others' GWAS results reasonably well (EUNAM-Flint) or very well (Ames and USNAM) do not share more significant loci in the mapping results. The good cross-population predictions and machine learning results suggest that these three populations share or partially share a common genetic architecture, and that the lack of overlap is an artifact of our inability to isolate signals within the data. Differences in population structure, and how LD shifts the

signal in mapping results from true causal loci, certainly contribute to lack of overlap in these structurally diverse populations. Environment, and genotype-by-environment interactions in the testing environments may also explain some of this discrepancy. The construction of the NAM designs reduce the spread of days to flowering across the population to those that can be crossed, reducing the total variance in ways which may be subjected to similarly unique pressures, even across multi-year/location studies. Likewise, the EUNAM trials were all located in Europe, much colder and further north than most of the NAM or Ames environments; the EUNAM populations predict B73 to flower a full two weeks later than in Ames, due to reduced GDD accumulation, and systemic genotype by environment interactions may affect associated regions. Finally, it is possible that some of the differences between EUNAM-Flint and the USNAM result from partially different genetic architecture derived from early Spanish selection for temperate adaption in early-introduced Caribbean germplasm.

These results raise the question of the suitability of simple threshold values for detecting similarity in complex, population structure confounded traits like flowering time (*88*).. Significance or heritability thresholds are modulated by population diversity, population allele frequency, LD with neighboring SNPs, population structure of the founders, and the unique set of environments the populations were evaluated in. This almost guarantees that there will not be a simple threshold to look for meaningful overlap. Machine learning provides an opportunity to look for site overlap in much more rigorous way, by allowing threshold values to vary we can better model non-linear relationships between power, true causal variants and

population structure to understand genetic architecture and target causal variation more precisely for breeding. Finally, this study highlights in CNNAM a unique set of germplasm not extensively utilized outside of China, which may provide a new source of novel alleles for breeding programs.

Supplemental figures

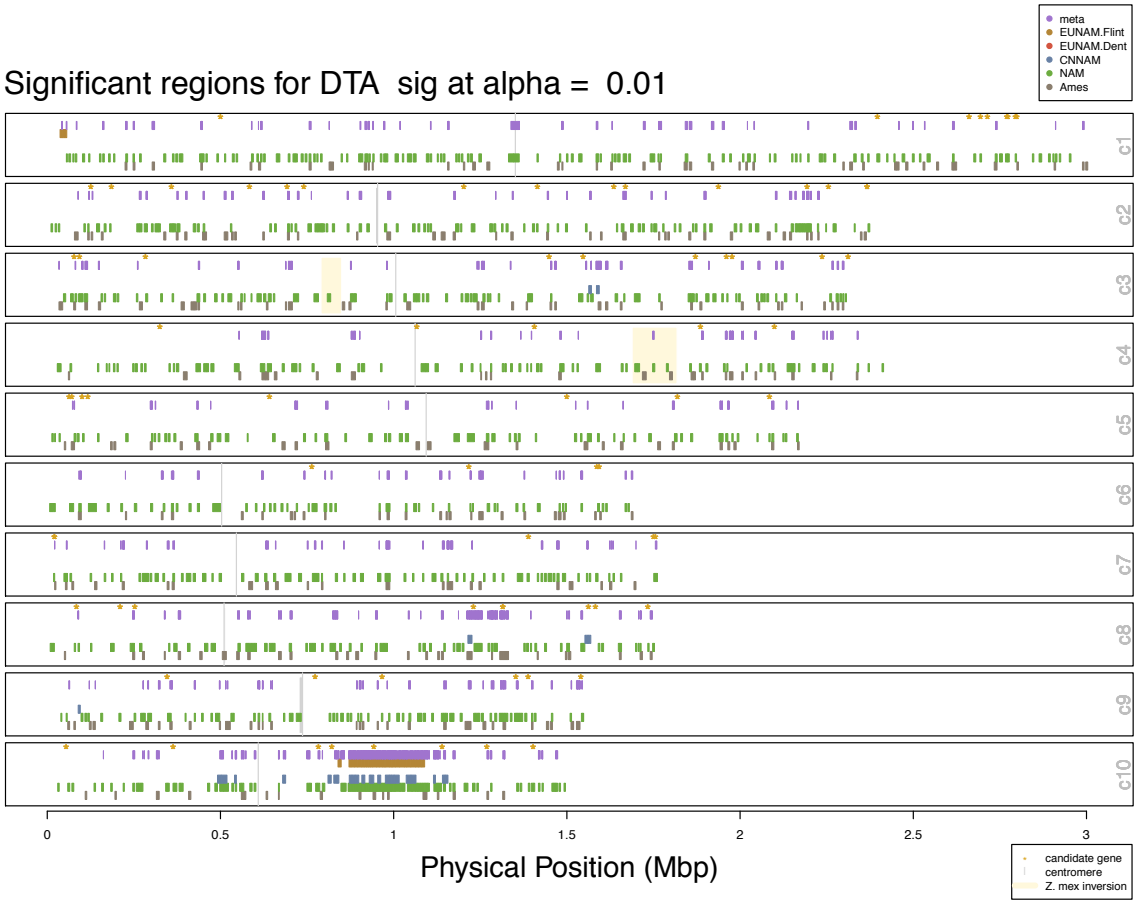


Figure S1. Significant regions and meta-analysis results for RHM method, DTA

Figure S2. Significant regions and meta-analysis results for RHM method, DTS

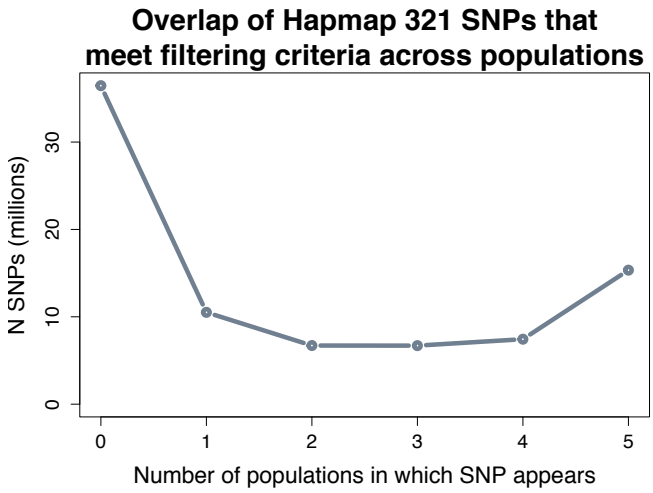


Figure S3. SNP overlap between populations in filtered projected Hapmap3.21 genotypes

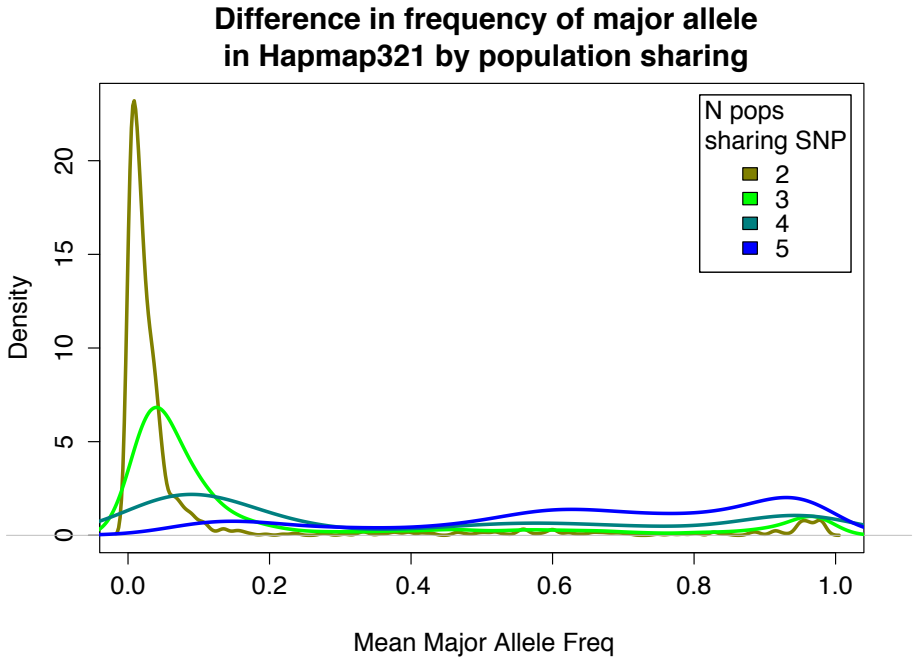


Figure S4. Density of the difference between the minimum and maximum Hapmap 3.21 major allele frequencies in shared filtered projected Hapmap3.21 genotypes

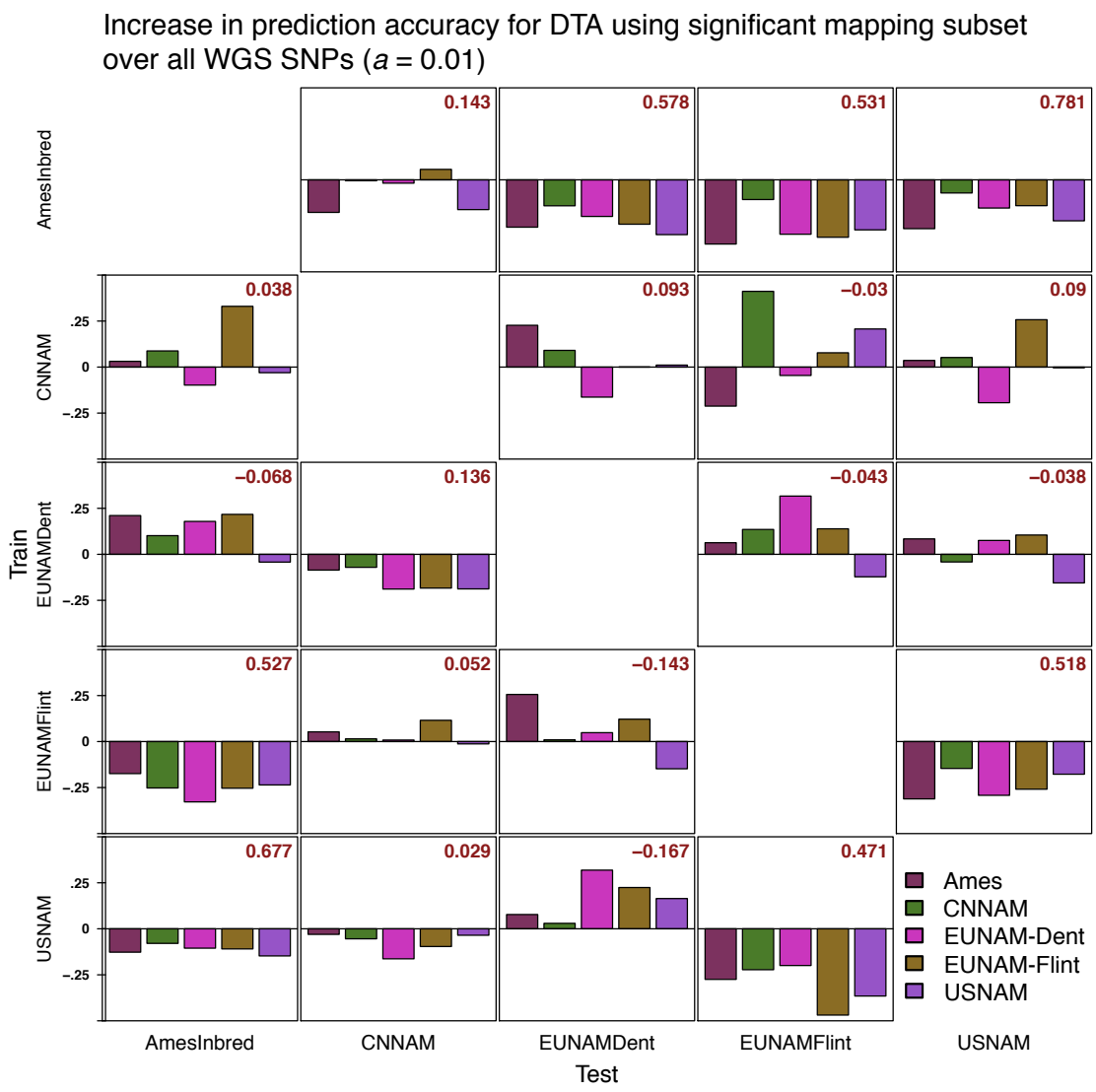


Figure S5. Increase in prediction accuracy for DTA using subsets of SNPs from significant regions in BRHM results. Whole genome accuracies reported in red.

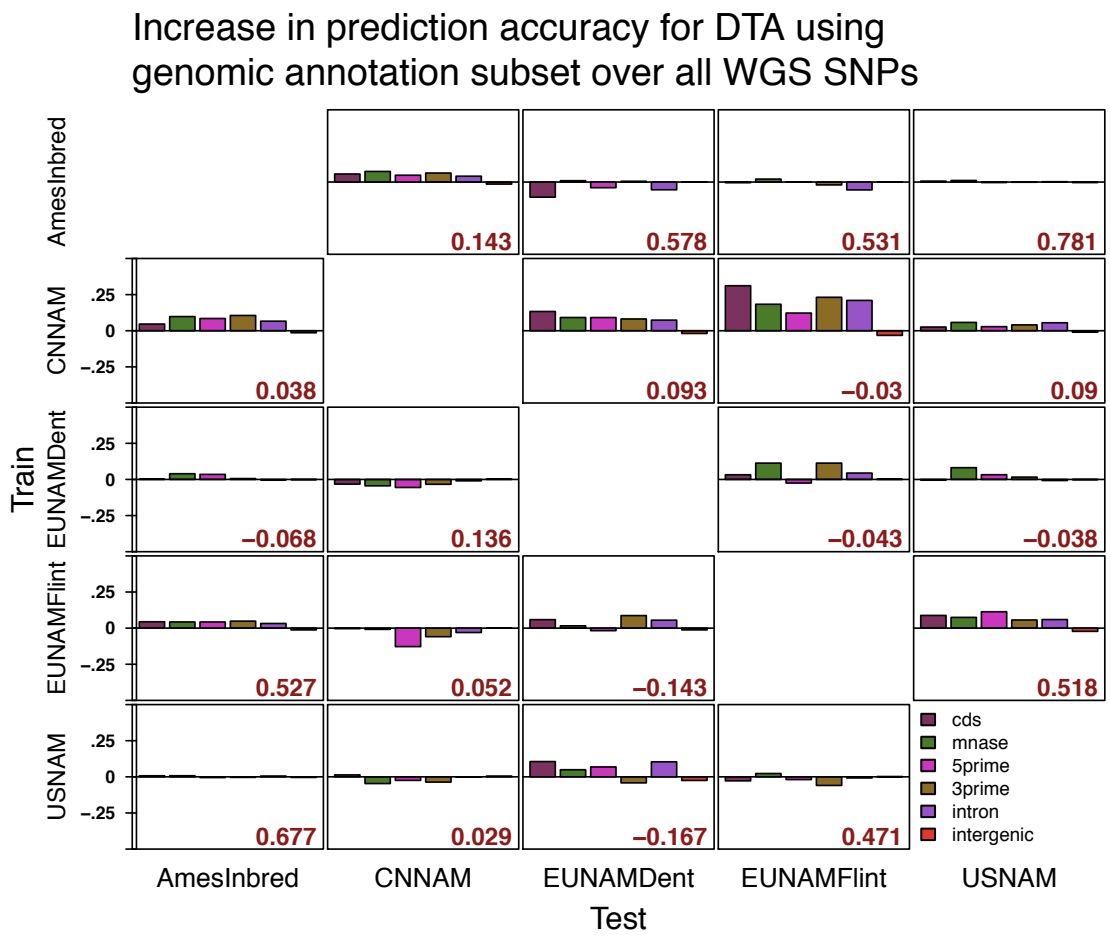


Figure S6. Increase in prediction accuracy for DTA using subsets of SNPs based on genomic annotations.

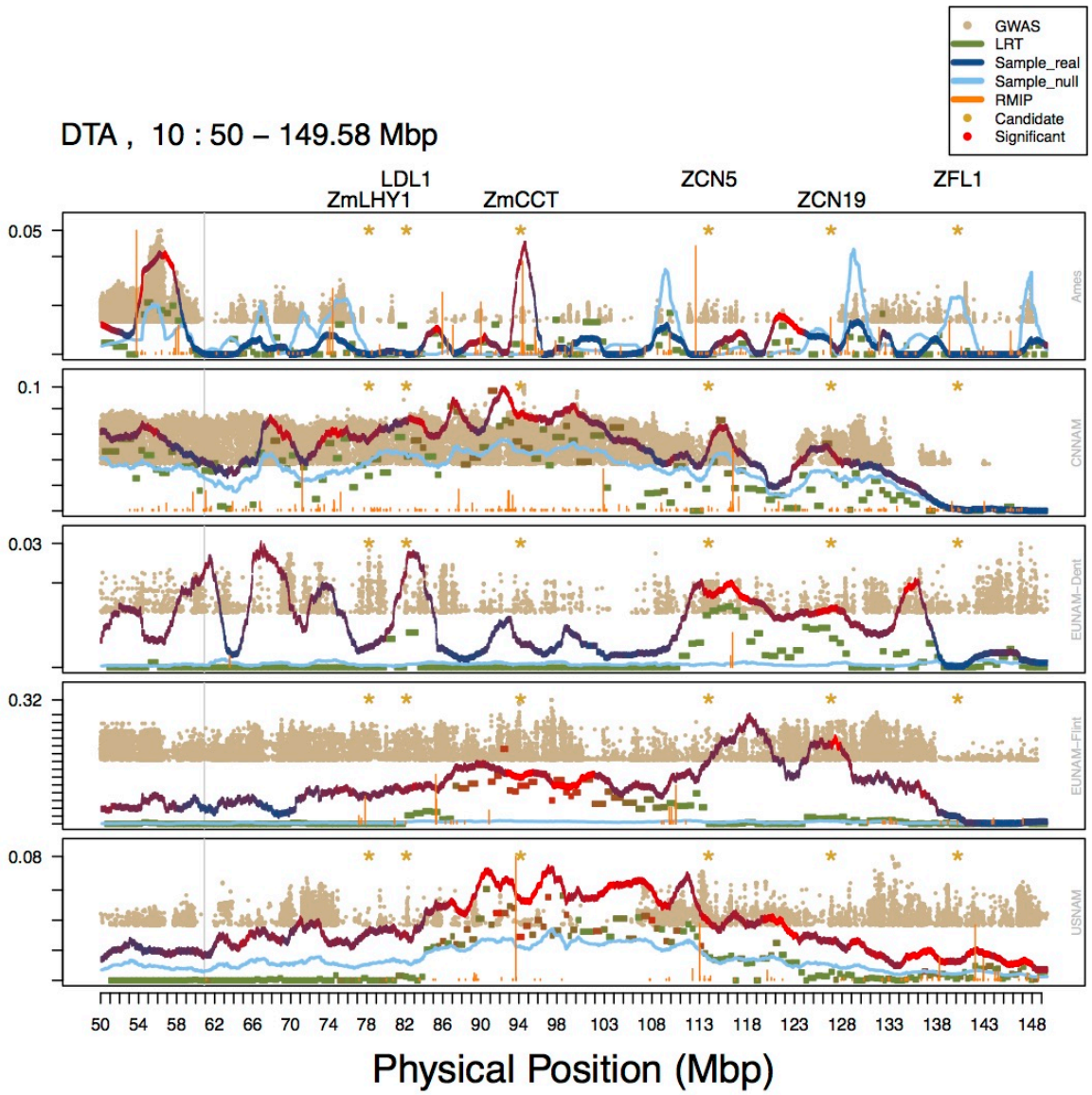


Figure S7. Combined results for all methods (GWAS (no population structure control), resampling GWAS (RMIP), RHM, and BRHM) for 50 Mbp region on Chromosome 10. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

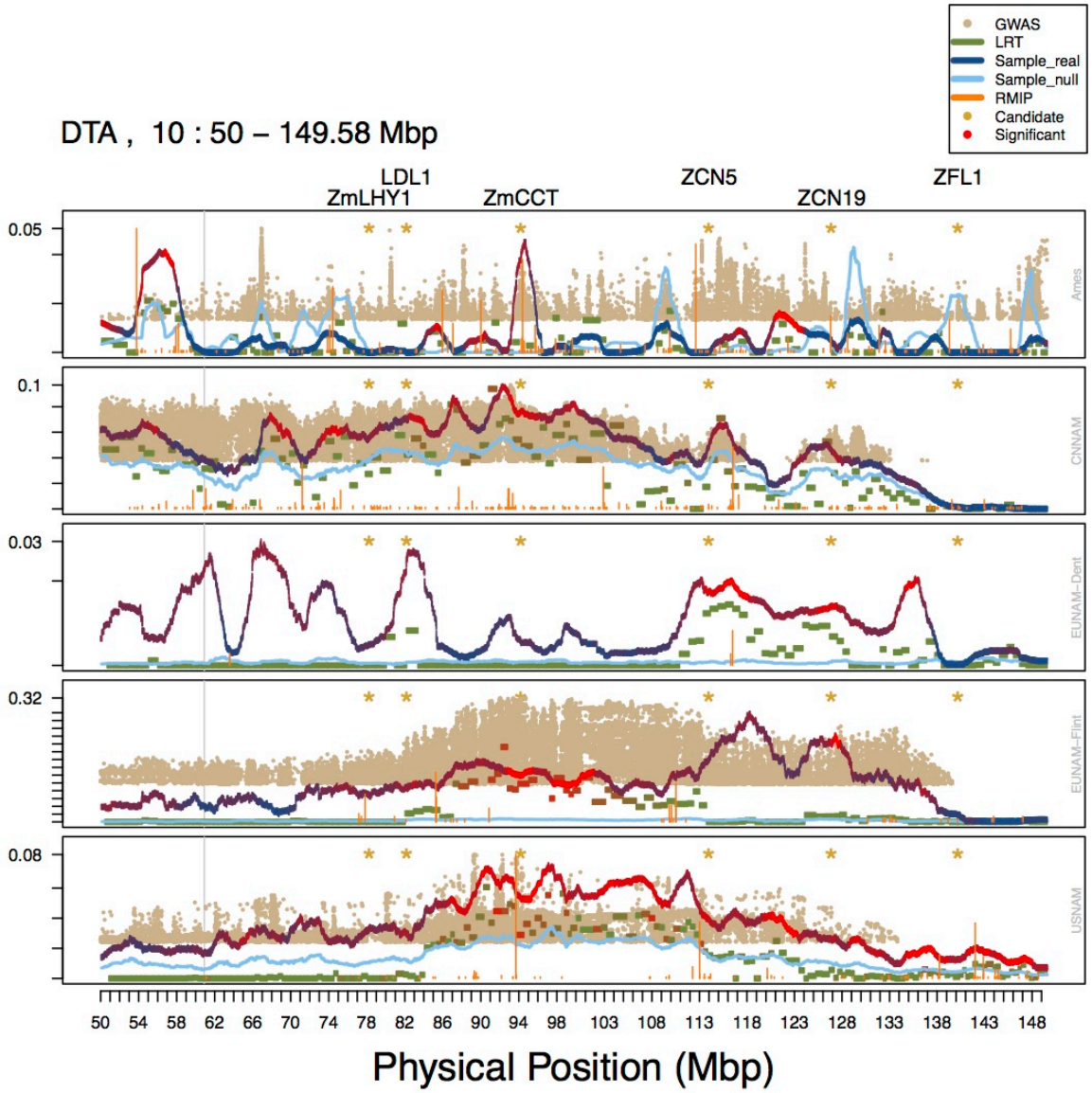


Figure S8. Combined results for all methods (GWAS (Family term or 5PCs for Ames), resampling GWAS (RMIP), RHM, and BRHM) for 50 Mbp region on Chromosome 10. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

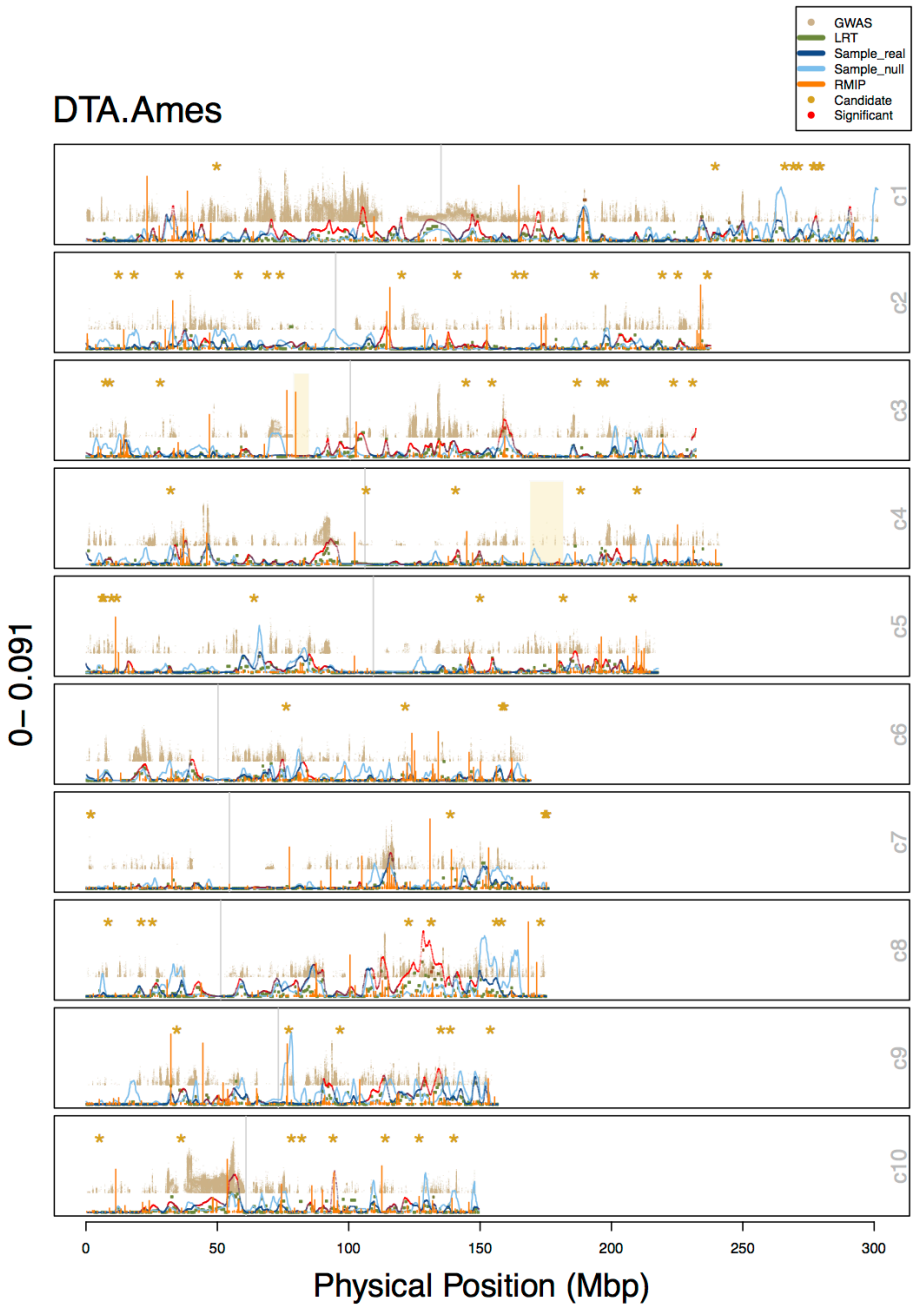


Figure S9. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for Ames, DTA. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

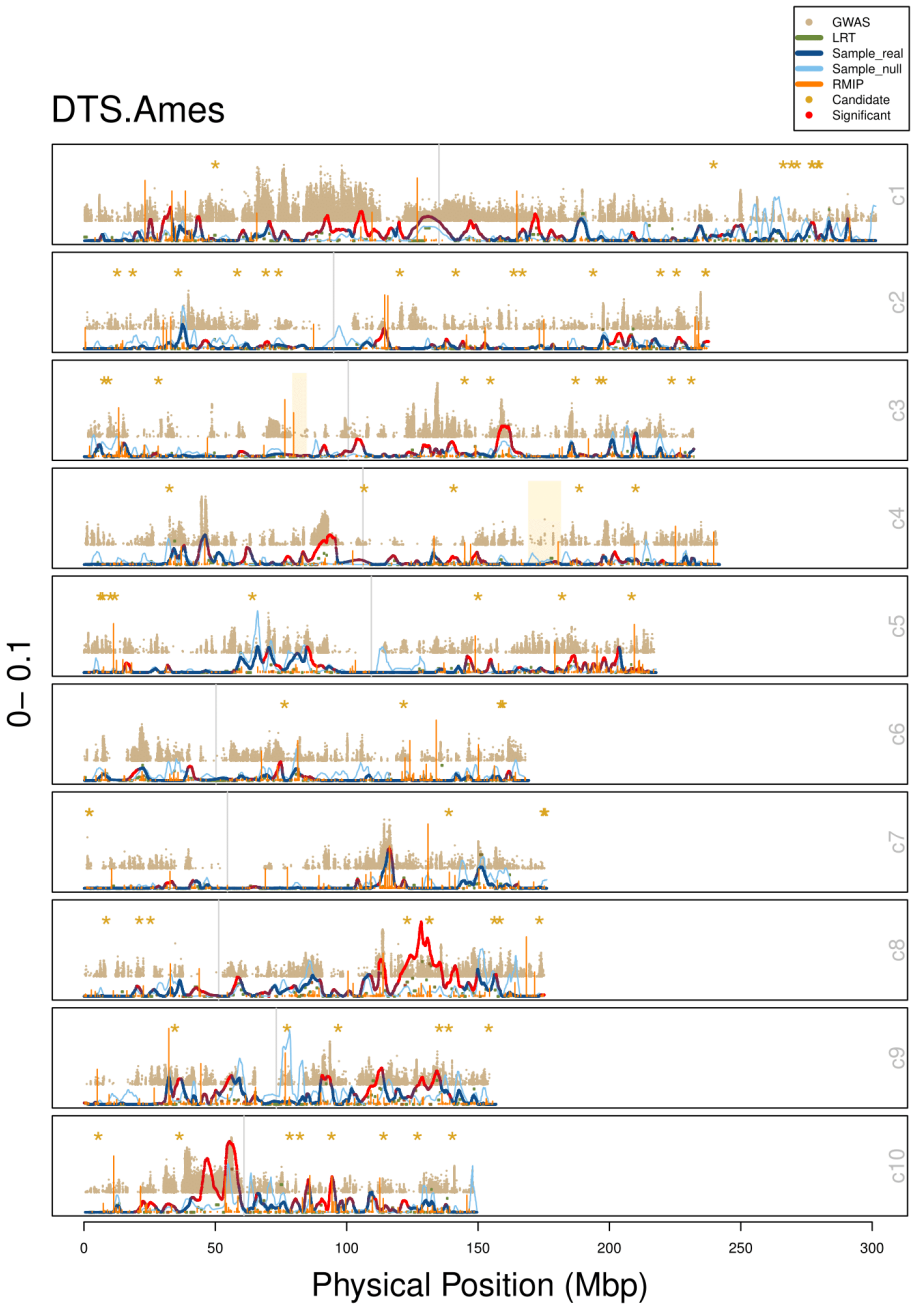


Figure S10. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for Ames, DTS. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

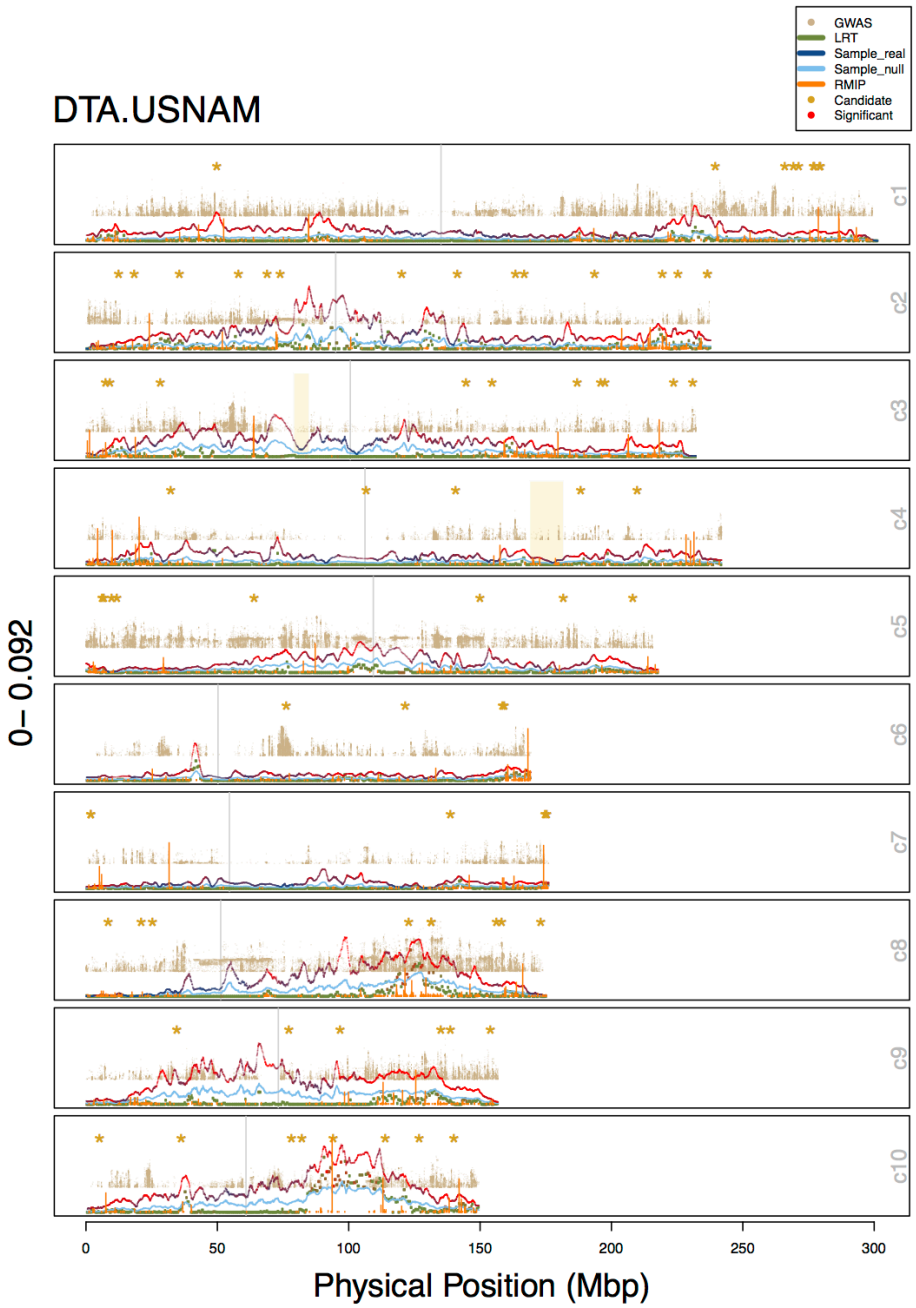


Figure S11. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for USNAM, DTA. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

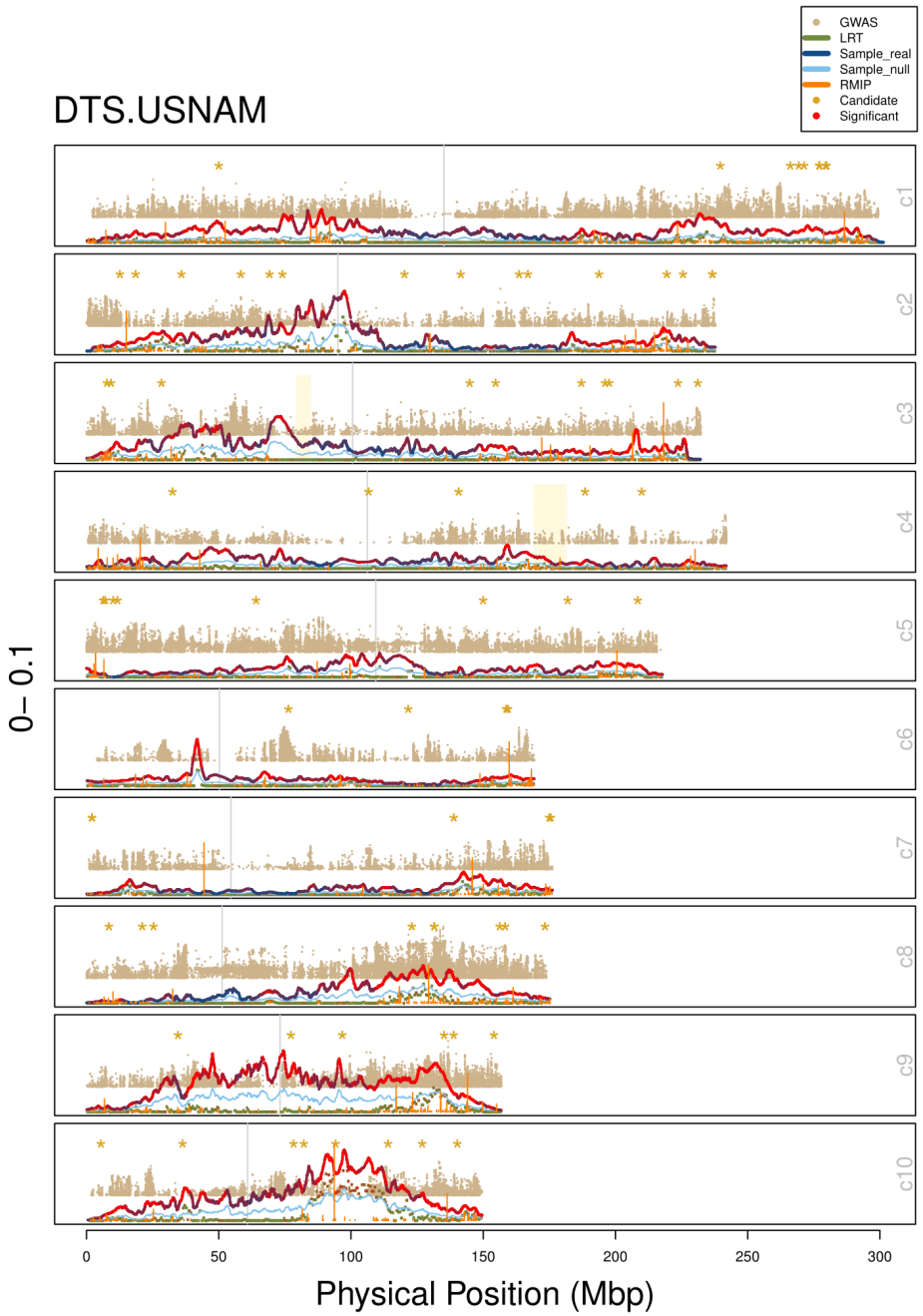


Figure S12. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for USNAM, DTS. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

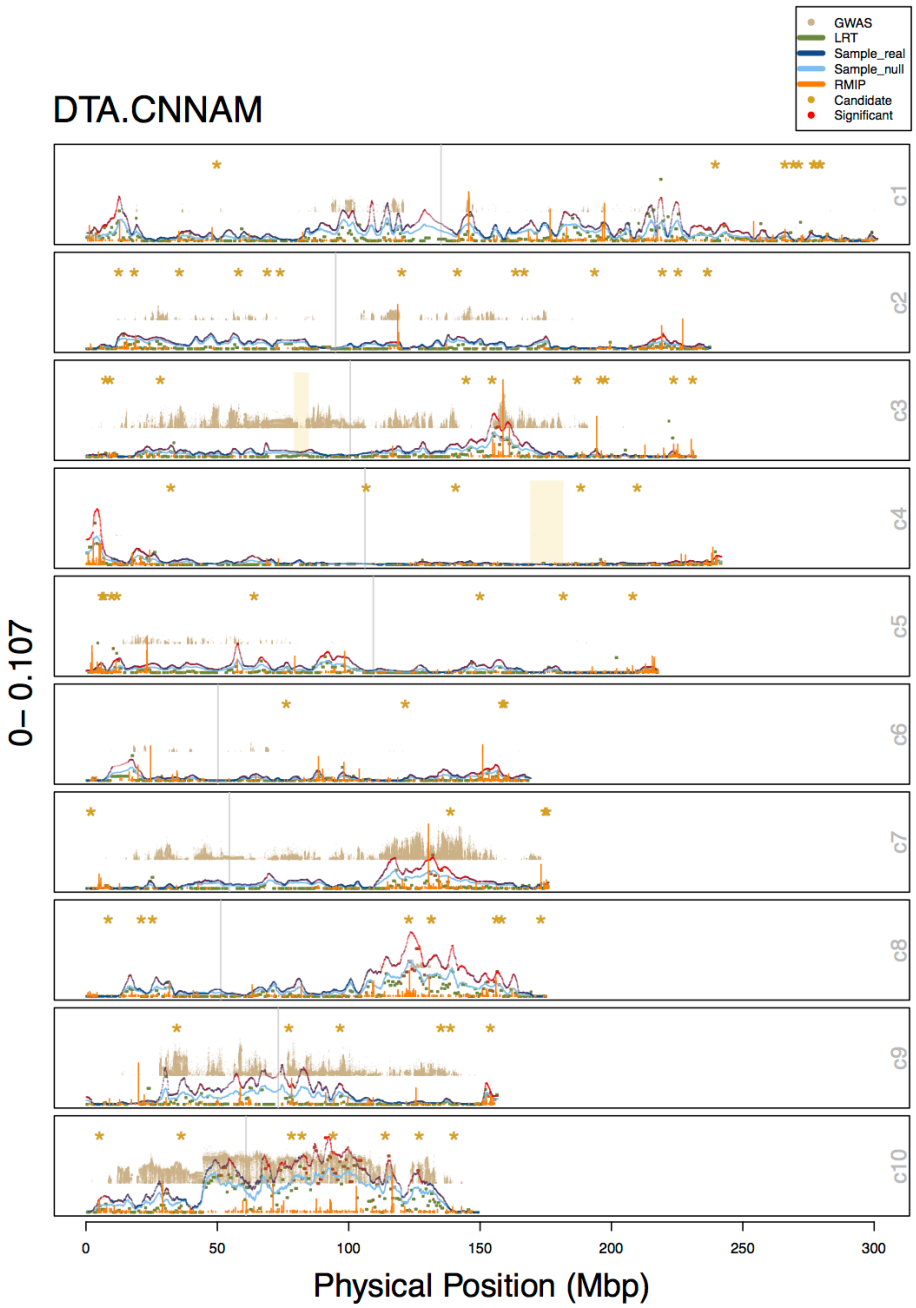


Figure S13. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for CNNAM, DTA. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

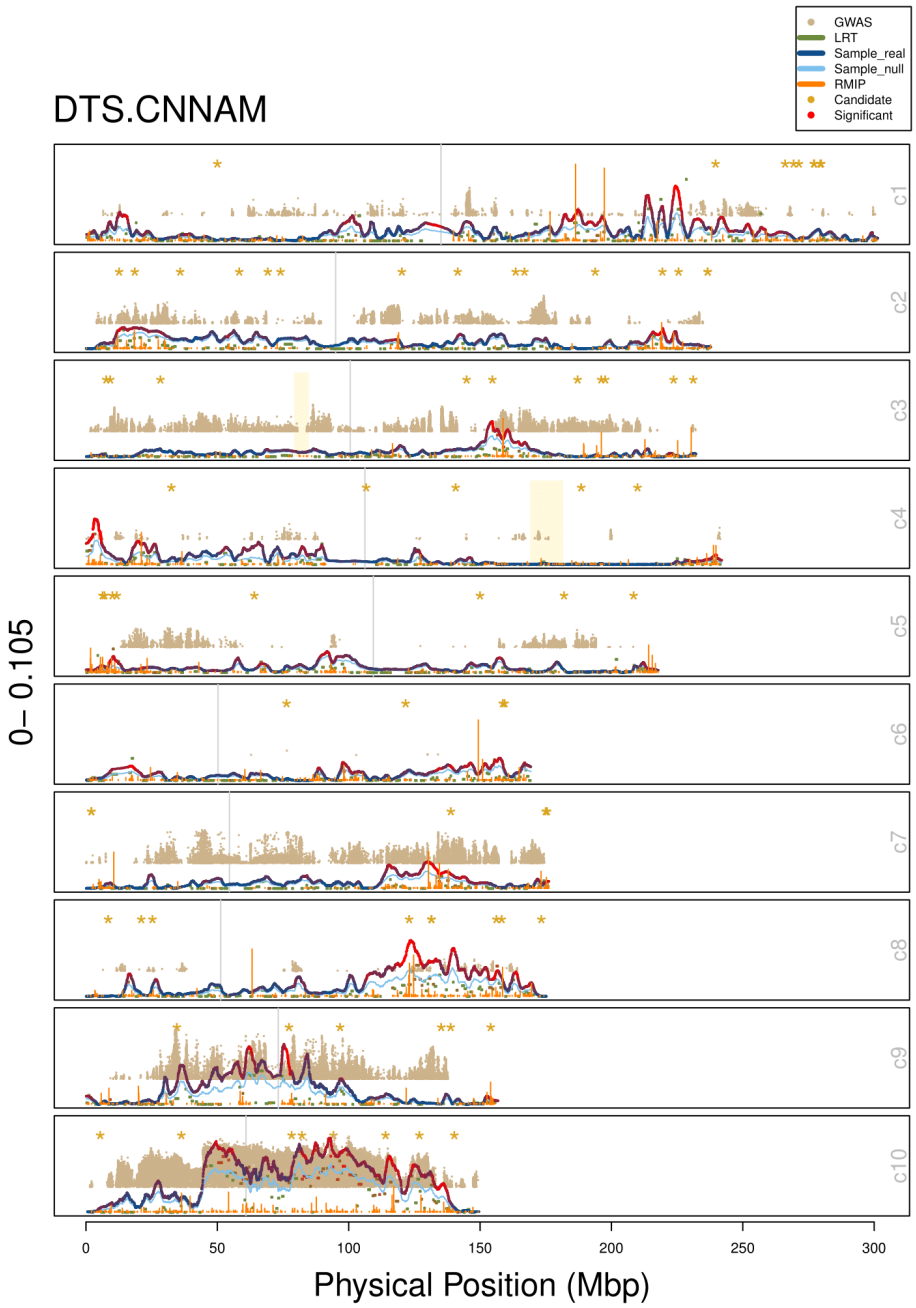


Figure S14. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for CNNAM, DTS. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

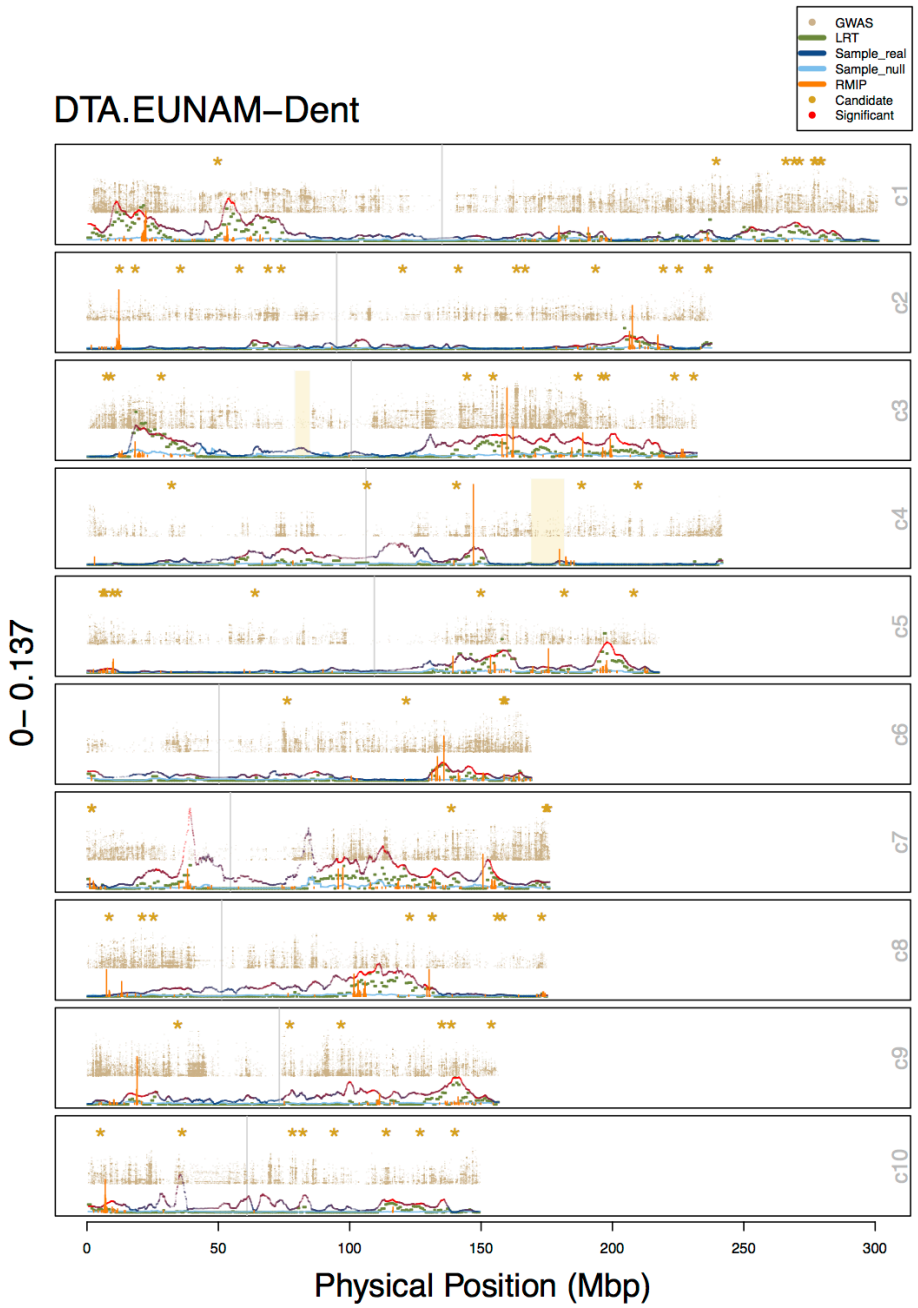


Figure S15. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for EUNAM-Dent, DTA. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

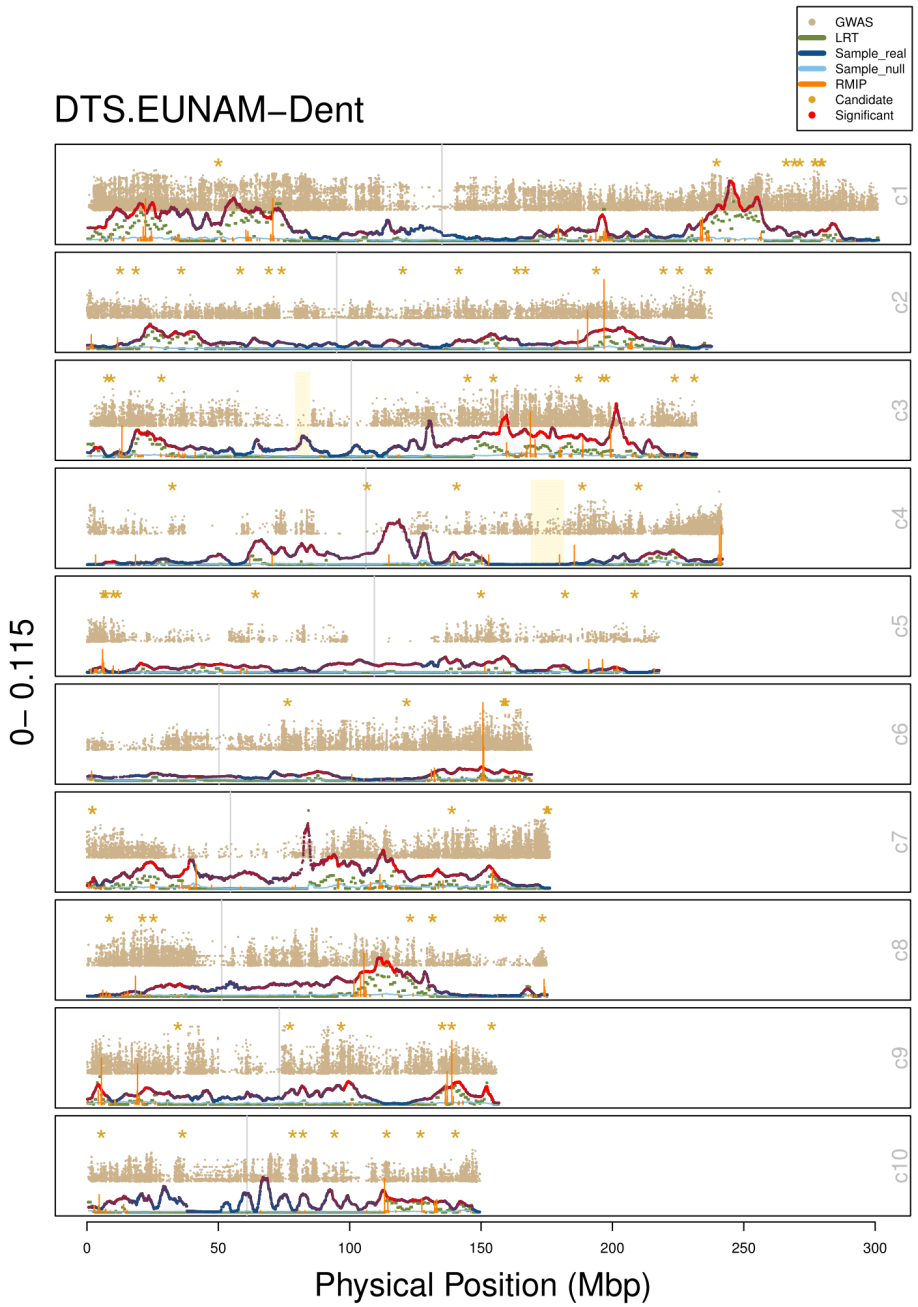


Figure S16. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for EUNAM-Dent, DTS. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

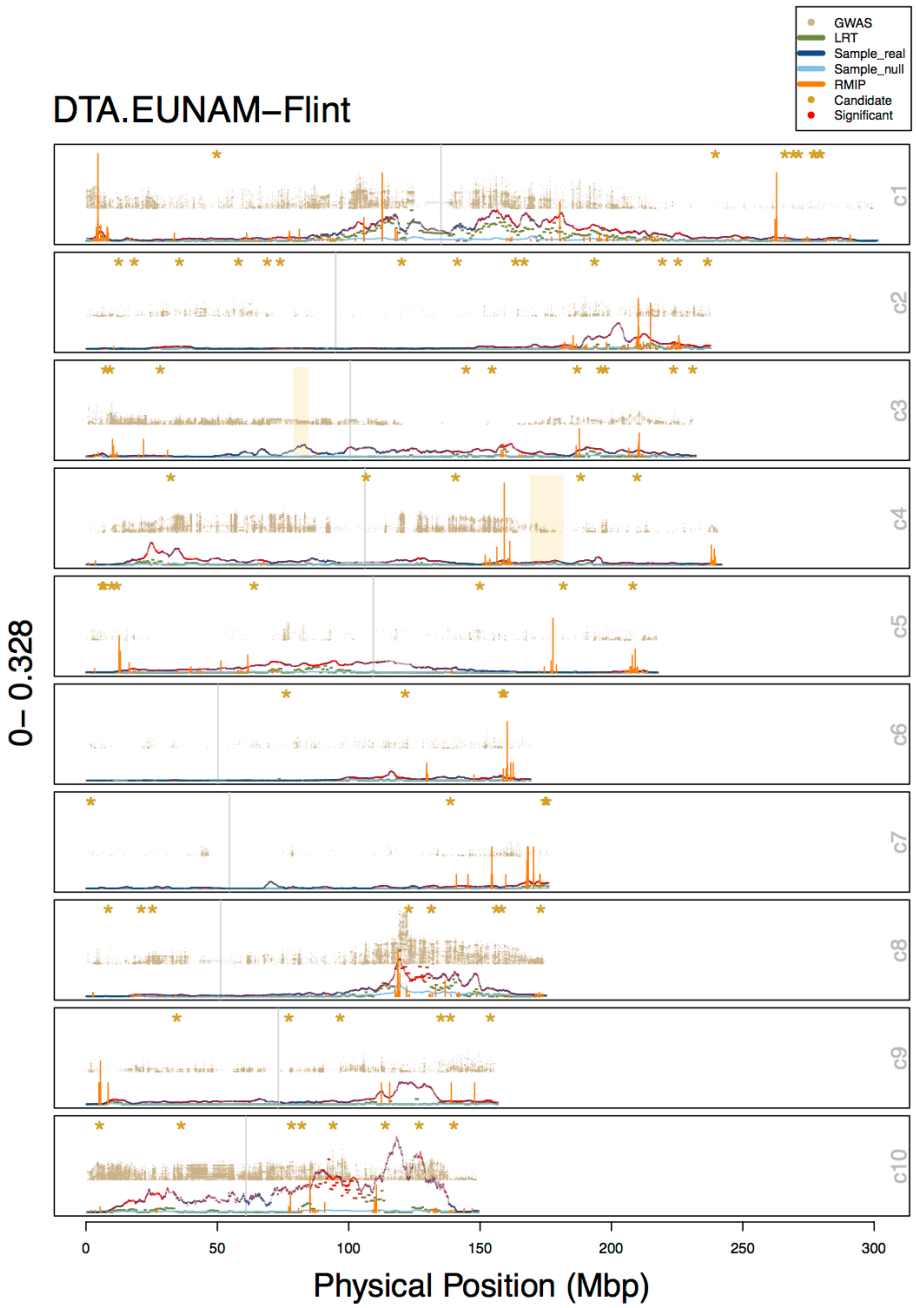


Figure S17. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for EUNAM-Flint, DTA. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

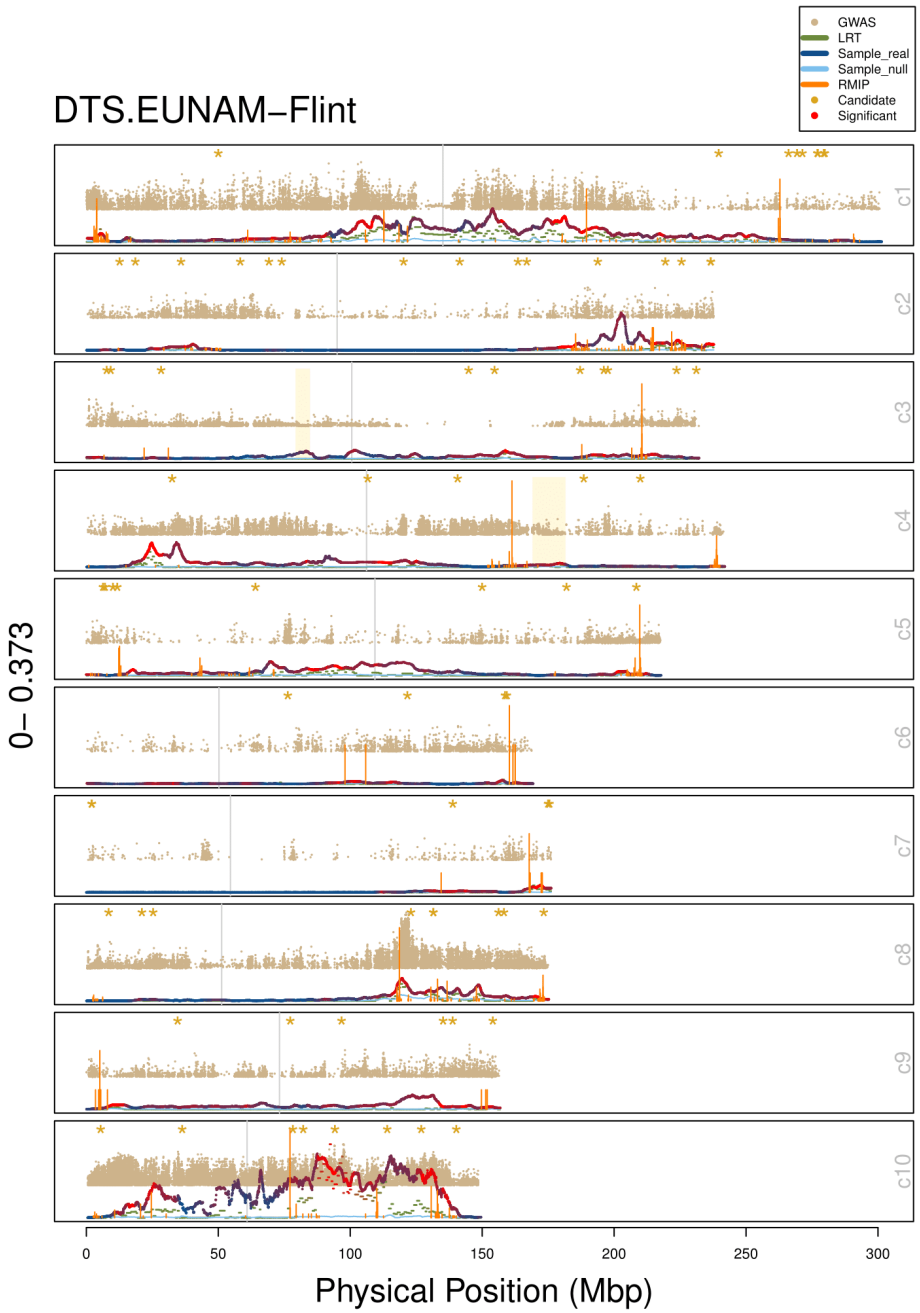


Figure S18. Combined results for all methods (GWAS (uncorrected), resampling GWAS (RMIP), RHM, and BRHM) for EUNAM-Flint, DTS. RHM and BRHM variance estimates are unmodified. GWAS –log10 p-values and the proportion of models for which a SNP was chosen in resampling GWAS rescaled to the maximum value of the variances in the window, and only the top 5,000 SNPs across the genome plotted in a window. Both the median estimated heritability for the null phenotypes (light blue) and the real estimated heritability (blue to red) is shown for BRHM. RMIP bars in orange represent the proportion of models in which that SNP was chosen. Candidate genes (based on Dong et al./ Danilevskya et al) are noted with a yellow asterisk and listed across the top.

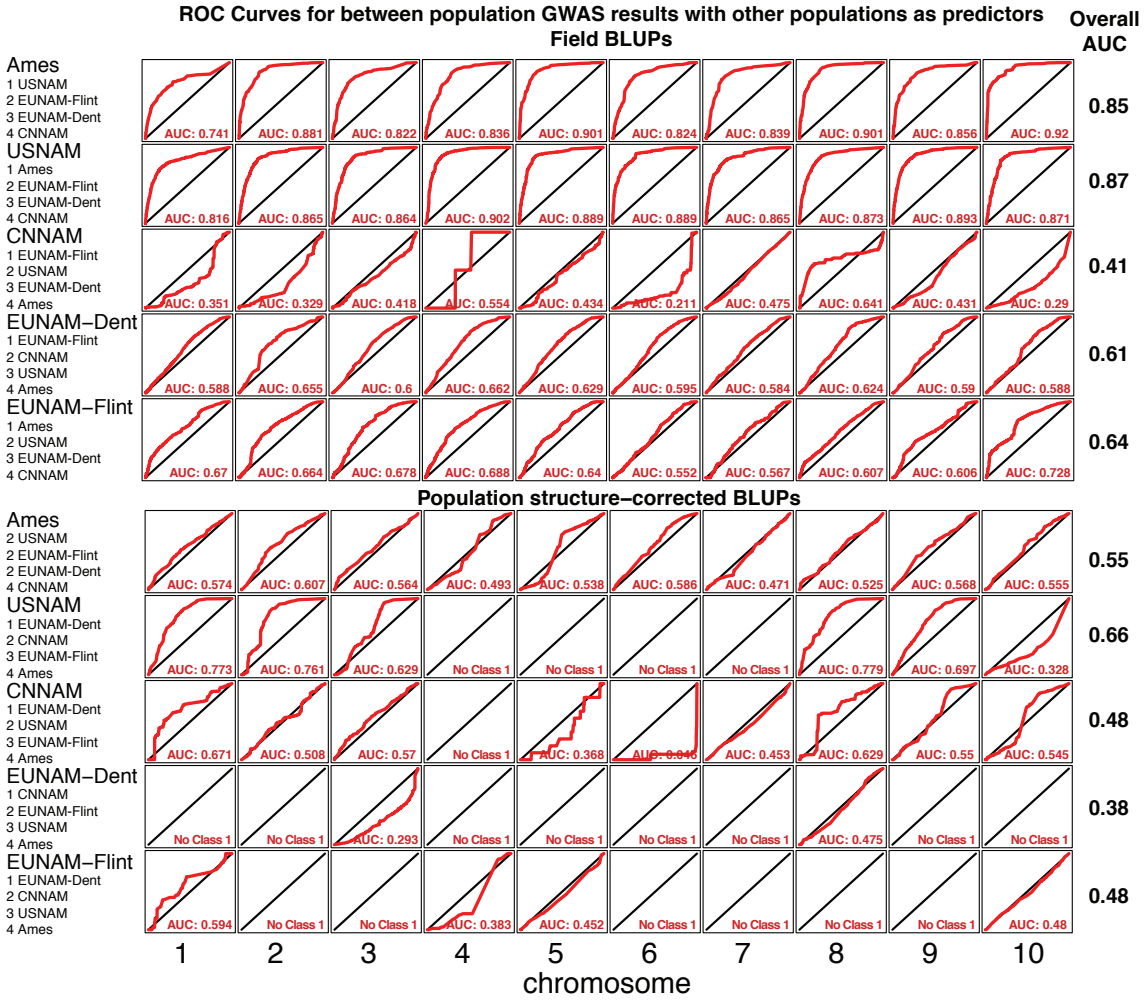


Figure S19. Random forest classifier for GWAS results between populations to evaluate overlap in GWAS results between populations. Predictors are equivalent GWAS (with or without population structure correction) for the other populations. Trained on nine chromosomes and tested on the 10th. Not all chromosomes for the population structure corrected results have top predictors, and are excluded from overall AUC calculations.

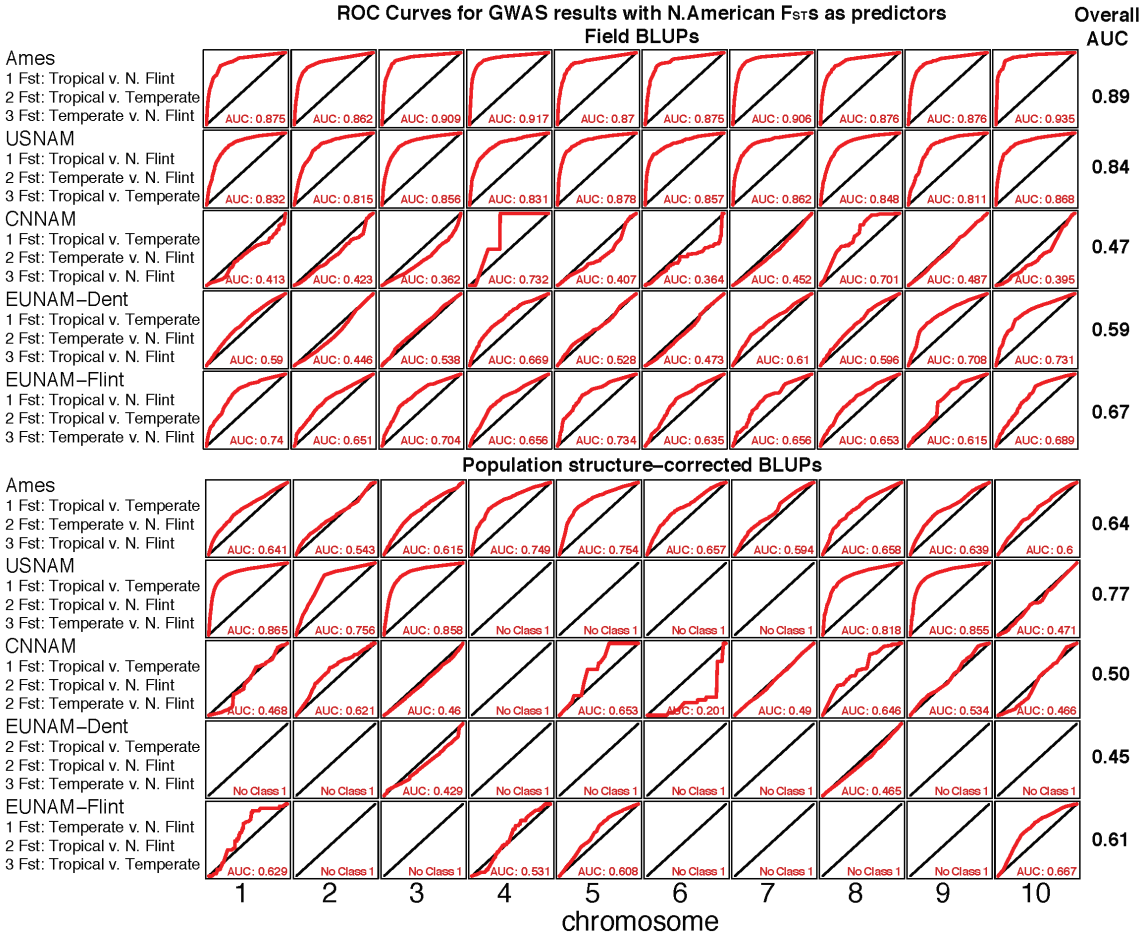


Figure S20. Random forest classifier with GWAS additive p-values as the response, and N. American F_{ST}s as the predictors. Trained on nine chromosomes and tested on the 10th. Not all chromosomes for the population structure corrected results have top predictors, and are excluded from overall AUC calculations.

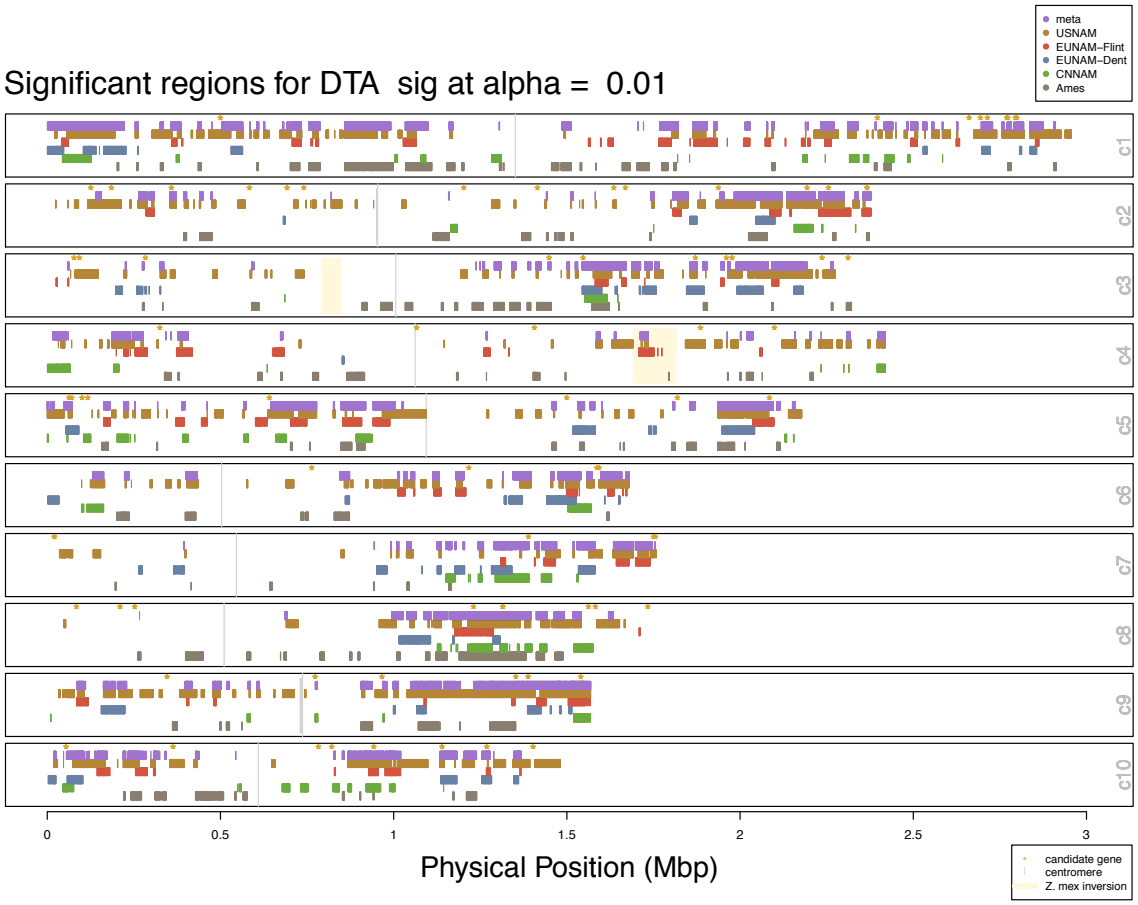


Figure S21. Significant regions and metaanalysis for BRHM results DTA

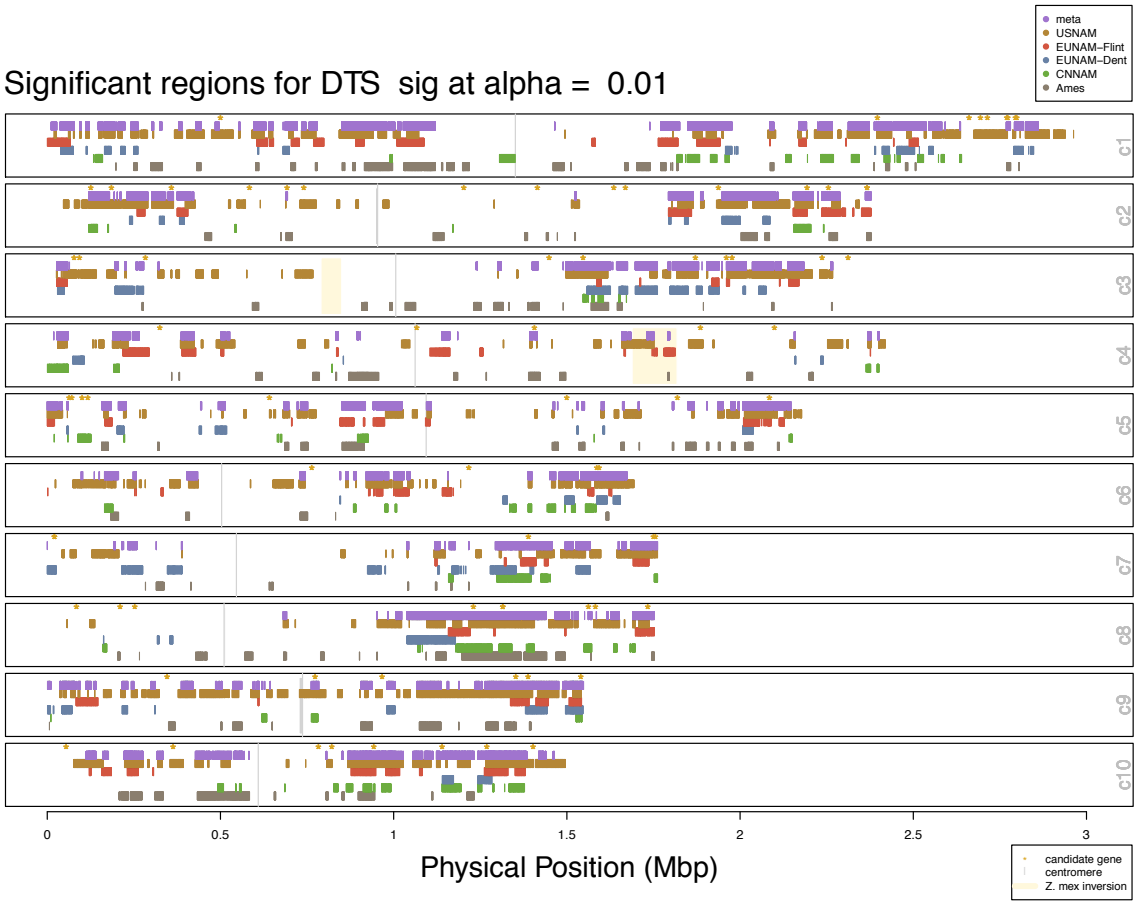
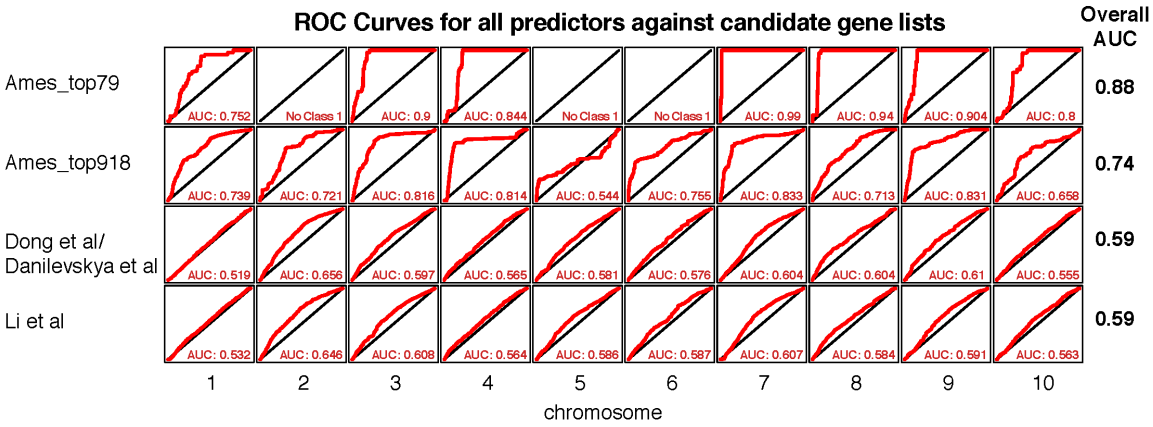


Figure S22. Significant regions and metaanalysis for BRHM results DTS



Top Predictor rankings

Ames_top79	Ames_top918	Dong et al/Danilevskya et al	Li et al
1 Ames BRHM median h2	1 Ames BRHM median h2	1 USNAM BRHM bon-corr p-value	1 USNAM BRHM bon-corr p-value
4 Ames BRHM bon-corr p-value	2 Ames BRHM bon-corr p-value	4 CNNAM BRHM median h2	4 CNNAM BRHM median h2
4 EUNAM-Flint BRHM median h2	4 EUNAM-Flint BRHM median h2	4 Ames coverage	4 Ames coverage
5 USNAM BRHM median h2	4 EUNAM-Dent BRHM median h2	5 USNAM BRHM median h2	5 EUNAM-Flint BRHM median h2
5 CNNAM BRHM median h2	6 CNNAM BRHM median h2	5 EUNAM-Flint BRHM median h2	5 USNAM BRHM median h2
6 EUNAM-Dent BRHM median h2	7 USNAM BRHM median h2	6 CNNAM coverage	6 CNNAM coverage
7 EUNAM-Flint BRHM bon-corr p-value	7 EUNAM-Flint coverage	6 EUNAM-Dent BRHM median h2	7 EUNAM-Dent BRHM median h2
8 EUNAM-Flint coverage	9 EUNAM-Dent BRHM bon-corr p-value	8 EUNAM-Dent coverage	8 EUNAM-Dent coverage
10 CNNAM BRHM bon-corr p-value	9 CNNAM BRHM bon-corr p-value	11 Ames N major allele	10 EUNAM-Dent BRHM bon-corr p-value
11 USNAM BRHM bon-corr p-value	10 Ames uncorrected GLM add-p	11 EUNAM-Flint BRHM bon-corr p-value	11 EUNAM-Flint BRHM bon-corr p-value

Figure S23. Machine learning using candidate genes as the response, and GWAS and BRHM mapping, power estimators, and F_{ST}s as the predictors. Candidate genes were classified as 1, with a 20kbp buffer around the gene start and end position (maize AGP v3). AUCs of 0 result when no candidates are found on the chromosome tested. Not all chromosomes for the Ames top 79 candidates have top predictors, and are excluded from overall AUC calculations.

Supplemental tables

Table S1. Number of populations significant for autonomous flowering time candidates (after Dong et al) based on Bonferroni corrected significance for BRHM method

Gene Model	Locus Name	Chr	Start	End	0.001	0.01	0.05
GRMZM2G381691	ZmCCT	10	94262291	94264845	3	4	4
GRMZM2G171365	ZMM5	9	154082657	154102449	2	3	4
GRMZM2G092174	PhyB2	9	135245567	135251882	2	2	4
GRMZM2G095727	ZmPRR73	9	138861929	138870925	3	3	3
GRMZM2G179264	ZCN8	8	123030387	123032175	3	3	3
GRMZM2G700665	ZmRAP2.7	8	131576889	131580316	3	3	3

MITE	MITE	8	131517973	131520473	3	3	3
GRMZM2G062052	ZCN19	10	126937360	126941217	2	3	3
GRMZM2G033962	ZmPRR37.1	2	219435832	219441286	2	2	3
GRMZM2G427618	GA2ox1	6	158703023	158704867	2	2	3
GRMZM2G047855	ZmHD6	1	279484878	279490883	2	2	2
GRMZM2G059958	LDL1	5	208475652	208478520	2	2	2
GRMZM2G141756	ZCN7	6	159399177	159400713	2	2	2
GRMZM2G143602	ZmHD6	2	236711216	236716685	2	2	2
GRMZM2G148693	ZAP1	2	236669675	236679078	2	2	2
GRMZM2G180190	ZFL2	2	12649206	12652213	2	2	2
GRMZM2G400167	ZCN26	9	77335866	77339429	2	2	2
AC217051.3	ZCN5	10	114050161	114051244	1	2	2
GRMZM2G031724	GA2ox1	8	173402117	173404386	1	2	2
GRMZM2G106613	ZmLD	3	154702548	154712172	1	2	2
GRMZM2G067921	DLF1	7	175583965	175585451	0	2	2
GRMZM2G440005	ZCN24	7	138923902	138933246	0	2	2
GRMZM2G011357	ID1	1	239667869	239671192	1	1	2
GRMZM2G022679	GA2ox1	3	196126846	196128900	1	1	2
GRMZM2G057935	PhyC1	1	277059620	277064623	1	1	2
GRMZM2G145058	ZmTOC1.1	3	187143863	187147433	1	1	2
GRMZM2G032339	ZMM4	1	277218151	277298925	0	1	2
AC214791.2	ZCN20	10	36388693	36389588	1	1	1
GRMZM2G017087	KN1	1	271407490	271415210	1	1	1
GRMZM2G019863	LDL1	2	18623440	18626564	1	1	1
GRMZM2G020081	ZmTOC1.5	4	140726569	140729617	1	1	1
GRMZM2G021560	ZCN25	2	35895307	35898831	1	1	1
GRMZM2G022489	miR156	3	7774552	7776125	1	1	1
GRMZM2G059358	ZCN10	2	69266391	69268157	1	1	1
GRMZM2G067702	LUX	8	156260024	156261399	1	1	1
GRMZM2G072582	ZAP1b	7	2065444	2076981	1	1	1
GRMZM2G103666	ZCN12	3	197706353	197708239	1	1	1
GRMZM2G107945	ZmFKF1b	2	225575240	225579603	1	1	1
GRMZM2G124532	PhyB1	1	50023180	50034523	1	1	1
GRMZM2G132880	ZCN6	4	188561914	188563106	1	1	1
GRMZM2G157727	PhyA1	1	269388274	269395277	1	1	1
GRMZM2G174083	ZmTOC1.3	1	279875580	279878961	1	1	1
GRMZM2G365688	ZmTOC1.4	3	223690797	223697299	1	1	1
GRMZM5G845755	ZmHD6	7	2138862	2144213	1	1	1
GRMZM5G861678	ZmHy2	8	158299004	158303170	1	1	1
GRMZM2G005732	ZmPRR37	7	174996931	175030487	0	1	1
GRMZM2G019993	ZCN21	2	74102888	74107552	0	1	1

GRMZM2G024973	D9	5	11793473	11795945	0	1	1
GRMZM2G145041	ZmCCA1	5	64145458	64148638	0	1	1
GRMZM2G158809	ZCN18	2	193839589	193845094	0	1	1
GRMZM2G181028	PhyA2	5	10111505	10117216	0	1	1
GRMZM2G107101	GIGZ1A	8	21104080	21112365	0	0	1
GRMZM2G108016	ZCN13	5	150036613	150037929	0	0	1
GRMZM2G152689	ZCN10	3	28400356	28402122	0	0	1
GRMZM2G160730	GL15	9	96744684	96748035	0	0	1
GRMZM2G405368	CONZ1	9	34628528	34633066	0	0	1
GRMZM5G844173	GIGZ1B	3	9306265	9317285	0	0	1
GRMZM2G014902	ZmLHY2	4	32527798	32560791	0	0	0
GRMZM2G021614	ZCN9	8	8472799	8475229	0	0	0
GRMZM2G051338	ZCN15	6	76400211	76402023	0	0	0
GRMZM2G065276	FCA	2	163595658	163616052	0	0	0
GRMZM2G066638	ZmTOC1.2	2	166953806	166957187	0	0	0
GRMZM2G075081	ZCN4	2	58373992	58376007	0	0	0
GRMZM2G075215	ZCN17	2	120292154	120293853	0	0	0
GRMZM2G092008	ZCN1	3	231199114	231200263	0	0	0
GRMZM2G106363	ZmFKF1a	4	209980833	209983942	0	0	0
GRMZM2G117057	ZCN11	6	121720043	121721582	0	0	0
GRMZM2G127121	ZCN16	5	181990493	181993244	0	0	0
GRMZM2G129889	PhyC2	5	7129937	7134953	0	0	0
GRMZM2G135446	ZmPRR59	2	141518105	141523326	0	0	0
GRMZM2G141903	ZmHD6	5	6381908	6387449	0	0	0
GRMZM2G144744	D8	1	266160101	266163168	0	0	0
GRMZM2G149786	LDL1	10	82218460	82222873	0	0	0
GRMZM2G156079	ZCN2	4	106696523	106697982	0	0	0
GRMZM2G338454	ZCN3	10	5465315	5465881	0	0	0
GRMZM2G373928	ZCN14	8	25338663	25347919	0	0	0
GRMZM2G474769	ZmLHY1	10	78333991	78337257	0	0	0
GRMZM5G803935	miR172	3	144918011	144918720	0	0	0

Table S2. New candidate genes derived from Ames and combined GWAS analysis, based on genes found within 100kbp of the top GWAS hits sorted by additive p-value. The ranking within each subset is given. The first 79 by rank are used in comparison to the shorter Dong et al/ZCN list and the full 918 by rank for each results list are compared with Li et al.

gene	Ames uncorr	Ames uncorr	Ames corr DTS	Ames corr DTA	Combined	gene	Ames uncorr	Ames uncorr	Ames corr DTS	Ames corr DTA	Combined
GRMZM2G000129	88	67			260	GRMZM2G140082					881
GRMZM2G000209					704	GRMZM2G140083			75	173	96
GRMZM2G000219					705	GRMZM2G140298	151	213			44
GRMZM2G000326			334	507		GRMZM2G140448					477
GRMZM2G000353	600	474				GRMZM2G140500				483	
GRMZM2G000376	599	473				GRMZM2G140755			518	458	
GRMZM2G000380			407	189		GRMZM2G140758			515	455	
GRMZM2G000471			1	11	309	GRMZM2G140811			514	454	
GRMZM2G000481			2	12	310	GRMZM2G140824			674	612	
GRMZM2G000632	627	548				GRMZM2G141241			797		
GRMZM2G000781	321	326			251	GRMZM2G141256			796		
GRMZM2G000801	320	325			250	GRMZM2G141288			795		
GRMZM2G001004			715	244		GRMZM2G141355			273	94	
GRMZM2G001139					325	GRMZM2G141383			943		
GRMZM2G001145	635	612				GRMZM2G141605			891		
GRMZM2G001205			393		163	GRMZM2G141618			894		
GRMZM2G001243			481	250		GRMZM2G141665					872
GRMZM2G001247					893	GRMZM2G141667			922		
GRMZM2G001297					894	GRMZM2G141723					871
GRMZM2G001541	937	905				GRMZM2G141784	226	168			
GRMZM2G001645					161	GRMZM2G142043				951	
GRMZM2G001648					791	GRMZM2G142334					891
GRMZM2G001816					54	GRMZM2G142352					191
GRMZM2G001895	14	10				GRMZM2G142366				818	
GRMZM2G001930	12	9				GRMZM2G142390					192
GRMZM2G002115	496	545				GRMZM2G142507					959
GRMZM2G002149			135	114		GRMZM2G142565					816
GRMZM2G002173			104	90		GRMZM2G142597					961
GRMZM2G002199			790	647		GRMZM2G142620					815
GRMZM2G002240			965	646		GRMZM2G142649					12
GRMZM2G002297	732	571				GRMZM2G142660					14
GRMZM2G002559			175	725		GRMZM2G142667	837	662			

GRMZM2G002626			547	481		GRMZM2G142718	731	663			
GRMZM2G003096			704			GRMZM2G142743	726	664			
GRMZM2G003186			605	903		GRMZM2G142757	727	665			
GRMZM2G003318	311	293				GRMZM2G142802	728	666			
GRMZM2G003488			900			GRMZM2G142820	730	668			
GRMZM2G003530	452	268				GRMZM2G142922	758				
GRMZM2G003595			124	362		GRMZM2G142927	759				
GRMZM2G007475					832	GRMZM2G142932	760				
GRMZM2G007514					831	GRMZM2G143025	180	170			385
GRMZM2G007530			974	574		GRMZM2G143075	92	117			383
GRMZM2G007587			973	573		GRMZM2G143082	93	118			384
GRMZM2G003607	906	846				GRMZM2G143202	666	685			
GRMZM2G003631			264	594		GRMZM2G143210	251	270			1
GRMZM2G003640	907	847				GRMZM2G143473			810		
GRMZM2G003734				652		GRMZM2G143747		910	374		
GRMZM2G003755			937			GRMZM2G144088			336	188	
GRMZM2G003979	68	58				GRMZM2G144155	505	285			152
GRMZM2G004298	199	150				GRMZM2G144173					297
GRMZM2G004500					502	GRMZM2G144398				430	22
GRMZM2G004511					501	GRMZM2G144451			607		
GRMZM2G004732	163	244				GRMZM2G144480			608		
GRMZM2G004898			549	504	584	GRMZM2G144742					377
GRMZM2G004959	935					GRMZM2G144821		976			
GRMZM2G005163				696	132	GRMZM2G144843		975			
GRMZM2G005205	756	623				GRMZM2G144890		973			
GRMZM2G005253				517		GRMZM2G145012					921
GRMZM2G005350				695	131	GRMZM2G145027			879	381	
GRMZM2G005353			298	689		GRMZM2G145054			878	380	
GRMZM2G005419		978				GRMZM2G145104			415	502	
GRMZM2G005480	285	353			104	GRMZM2G145346				900	
GRMZM2G005499			561			GRMZM2G145374				899	
GRMZM2G005552				694	130	GRMZM2G145390				897	
GRMZM2G005680	508					GRMZM2G145412			87	755	
GRMZM2G005710				693	129	GRMZM2G145458			88	756	
GRMZM2G005732			325	182		GRMZM2G145573					766
GRMZM2G005771			927	678		GRMZM2G145579			68	60	
GRMZM2G006121					339	GRMZM2G145752	955				
GRMZM2G006429					199	GRMZM2G145951				912	
GRMZM2G006672				883		GRMZM2G145972				911	
GRMZM2G006704					976	GRMZM2G146108					689

GRMZM2G006714	735	875			200	GRMZM2G146143					691
GRMZM2G006721					671	GRMZM2G146449			925		
GRMZM2G006727			820			GRMZM2G147158			688		
GRMZM2G006830	738	878			203	GRMZM2G147475					259
GRMZM2G006884	737	877			202	GRMZM2G147534					327
GRMZM2G006953				991		GRMZM2G147671					328
GRMZM2G006977			404			GRMZM2G147683	335	392			
GRMZM2G007130			403			GRMZM2G147712				592	
GRMZM2G007288				691		GRMZM2G147724					731
GRMZM2G007404		922				GRMZM2G147756			740	329	
GRMZM2G007590					830	GRMZM2G147775					329
GRMZM2G007630					115	GRMZM2G147787					330
GRMZM2G007633				981		GRMZM2G147819	794	605			
GRMZM2G007848					114	GRMZM2G147891	793	604			
GRMZM2G007953				865		GRMZM2G147908	795	606			
GRMZM2G007981			972	572		GRMZM2G147937	157	174			
GRMZM2G008273			853	570		GRMZM2G148074	205	406			66
GRMZM2G008287				792		GRMZM2G148355				552	
GRMZM2G008316			946			GRMZM2G148400			421	558	
GRMZM2G008482			852	569		GRMZM2G148485					563
GRMZM2G008583				793		GRMZM2G148561			420	197	
GRMZM2G008622				791		GRMZM2G148695					920
GRMZM2G008714				864		GRMZM2G148867	805				
GRMZM2G008751					672	GRMZM2G148896	804				
GRMZM2G009019				733		GRMZM2G149038			163	850	
GRMZM2G009114					888	GRMZM2G149153				732	
GRMZM2G009214			391	617		GRMZM2G149211	79	83			257
GRMZM2G009344				857		GRMZM2G149216	136	114			
GRMZM2G009365					515	GRMZM2G149224	437	558			56
GRMZM2G009435	193	192			138	GRMZM2G149236	135	113			
GRMZM2G009571			528			GRMZM2G149286	78	82			256
GRMZM2G009724					847	GRMZM2G149681					935
GRMZM2G009876			529			GRMZM2G149903	258	490			429
GRMZM2G009888			503	350		GRMZM2G150024			577	500	
GRMZM2G010017					908	GRMZM2G150474					397
GRMZM2G010095	82	75				GRMZM2G150485	812	688			
GRMZM2G010136			530			GRMZM2G150688	557	580			
GRMZM2G010338			139	475		GRMZM2G150772	556	579			
GRMZM2G010406	60	61				GRMZM2G151041			105	663	
GRMZM2G010490					909	GRMZM2G151230				491	

GRMZM2G010518				723		GRMZM2G151319			314	438	
GRMZM2G010628	182	154				GRMZM2G151496	617				917
GRMZM2G010640	351	312				GRMZM2G151519			351	235	
GRMZM2G010714	352	313				GRMZM2G151542	606	671			
GRMZM2G011071					468	GRMZM2G151656		774			840
GRMZM2G011422			632	473		GRMZM2G151668		859			
GRMZM2G011491					363	GRMZM2G151706					639
GRMZM2G011518	192	203				GRMZM2G151717		673			
GRMZM2G011590				598		GRMZM2G151992	763	837			
GRMZM2G011592			313	437		GRMZM2G152051			698		
GRMZM2G011622					282	GRMZM2G152417		902			
GRMZM2G011998			161	773		GRMZM2G152549			276		
GRMZM2G012041					280	GRMZM2G152573			457	328	635
GRMZM2G012176	59	60				GRMZM2G152631			35	32	215
GRMZM2G012306	191	202				GRMZM2G152655			36	33	216
GRMZM2G012342			802		279	GRMZM2G152661					570
GRMZM2G012412				549		GRMZM2G152663			277		
GRMZM2G012434				927		GRMZM2G152708					517
GRMZM2G012546	254	367				GRMZM2G152764					470
GRMZM2G012631	563	442				GRMZM2G152777	896	841			
GRMZM2G012756	50	53				GRMZM2G152808	897	842			
GRMZM2G012874					304	GRMZM2G152877	165	138			348
GRMZM2G012932					43	GRMZM2G152925					561
GRMZM2G013002					656	GRMZM2G152975					918
GRMZM2G013079			683	298		GRMZM2G152981			321	263	
GRMZM2G013100			423	333		GRMZM2G153240	969	814			
GRMZM2G013324	540	697				GRMZM2G153359	581	516			
GRMZM2G013331	371	498				GRMZM2G153368	580	515			
GRMZM2G013650					657	GRMZM2G153438	879	734			
GRMZM2G013671					943	GRMZM2G153454	878	684			
GRMZM2G013814			962			GRMZM2G153488	579	514			
GRMZM2G013848			769			GRMZM2G153552	903				
GRMZM2G013981			763			GRMZM2G154114	734	855			
GRMZM2G013987	447	613			92	GRMZM2G154165			606		
GRMZM2G014089					270	GRMZM2G154169	957				
GRMZM2G014233			729	415		GRMZM2G154216					519
GRMZM2G014356			408	168		GRMZM2G154290	100	103			87
GRMZM2G014444			643	448		GRMZM2G154366	959				
GRMZM2G014705		883				GRMZM2G154394	960				
GRMZM2G014839					368	GRMZM2G154414	961				

GRMZM2G014902	683					GRMZM2G154487	56	73			50
GRMZM2G015190			821	465		GRMZM2G154549	323	283			
GRMZM2G015568			520			GRMZM2G154565	55	72			49
GRMZM2G015610			693			GRMZM2G154578			81	41	
GRMZM2G015844			692			GRMZM2G154580	54	71			2
GRMZM2G015967		777				GRMZM2G154830	939				
GRMZM2G015972				873		GRMZM2G155252	806	557			
GRMZM2G016020	643	748				GRMZM2G155274					378
GRMZM2G016275			897	630		GRMZM2G155380	558	578			571
GRMZM2G016318			652	252		GRMZM2G155491			699	787	
GRMZM2G016393			806	423		GRMZM2G155675			192	355	
GRMZM2G016734	820	891				GRMZM2G155816				975	
GRMZM2G016890	553	757				GRMZM2G155931			59	20	
GRMZM2G017011	741	881				GRMZM2G155974	238	161			170
GRMZM2G017044	821	892				GRMZM2G155998	240	160			169
GRMZM2G017142					623	GRMZM2G156013	400	531			
GRMZM2G017321			65	22		GRMZM2G156105				790	
GRMZM2G017400	641	746				GRMZM2G156145	399	530			
GRMZM2G017405	424	519			620	GRMZM2G156174	398	529			
GRMZM2G017486					619	GRMZM2G156392	636	483	977		23
GRMZM2G017533			186	96		GRMZM2G156451	637	484	978		24
GRMZM2G017536	640	745				GRMZM2G156470	638	485	979		25
GRMZM2G017671				982		GRMZM2G156803			593	233	
GRMZM2G018059					182	GRMZM2G157016			61	108	
GRMZM2G018223					183	GRMZM2G157019			60	107	
GRMZM2G018372		756				GRMZM2G157127					450
GRMZM2G018593		755				GRMZM2G157164	317	470			
GRMZM2G018619					742	GRMZM2G157207			84		
GRMZM2G018770			564		680	GRMZM2G157252				565	777
GRMZM2G018782				462		GRMZM2G157306			185	217	10
GRMZM2G019090		971				GRMZM2G157343			611	302	
GRMZM2G019358			563		679	GRMZM2G157357			612	303	
GRMZM2G019468			905	932		GRMZM2G157456	365	263			
GRMZM2G019562	222	258				GRMZM2G157458				566	
GRMZM2G019971				946		GRMZM2G157470	364	262			
GRMZM2G020081				622		GRMZM2G157677			521	405	
GRMZM2G020131	777	831				GRMZM2G157722			202	169	
GRMZM2G020196			562		678	GRMZM2G157824	294	238			18
GRMZM2G020302			201	176	176	GRMZM2G157878	343	298			20
GRMZM2G020468	183	286				GRMZM2G157925			869	294	

GRMZM2G020523	653	678				GRMZM2G158328	918				
GRMZM2G020610	717	744				GRMZM2G158616	523	508			
GRMZM2G020620					812	GRMZM2G158629					126
GRMZM2G020676	184	287				GRMZM2G158679					127
GRMZM2G020721	185	288				GRMZM2G158682	663	563			303
GRMZM2G020828			71	65		GRMZM2G158998					980
GRMZM2G021225			16	28		GRMZM2G159016					979
GRMZM2G021243			344	532		GRMZM2G159028					978
GRMZM2G021256			679	288		GRMZM2G159049				672	
GRMZM2G021378	85	38			94	GRMZM2G159145					356
GRMZM2G021410			343	531		GRMZM2G159285					934
GRMZM2G021427	742				26	GRMZM2G159475		888			
GRMZM2G021471	53	24			46	GRMZM2G159691				722	
GRMZM2G021621	124	251			100	GRMZM2G159700				637	
GRMZM2G021879			760	749	481	GRMZM2G159811			711		
GRMZM2G022061	107	149				GRMZM2G159849			532	516	
GRMZM2G022095	268	317				GRMZM2G160174					733
GRMZM2G022120	231	184			627	GRMZM2G160178	119	88			508
GRMZM2G022634					590	GRMZM2G160454	221	171			
GRMZM2G022694			270	248		GRMZM2G160609			329	223	
GRMZM2G022740			812			GRMZM2G160730			454	498	
GRMZM2G022915		816				GRMZM2G160738	15	17			
GRMZM2G023003				786		GRMZM2G160770	16	18			
GRMZM2G023325	3	4				GRMZM2G160862	47	25			47
GRMZM2G023811			53	98	621	GRMZM2G160990					82
GRMZM2G024038			269	247		GRMZM2G161087			856	845	
GRMZM2G024109			483	307		GRMZM2G161293					845
GRMZM2G024641		931				GRMZM2G161459	26	47			51
GRMZM2G024811					380	GRMZM2G161787		823			
GRMZM2G024823	350	277			559	GRMZM2G162356	360	401			549
GRMZM2G024838			39	178	57	GRMZM2G162388			331	214	
GRMZM2G025026			262	511		GRMZM2G162426			495	657	461
GRMZM2G025157	236	194			668	GRMZM2G162451			494	656	460
GRMZM2G025175		949				GRMZM2G162460			493	655	459
GRMZM2G025303	235	193			667	GRMZM2G162467			492	654	458
GRMZM2G025396	783	907				GRMZM2G162486			491	653	457
GRMZM2G025458	664	566			121	GRMZM2G162497	375	340			
GRMZM2G025480			742			GRMZM2G162537	303	405			80
GRMZM2G025579	381	437				GRMZM2G162623			886		
GRMZM2G025640	825	766				GRMZM2G162637			207	125	

GRMZM2G025748					283	GRMZM2G162670			208	126	
GRMZM2G025895	675	539				GRMZM2G162917		961			
GRMZM2G025903			550	505	809	GRMZM2G163045	909	781			
GRMZM2G025954					419	GRMZM2G163095	911	783			
GRMZM2G026098	243	201			774	GRMZM2G163189	207	374			30
GRMZM2G026135	339	255				GRMZM2G163193				958	
GRMZM2G026147	179	211				GRMZM2G163233	822	787			
GRMZM2G026447	510	475				GRMZM2G163297	823	788			
GRMZM2G026490				989		GRMZM2G163307	824	789			
GRMZM2G026641	549	408				GRMZM2G163398				707	
GRMZM2G026767	288	275				GRMZM2G163519			527		
GRMZM2G026900				953		GRMZM2G163798	273	241			
GRMZM2G027472			34	31		GRMZM2G163841	947	803			173
GRMZM2G027476			95	46		GRMZM2G163843	946	802			172
GRMZM2G027551			603			GRMZM2G163848	945	801			171
GRMZM2G027563		927				GRMZM2G163905			540	806	
GRMZM2G027571				937		GRMZM2G164263			800	925	
GRMZM2G027594					754	GRMZM2G164696					582
GRMZM2G027673			268	196		GRMZM2G164735			426	192	
GRMZM2G027972	771					GRMZM2G164759			425	191	
GRMZM2G028033	684	815				GRMZM2G165011	518				
GRMZM2G028089	770					GRMZM2G165308	465	924			366
GRMZM2G028114					466	GRMZM2G165351			629	668	
GRMZM2G028139			229	522		GRMZM2G165387			587	162	781
GRMZM2G028258			228	521		GRMZM2G165506	900				509
GRMZM2G028306	41	46				GRMZM2G165511	687	729			
GRMZM2G028386	357	790				GRMZM2G165567			434	929	
GRMZM2G028389	888				48	GRMZM2G165734				969	
GRMZM2G028543			672	863		GRMZM2G165778			545	716	
GRMZM2G028608			226	519		GRMZM2G165867					252
GRMZM2G028640			225	518		GRMZM2G165930					393
GRMZM2G028665					407	GRMZM2G166537	363	358			
GRMZM2G028745			923			GRMZM2G166572	362	357			
GRMZM2G028766			267	195		GRMZM2G166671			367	266	
GRMZM2G028834		864				GRMZM2G166686			604		
GRMZM2G029001					710	GRMZM2G166701			368	267	
GRMZM2G029048	661	732				GRMZM2G166711			992	888	
GRMZM2G029151			18	6		GRMZM2G166745			371	270	
GRMZM2G029184			578	255		GRMZM2G166753			370	269	
GRMZM2G029223	395	511				GRMZM2G166889			907		

GRMZM2G029300	173	148				GRMZM2G166979	289	163			564
GRMZM2G029356					711	GRMZM2G167088	916	818			
GRMZM2G029518			737			GRMZM2G167613			110	122	
GRMZM2G029519					985	GRMZM2G167658			111	123	
GRMZM2G029527					591	GRMZM2G167941					587
GRMZM2G029536			197	408		GRMZM2G168299					673
GRMZM2G029690					721	GRMZM2G168392			785		
GRMZM2G030128			871	427		GRMZM2G168849	471	463			748
GRMZM2G030426					713	GRMZM2G168886	533	576			
GRMZM2G030436					529	GRMZM2G168890					824
GRMZM2G030688	944					GRMZM2G168893	470	462			747
GRMZM2G030717	943					GRMZM2G168920					823
GRMZM2G030831			117	79		GRMZM2G168987	87	119			
GRMZM2G031331			253	431		GRMZM2G169064	201	212			
GRMZM2G031394			935			GRMZM2G169080	175	176			
GRMZM2G031447			934			GRMZM2G169149		938			
GRMZM2G031560			753			GRMZM2G169167	509	496			
GRMZM2G031591	304	274				GRMZM2G169173	624	632			
GRMZM2G031607	891	869				GRMZM2G169270					746
GRMZM2G031724	893	871				GRMZM2G169301			251	397	
GRMZM2G031941					473	GRMZM2G169593			968		
GRMZM2G031954			460	285		GRMZM2G169726			129	338	
GRMZM2G032024					758	GRMZM2G169734				876	
GRMZM2G032821	61	62				GRMZM2G169812	94	74			
GRMZM2G032847			656	763		GRMZM2G169848			794	383	
GRMZM2G033175					950	GRMZM2G169931	181	242			558
GRMZM2G033228	611	399				GRMZM2G170201			346		
GRMZM2G033236	608	396				GRMZM2G170232			303	741	
GRMZM2G033406			655	762		GRMZM2G170262			304	742	
GRMZM2G033478	282	292			8	GRMZM2G170628			557		
GRMZM2G033489					136	GRMZM2G170632			558		
GRMZM2G033519	234	425				GRMZM2G170727			959		
GRMZM2G033612					137	GRMZM2G170766					851
GRMZM2G033653			233	748		GRMZM2G170798					277
GRMZM2G033665				797		GRMZM2G170826			112	171	
GRMZM2G033767			261	840		GRMZM2G170842			113	172	
GRMZM2G033785					448	GRMZM2G170969	450	538			
GRMZM2G033820					449	GRMZM2G170975	448	536			
GRMZM2G033871					333	GRMZM2G171028	376	341			
GRMZM2G033876	493	362			538	GRMZM2G171031					258

GRMZM2G034015					869	GRMZM2G171078	377	342			
GRMZM2G034083			149	286		GRMZM2G171080			32	26	
GRMZM2G034511			150	287		GRMZM2G171367			156	110	
GRMZM2G034551					688	GRMZM2G171507	159	134			
GRMZM2G034563					198	GRMZM2G171600	242	110		868	34
GRMZM2G034623			750			GRMZM2G171616	262	243			
GRMZM2G034645					196	GRMZM2G171622	105	40	5	3	15
GRMZM2G034690					195	GRMZM2G171650	106	41	6	4	16
GRMZM2G034697				703		GRMZM2G171702			826		
GRMZM2G034727	89	91				GRMZM2G171781			882		167
GRMZM2G034797				705		GRMZM2G172081					808
GRMZM2G034804				704		GRMZM2G172139					931
GRMZM2G035134					906	GRMZM2G172153					932
GRMZM2G035189			488	460		GRMZM2G172581					922
GRMZM2G035285			781	809	418	GRMZM2G172794	402	543			17
GRMZM2G035444	561	624				GRMZM2G172834			119	77	
GRMZM2G035506			474	271		GRMZM2G173030	123	198			
GRMZM2G035557	257	235				GRMZM2G173137					308
GRMZM2G035563			362	207		GRMZM2G173198					553
GRMZM2G035701	530	552				GRMZM2G173209	146	124			
GRMZM2G035826	789					GRMZM2G173344				817	
GRMZM2G035839					435	GRMZM2G173479			172	139	
GRMZM2G036502			654	254		GRMZM2G173668			499	466	
GRMZM2G036619			452			GRMZM2G173674			498	514	
GRMZM2G036980	198	153				GRMZM2G173747			116	74	
GRMZM2G037204					193	GRMZM2G173763			79	194	59
GRMZM2G037255	429	742			65	GRMZM2G174255					462
GRMZM2G037286			889		761	GRMZM2G174517	968	834			518
GRMZM2G037334					562	GRMZM2G174650	11	1			
GRMZM2G037343			888		760	GRMZM2G174708	587	591			
GRMZM2G037350					315	GRMZM2G174719				960	
GRMZM2G037413			576			GRMZM2G174732			70	83	530
GRMZM2G037496			441	229		GRMZM2G174757	326	394			
GRMZM2G037685			181	421		GRMZM2G174860			217	156	179
GRMZM2G037781			372	282		GRMZM2G174938	693	631			
GRMZM2G037910			180	420		GRMZM2G174984				882	
GRMZM2G038034	715				41	GRMZM2G175134	869				
GRMZM2G038243					326	GRMZM2G175188					569
GRMZM2G038340			776			GRMZM2G175236			199	131	
GRMZM2G038449	910	782				GRMZM2G175549			822		

GRMZM2G038811	25	16				GRMZM2G175661					779
GRMZM2G038846					227	GRMZM2G175761	7	3			
GRMZM2G038988	718	963				GRMZM2G175827			380	372	
GRMZM2G039106	463	620				GRMZM2G176595	757	863			105
GRMZM2G039711		893				GRMZM2G177052			373	206	
GRMZM2G039845		953				GRMZM2G177061			880	539	
GRMZM2G039870					567	GRMZM2G177104			784	962	
GRMZM2G039934			130	101	134	GRMZM2G177284			601	325	
GRMZM2G039968			807			GRMZM2G177391	591	681			39
GRMZM2G040052			808			GRMZM2G177400	785	839			38
GRMZM2G040094	690	687				GRMZM2G177447	977				
GRMZM2G040121			809			GRMZM2G177461					476
GRMZM2G040158	803	912				GRMZM2G177866			83		
GRMZM2G040236			82	155	632	GRMZM2G177883					580
GRMZM2G040320	689	686				GRMZM2G177929	849	942			578
GRMZM2G040326			624	719		GRMZM2G177940	848	941			577
GRMZM2G040508			586	161		GRMZM2G177970	847	940			576
GRMZM2G040555			971			GRMZM2G178025				915	
GRMZM2G040587					388	GRMZM2G178079	293	356			
GRMZM2G040630					391	GRMZM2G178102	765	524			
GRMZM2G040706			701			GRMZM2G178356				874	
GRMZM2G040750	368	452				GRMZM2G178435					446
GRMZM2G040838			716			GRMZM2G178576					254
GRMZM2G041028	500	403				GRMZM2G178801					482
GRMZM2G041050					123	GRMZM2G178886		885			
GRMZM2G041068					651	GRMZM2G178894	195	303			
GRMZM2G041167	887	713				GRMZM2G179253	932	587			
GRMZM2G041344			870	346		GRMZM2G179264				604	261
GRMZM2G041497					475	GRMZM2G179268	934	589			
GRMZM2G041554					122	GRMZM2G179274				606	263
GRMZM2G041561					560	GRMZM2G179758					513
GRMZM2G041613	356	617				GRMZM2G179789	348	300			862
GRMZM2G041831	488	555			795	GRMZM2G179792	548	575			861
GRMZM2G041963			123	464		GRMZM2G179976			582	443	
GRMZM2G041980			473	225		GRMZM2G179984			581	442	
GRMZM2G042006	457	795				GRMZM2G180408					91
GRMZM2G042118	739	806				GRMZM2G180775			241	141	81
GRMZM2G042181					463	GRMZM2G180909	140	229			35
GRMZM2G042442			80	58		GRMZM2G180916	142	231			37
GRMZM2G042443	933	588				GRMZM2G180983					802

GRMZM2G042538					209	GRMZM2G181491	610	398			
GRMZM2G042712	646	476			68	GRMZM2G181566					444
GRMZM2G042741	648	478			70	GRMZM2G181568					445
GRMZM2G042765	649	479			71	GRMZM2G300163					706
GRMZM2G042776					290	GRMZM2G300258			506	291	
GRMZM2G042855		908				GRMZM2G300348	598	472			
GRMZM2G042865	966	848				GRMZM2G300692	626	547			
GRMZM2G043069	967	849				GRMZM2G300916			126	64	
GRMZM2G043162			509	335		GRMZM2G300924			215	174	
GRMZM2G043477					539	GRMZM2G301853	529	551			
GRMZM2G043855					783	GRMZM2G302160	725	570	306	579	
GRMZM2G043983					946	GRMZM2G302259	724	569	307	580	
GRMZM2G044246			478	842	317	GRMZM2G302322				746	
GRMZM2G044481	430	382				GRMZM2G302373	291	354			
GRMZM2G044548	550	445				GRMZM2G302701			174	724	
GRMZM2G044684				702		GRMZM2G302712			176	726	
GRMZM2G044744					820	GRMZM2G302866			533		
GRMZM2G044771					819	GRMZM2G302876			638		
GRMZM2G044797					818	GRMZM2G303241	904	753			
GRMZM2G044832			574	238		GRMZM2G303752			705		
GRMZM2G044866					817	GRMZM2G303909					615
GRMZM2G044882				744	805	GRMZM2G304010					307
GRMZM2G044902	113	217				GRMZM2G304049			551	506	585
GRMZM2G044908	432	384				GRMZM2G304132	283	351			128
GRMZM2G045005	629					GRMZM2G305154	160	131			
GRMZM2G045435		785				GRMZM2G305400	261	228			575
GRMZM2G045668					351	GRMZM2G306172			107	308	
GRMZM2G045944			706	700		GRMZM2G306412					338
GRMZM2G045971			394		164	GRMZM2G306482	736	876			201
GRMZM2G046326					957	GRMZM2G306851		921			
GRMZM2G046480	287	221				GRMZM2G307262			504	351	
GRMZM2G046520				709	548	GRMZM2G307437			358	183	
GRMZM2G046537		950				GRMZM2G308799					952
GRMZM2G046743				777		GRMZM2G308860	702	715			
GRMZM2G046804				778		GRMZM2G308944			851	568	
GRMZM2G046848	762	762			60	GRMZM2G309248	96	99			27
GRMZM2G046947	324	332				GRMZM2G309495					514
GRMZM2G047167					527	GRMZM2G310652				548	
GRMZM2G047187	187	204			347	GRMZM2G310944					637
GRMZM2G047208					526	GRMZM2G311036	33	20			

GRMZM2G047434	18	14				GRMZM2G311401				461	
GRMZM2G047533			354	280		GRMZM2G311465	644	749			
GRMZM2G047607	390	561				GRMZM2G312026				752	
GRMZM2G047715			74	84		GRMZM2G312806			761	312	
GRMZM2G047763				730		GRMZM2G313009					506
GRMZM2G047813	714					GRMZM2G313305					306
GRMZM2G047867			259	433		GRMZM2G313327		972			807
GRMZM2G047894					724	GRMZM2G313341			160	772	
GRMZM2G047968	245	205				GRMZM2G313643	353	314			
GRMZM2G048136	704	717				GRMZM2G313672	354	315			
GRMZM2G048140					441	GRMZM2G313703			644	449	
GRMZM2G048297			525	221		GRMZM2G313707			645	499	
GRMZM2G048324					403	GRMZM2G313750				644	
GRMZM2G048665	589	960				GRMZM2G314297			653	253	
GRMZM2G048672	889	861				GRMZM2G314647					970
GRMZM2G048800	881	856				GRMZM2G314692					969
GRMZM2G049070	83	97			859	GRMZM2G315506				890	
GRMZM2G049091			657	764		GRMZM2G315786	560	773			
GRMZM2G049127			299	779		GRMZM2G316275	686	721			
GRMZM2G049173			300	780		GRMZM2G316447			353	279	
GRMZM2G049185			302	782		GRMZM2G316595	456	493			
GRMZM2G049424				846		GRMZM2G316827					868
GRMZM2G049549			141	477		GRMZM2G316904					269
GRMZM2G049568			140	476		GRMZM2G316907					272
GRMZM2G049672	535	638				GRMZM2G317262					885
GRMZM2G049895	536	639				GRMZM2G317267					886
GRMZM2G049921			994	980		GRMZM2G317270	564	443			
GRMZM2G049952	538	641				GRMZM2G317386	196	359			31
GRMZM2G049954	539	642				GRMZM2G317900	483	826			
GRMZM2G049990			864		221	GRMZM2G321020		627			
GRMZM2G050089	573	657				GRMZM2G321023		626			
GRMZM2G050268			637			GRMZM2G321753	818	889			
GRMZM2G050305			635			GRMZM2G321839					926
GRMZM2G050561	389	308				GRMZM2G321870	819	890			
GRMZM2G050705	388	307				GRMZM2G321940					811
GRMZM2G050851	813	882				GRMZM2G322672					88
GRMZM2G050867	248	191				GRMZM2G323925			280	439	913
GRMZM2G050912	247	190				GRMZM2G323936	52	23			45
GRMZM2G050961					292	GRMZM2G323971				761	
GRMZM2G051311			13	8		GRMZM2G324340			489	339	

GRMZM2G051367	569	573				GRMZM2G324643			792	589	
GRMZM2G051697			584	614		GRMZM2G324671			200	175	197
GRMZM2G051787			583	613		GRMZM2G325008			948		
GRMZM2G051795		918				GRMZM2G325131					712
GRMZM2G051879	775	567				GRMZM2G325580			19	42	
GRMZM2G051917		796				GRMZM2G326223					722
GRMZM2G051974		797				GRMZM2G326227	310	348			
GRMZM2G052111			618	525		GRMZM2G326631	148	143			
GRMZM2G052129			619	526		GRMZM2G326783	407	494			
GRMZM2G052178		800				GRMZM2G327427	892	870			
GRMZM2G052344	177	177				GRMZM2G327907	217	233			
GRMZM2G052357			248	89		GRMZM2G329559					228
GRMZM2G052442			345	533		GRMZM2G329962			484	177	
GRMZM2G052619	616					GRMZM2G330095	244	222			775
GRMZM2G052630	971					GRMZM2G330218			196	407	
GRMZM2G052666					428	GRMZM2G330379	744				
GRMZM2G053066	511	542				GRMZM2G330430				938	
GRMZM2G053397	614	597				GRMZM2G330772		865			
GRMZM2G053531					557	GRMZM2G330792				878	
GRMZM2G053639					491	GRMZM2G331614			231	524	
GRMZM2G053652					492	GRMZM2G331701			227	520	
GRMZM2G053720	104	81			217	GRMZM2G332412			487	459	
GRMZM2G053742	103	80			265	GRMZM2G332703	215	172			
GRMZM2G053757	425	409				GRMZM2G332809	607	395			
GRMZM2G053790					493	GRMZM2G332829			96	47	
GRMZM2G053908					494	GRMZM2G333049					725
GRMZM2G053925				944		GRMZM2G333083					700
GRMZM2G053952					511	GRMZM2G333183			774	468	
GRMZM2G053991			669	676	275	GRMZM2G333377	917				
GRMZM2G054023				945		GRMZM2G333641					364
GRMZM2G054040					394	GRMZM2G334165		901			
GRMZM2G054247					224	GRMZM2G334628					387
GRMZM2G054507	249	296			244	GRMZM2G334660	662	733			
GRMZM2G054588		767			797	GRMZM2G334722					156
GRMZM2G054658					854	GRMZM2G334740					389
GRMZM2G054689					226	GRMZM2G334741					495
GRMZM2G055141				832		GRMZM2G335604			401	685	
GRMZM2G055309				833		GRMZM2G336824	28	32			
GRMZM2G055437			522	575		GRMZM2G336839			453		
GRMZM2G055499			378	310	849	GRMZM2G336962			451		

GRMZM2G055520			377	309	848	GRMZM2G337621			731	923	
GRMZM2G055782			384	163		GRMZM2G337633		811			
GRMZM2G056039		937				GRMZM2G338049			157	121	
GRMZM2G056078			218	366	146	GRMZM2G339151			245	142	
GRMZM2G056093	328	380				GRMZM2G339367				822	
GRMZM2G056120	842	750				GRMZM2G339725					629
GRMZM2G056151	815	887				GRMZM2G340224	908	964			
GRMZM2G056166	327	379				GRMZM2G340279	405	334			
GRMZM2G056224				603		GRMZM2G340286	404	333			
GRMZM2G056270	929	526				GRMZM2G341723	369	453			
GRMZM2G056350				602		GRMZM2G341957					803
GRMZM2G056442				875		GRMZM2G342246	188	347			9
GRMZM2G056573			868	293		GRMZM2G342738	408	525			796
GRMZM2G056686	467					GRMZM2G343048			272	432	
GRMZM2G056851	72	51				GRMZM2G343351	647	477			69
GRMZM2G056884			982			GRMZM2G343360	650	480			72
GRMZM2G056889			981			GRMZM2G343365	651	481			73
GRMZM2G056903			980			GRMZM2G343688	431	383			
GRMZM2G056961	522	646				GRMZM2G344279	709	438			
GRMZM2G057000			159	771		GRMZM2G345155	296	375			
GRMZM2G057026	521	645				GRMZM2G345493	681	503			
GRMZM2G057091	520	644				GRMZM2G345544	682	504			
GRMZM2G057186		916				GRMZM2G345717			985		
GRMZM2G057258	519	643				GRMZM2G346278				745	806
GRMZM2G057412	772	725				GRMZM2G346639					956
GRMZM2G057466	491	328				GRMZM2G346750			339	956	
GRMZM2G057571			158	770		GRMZM2G346837	882	857			
GRMZM2G057611				992		GRMZM2G347583	122	197			
GRMZM2G057768	393	501				GRMZM2G347808	286	220			839
GRMZM2G057841	886	712				GRMZM2G347983	19	15			
GRMZM2G057959					636	GRMZM2G348780					442
GRMZM2G058081	941					GRMZM2G348956		798			
GRMZM2G058149	588				7	GRMZM2G348959		799			
GRMZM2G058374					771	GRMZM2G349185	412	413			
GRMZM2G058472					751	GRMZM2G349352			237	145	
GRMZM2G059138	490	499				GRMZM2G349556			628	667	
GRMZM2G059167			58	27		GRMZM2G349709			405		
GRMZM2G059179				812		GRMZM2G350210	612	599			
GRMZM2G059266				813		GRMZM2G350399	572	656			
GRMZM2G059325				815		GRMZM2G350410	571	655			

GRMZM2G059393					794	GRMZM2G350633			568	384	
GRMZM2G059445				836		GRMZM2G351018	575	709			
GRMZM2G059538				837		GRMZM2G351373		761		740	
GRMZM2G059556				838		GRMZM2G351382		760		739	
GRMZM2G059618			64	55		GRMZM2G351505					395
GRMZM2G059643			122	69		GRMZM2G351775		873			
GRMZM2G059663					401	GRMZM2G351921	975				
GRMZM2G059700					399	GRMZM2G351937	973				
GRMZM2G060023			710	731		GRMZM2G351941	972				
GRMZM2G060045			424	334		GRMZM2G352042	537	640			
GRMZM2G060118				931		GRMZM2G353250			670	677	276
GRMZM2G060170				674		GRMZM2G353313			863		220
GRMZM2G060206			106	106		GRMZM2G353734			661	344	
GRMZM2G060213			917	536		GRMZM2G353967	615	598			
GRMZM2G060290			73	220		GRMZM2G354618			681	283	
GRMZM2G060319			987	671		GRMZM2G355389					510
GRMZM2G060324				675		GRMZM2G355610					225
GRMZM2G060357				936		GRMZM2G355636		965			
GRMZM2G060373	699	621				GRMZM2G355667		966			
GRMZM2G060464			841	387		GRMZM2G356321		741			
GRMZM2G060720			187	230		GRMZM2G356556	426	410			
GRMZM2G060723					666	GRMZM2G358009		586			
GRMZM2G060937			266	186		GRMZM2G358050		585			
GRMZM2G060977	307	225			204	GRMZM2G358667			866	916	
GRMZM2G061099	914	958			268	GRMZM2G359102			243	170	
GRMZM2G061184			859	819		GRMZM2G359142	73	52			
GRMZM2G061283			930			GRMZM2G359589					640
GRMZM2G061287	807	735				GRMZM2G359986			63	54	
GRMZM2G061398	808	736				GRMZM2G360352				814	
GRMZM2G061403	308	226			205	GRMZM2G360374			220	595	
GRMZM2G061485					825	GRMZM2G360389			448	242	
GRMZM2G061495			222	597		GRMZM2G360519	809	737			
GRMZM2G061695			664			GRMZM2G360529	810	738			
GRMZM2G061723			663			GRMZM2G360541			696		
GRMZM2G061890			662			GRMZM2G360615	90	92			
GRMZM2G062009				543	887	GRMZM2G360794	864	935			
GRMZM2G062042		775				GRMZM2G361398			730	416	
GRMZM2G062091				544	425	GRMZM2G361423			860	820	
GRMZM2G062391					414	GRMZM2G361688	309	227			206
GRMZM2G062425					415	GRMZM2G361718	884	934			

GRMZM2G062706			153	527	187	GRMZM2G361855	334	451			
GRMZM2G062844				828		GRMZM2G361896	441	535	732		141
GRMZM2G062996	436	269			709	GRMZM2G361902	332	449			
GRMZM2G063188			256	118		GRMZM2G361917	333	450			
GRMZM2G063394					889	GRMZM2G363535	567				
GRMZM2G063478			805			GRMZM2G363540				785	
GRMZM2G063492			804			GRMZM2G363554	566				
GRMZM2G063688		440				GRMZM2G363583					865
GRMZM2G063792				753		GRMZM2G364285	69	59			
GRMZM2G063850			951			GRMZM2G365134	346	722			4
GRMZM2G063875					843	GRMZM2G365319	300	291			
GRMZM2G064008	630	669				GRMZM2G365423				616	
GRMZM2G064133				783		GRMZM2G365888			838		
GRMZM2G064145				967		GRMZM2G366485			182	97	
GRMZM2G064533				801		GRMZM2G366532					842
GRMZM2G064584			416	137		GRMZM2G366659			458	226	
GRMZM2G064605			823			GRMZM2G367023	24	13			
GRMZM2G064715					792	GRMZM2G367483	844	830			
GRMZM2G065021			240	154		GRMZM2G368632					876
GRMZM2G065050	867	944			264	GRMZM2G369069	954				
GRMZM2G065168	169	144				GRMZM2G369149	623	652			
GRMZM2G065174	171	146				GRMZM2G369799			469	618	
GRMZM2G065205	554	845			214	GRMZM2G369987			7	29	
GRMZM2G065225			296	129		GRMZM2G370519	712	897			
GRMZM2G065244			179	86		GRMZM2G370529	57	76			319
GRMZM2G065374			738			GRMZM2G370693	618	509			
GRMZM2G065478					541	GRMZM2G371137	802	911			
GRMZM2G065496					540	GRMZM2G371229			691	402	
GRMZM2G065718	409	416				GRMZM2G372171			410	132	
GRMZM2G065908			712	482		GRMZM2G372200					420
GRMZM2G065939			947			GRMZM2G372633			916	546	
GRMZM2G066171	314	500			64	GRMZM2G373329	528	534			948
GRMZM2G066326			219	320		GRMZM2G373341	492	455			545
GRMZM2G066480			258	181		GRMZM2G373607	544	523			973
GRMZM2G066551			257	180		GRMZM2G374313				567	
GRMZM2G066618			103	62		GRMZM2G374475			355	686	
GRMZM2G066683					497	GRMZM2G374574			357	688	
GRMZM2G067225					770	GRMZM2G375302			128	143	
GRMZM2G067257					291	GRMZM2G375733			359	330	
GRMZM2G067277			203	80		GRMZM2G376074	603	653			

GRMZM2G067298					695	GRMZM2G376282					954
GRMZM2G067320	120	129				GRMZM2G376927					499
GRMZM2G067388					769	GRMZM2G377131					643
GRMZM2G067402					768	GRMZM2G377161					904
GRMZM2G067426					767	GRMZM2G377217					32
GRMZM2G067555	768					GRMZM2G377887			595	620	
GRMZM2G067587	769					GRMZM2G377889					975
GRMZM2G067964			957			GRMZM2G379375					670
GRMZM2G067984				804		GRMZM2G379550	476	607			
GRMZM2G068239	305	304			342	GRMZM2G379835	577	512			
GRMZM2G068294		776				GRMZM2G380247			862	948	694
GRMZM2G068479	186	289				GRMZM2G380738	964				
GRMZM2G068506				708		GRMZM2G380777	475	336			
GRMZM2G068510			673	680		GRMZM2G381168			543	757	
GRMZM2G068547			777			GRMZM2G381691			305		184
GRMZM2G068590		898				GRMZM2G382077			573	887	
GRMZM2G068701	17	22				GRMZM2G382557					185
GRMZM2G068826	35	37				GRMZM2G382717	66	65			
GRMZM2G069025	631	608				GRMZM2G382774	811	967			864
GRMZM2G069082	632	609				GRMZM2G383303	755				
GRMZM2G069126	633	610				GRMZM2G383594	485	370			
GRMZM2G069146	634	611				GRMZM2G383631	796	520			300
GRMZM2G069176	74	116				GRMZM2G384564			837	803	
GRMZM2G069177					701	GRMZM2G384762			993		
GRMZM2G069201	621	650				GRMZM2G384884	499	378			
GRMZM2G069274	622	651				GRMZM2G384972					547
GRMZM2G069476	773	660				GRMZM2G385050	970				
GRMZM2G069542			771	413		GRMZM2G385428		954			
GRMZM2G069594	774	661				GRMZM2G385619			316	200	
GRMZM2G069606	80	70				GRMZM2G387471			526	607	
GRMZM2G069664			614	971		GRMZM2G387485			594	619	
GRMZM2G069672	601	752				GRMZM2G387576				877	
GRMZM2G069678			399	224		GRMZM2G388371	854	779			
GRMZM2G069765			285	585		GRMZM2G388420	468	614			
GRMZM2G069773			613	970		GRMZM2G388778			66	23	
GRMZM2G069928	355	316				GRMZM2G388915			168	593	
GRMZM2G070271	698	731				GRMZM2G389240	306	305			
GRMZM2G070295					910	GRMZM2G389462			630	774	
GRMZM2G070343			876	354		GRMZM2G389625					732
GRMZM2G070442			284	584		GRMZM2G389878	36	43			233

GRMZM2G070462	211	167			599	GRMZM2G389880	37	44			234
GRMZM2G070620	86	96				GRMZM2G390096				926	525
GRMZM2G070693	920					GRMZM2G390150			395	272	523
GRMZM2G070716			283	583		GRMZM2G390489			748	342	
GRMZM2G070865				615		GRMZM2G390691				884	
GRMZM2G070943					149	GRMZM2G391312				841	
GRMZM2G071042			332	215		GRMZM2G391833					604
GRMZM2G071059	421	391				GRMZM2G392037			468	645	
GRMZM2G071101	691	600				GRMZM2G392477	495	675			838
GRMZM2G071157	479	465			413	GRMZM2G392524	494	674			837
GRMZM2G071172	478	464			412	GRMZM2G392863			912		
GRMZM2G071208	527	533			947	GRMZM2G392956			915		
GRMZM2G071268			621	721		GRMZM2G393039			31	99	106
GRMZM2G071307			609			GRMZM2G393057			633	305	
GRMZM2G071322			610			GRMZM2G393146					286
GRMZM2G071396				872		GRMZM2G393843				561	504
GRMZM2G071448				669		GRMZM2G394403	613	393			
GRMZM2G071575	919					GRMZM2G394410	417	361			
GRMZM2G071688	63	86				GRMZM2G394500			819		
GRMZM2G071928			411	133		GRMZM2G394747					211
GRMZM2G072041			211	983		GRMZM2G394827	584	635			
GRMZM2G072088			212	984		GRMZM2G394945			47	78	
GRMZM2G072115			23	53		GRMZM2G395039					658
GRMZM2G072117	658	704			944	GRMZM2G395094					699
GRMZM2G072292			389			GRMZM2G395236	834	702			
GRMZM2G072350			759	375	219	GRMZM2G395244	833	701			
GRMZM2G072371	659	705			408	GRMZM2G395534			361	844	
GRMZM2G072388			798			GRMZM2G395853		813			
GRMZM2G072589			169	257		GRMZM2G396212	166	139			349
GRMZM2G072909			333	590		GRMZM2G396418					305
GRMZM2G072939					255	GRMZM2G397948					321
GRMZM2G073044					496	GRMZM2G398731		959			
GRMZM2G073826			538	274		GRMZM2G398795			444	231	
GRMZM2G073836				560		GRMZM2G398825	655	680			756
GRMZM2G073842					159	GRMZM2G399098					955
GRMZM2G073860			857			GRMZM2G399433		929			
GRMZM2G073908	545	553				GRMZM2G399541	902				148
GRMZM2G073954					430	GRMZM2G399584	766				
GRMZM2G074087			727	479		GRMZM2G400005	570	574			
GRMZM2G074094	733	854				GRMZM2G400129			28	14	

GRMZM2G074138	75	94				GRMZM2G400135		968	728	480	
GRMZM2G074173					438	GRMZM2G400173			940		
GRMZM2G074211			485	855	320	GRMZM2G400197					984
GRMZM2G074280			726	478		GRMZM2G401561			858		
GRMZM2G074282			356	687		GRMZM2G401997			650	905	
GRMZM2G074300				955		GRMZM2G402368					659
GRMZM2G074331	531	433				GRMZM2G402564	269	346			
GRMZM2G074377	532	434				GRMZM2G402653			143	120	
GRMZM2G074436			779	701		GRMZM2G402708					727
GRMZM2G074501			718	576		GRMZM2G403712			678	485	
GRMZM2G074514					953	GRMZM2G403915			255	76	
GRMZM2G074648			919	538		GRMZM2G404121				825	
GRMZM2G075000				976	189	GRMZM2G404377			625		
GRMZM2G075058	301	245				GRMZM2G404426			902		
GRMZM2G075092	302	246				GRMZM2G404530			390	292	884
GRMZM2G075101				954		GRMZM2G405307	214	224			
GRMZM2G075294					222	GRMZM2G405387					213
GRMZM2G075456					863	GRMZM2G405622	32	66			155
GRMZM2G075715		900				GRMZM2G406101				964	
GRMZM2G075851			33	38		GRMZM2G406119				919	
GRMZM2G075892	131	179				GRMZM2G406376					334
GRMZM2G075921					312	GRMZM2G407119					789
GRMZM2G076157		793				GRMZM2G407406			722	665	
GRMZM2G076392			596	621		GRMZM2G407522	277	454			
GRMZM2G076468	162	214			964	GRMZM2G407913	639	467	904	611	
GRMZM2G076683				895		GRMZM2G407969					336
GRMZM2G076747				893		GRMZM2G407996			936	301	
GRMZM2G076936			846	394		GRMZM2G408038			205	510	
GRMZM2G076962			442	245		GRMZM2G408598	838	601			
GRMZM2G077045	620	654				GRMZM2G408875	761				
GRMZM2G077258					905	GRMZM2G409265	828	726			
GRMZM2G077316	76	90			750	GRMZM2G409343			555		
GRMZM2G077317			189	148		GRMZM2G409430			155	738	
GRMZM2G077486			190	149		GRMZM2G409473			684		
GRMZM2G077488					669	GRMZM2G409753			476	318	
GRMZM2G077757					827	GRMZM2G411216			322		
GRMZM2G077897	118	79				GRMZM2G411653			328	683	
GRMZM2G077960	117	78				GRMZM2G411668			327	682	
GRMZM2G077991	484	306				GRMZM2G412161	602	689			
GRMZM2G078252					642	GRMZM2G412426	349	301			

GRMZM2G078275					644	GRMZM2G412430	155	132			
GRMZM2G078283			12	7		GRMZM2G413226	265	278			
GRMZM2G078294			599	487		GRMZM2G413996					719
GRMZM2G078297	176	215				GRMZM2G414007					718
GRMZM2G078368	108	120				GRMZM2G414225					294
GRMZM2G078382				564		GRMZM2G414278	487	560			649
GRMZM2G078806	200	151				GRMZM2G414331	925				
GRMZM2G078876	139	106				GRMZM2G414639					111
GRMZM2G079031	272	784			6	GRMZM2G414866	444	280			
GRMZM2G079470	474	568			318	GRMZM2G414995	267	339			
GRMZM2G079487	133	209			424	GRMZM2G415012	266	338			
GRMZM2G079559					664	GRMZM2G416156	128	158			617
GRMZM2G079568	134	210			426	GRMZM2G416184	178	200			
GRMZM2G079730	965					GRMZM2G416308	374	400			
GRMZM2G079754	963					GRMZM2G416652	927				
GRMZM2G079774	798					GRMZM2G417402			600	324	
GRMZM2G079817				513		GRMZM2G418258			214	111	
GRMZM2G079954			500	399		GRMZM2G419104			239	147	
GRMZM2G080169				889		GRMZM2G419563				737	
GRMZM2G080178			461	336		GRMZM2G420619					478
GRMZM2G080191			544	758		GRMZM2G420684					982
GRMZM2G080281			462	337		GRMZM2G420926			438		
GRMZM2G080530					752	GRMZM2G421779	233	424			
GRMZM2G080537	578	513				GRMZM2G421866					675
GRMZM2G080689			861	947	693	GRMZM2G422641	672	850			
GRMZM2G080731	197	152				GRMZM2G423025			944		
GRMZM2G080816			445			GRMZM2G423413					484
GRMZM2G080839					119	GRMZM2G423486			188		
GRMZM2G080843	676	592				GRMZM2G423669	174	196			
GRMZM2G080851	722	672	97	551		GRMZM2G423851		914			
GRMZM2G080858			191	213		GRMZM2G423861			147		
GRMZM2G080906				649		GRMZM2G423886	336	468			232
GRMZM2G080907					120	GRMZM2G423975			148		
GRMZM2G080930	677	593				GRMZM2G424075	298	377			
GRMZM2G081158					340	GRMZM2G424595	125	121			
GRMZM2G081350	546	364				GRMZM2G424625					901
GRMZM2G081359			459	185		GRMZM2G424628					902
GRMZM2G081441			510	205		GRMZM2G425719					75
GRMZM2G081486					856	GRMZM2G425736	931				
GRMZM2G081536				651		GRMZM2G425774			560		236

GRMZM2G081538	337	329				GRMZM2G426150	694	711			
GRMZM2G081582					628	GRMZM2G426511	711	896			
GRMZM2G081583			48	5		GRMZM2G426888			45	39	
GRMZM2G081626				650		GRMZM2G427635					313
GRMZM2G081712			680	289		GRMZM2G427685	167	128			
GRMZM2G081808	434	407			341	GRMZM2G427932			145		
GRMZM2G081928		822				GRMZM2G428071					175
GRMZM2G081935	31	21				GRMZM2G428197			290	277	
GRMZM2G081943					858	GRMZM2G429045	751	695			
GRMZM2G082180			829	486		GRMZM2G429540			641	446	
GRMZM2G082189					738	GRMZM2G429611		858			
GRMZM2G082362	515	420				GRMZM2G429992			91	49	
GRMZM2G082372	514	419				GRMZM2G430029			733	364	
GRMZM2G082376	513	418				GRMZM2G430463	901				
GRMZM2G082390	512	417				GRMZM2G430482					381
GRMZM2G082608	252	157	183	167	125	GRMZM2G430501					382
GRMZM2G082709	1	12				GRMZM2G430710	604	426			
GRMZM2G082836	65	64				GRMZM2G431006					928
GRMZM2G082916	716				42	GRMZM2G431524	153	105			609
GRMZM2G082940					345	GRMZM2G431691			15	35	685
GRMZM2G083091			569			GRMZM2G433528				986	
GRMZM2G083128	246	206				GRMZM2G433624	446	321			
GRMZM2G083156			570			GRMZM2G433731	788				
GRMZM2G083182			571			GRMZM2G433795			437	493	
GRMZM2G083301					784	GRMZM2G433855					974
GRMZM2G083309	936	904				GRMZM2G434839					951
GRMZM2G083367					785	GRMZM2G435244					616
GRMZM2G083402					786	GRMZM2G435290					879
GRMZM2G083411	565	444				GRMZM2G435294					878
GRMZM2G083642					736	GRMZM2G435824			758		
GRMZM2G083841					866	GRMZM2G436299	168	135			
GRMZM2G084063	48	56				GRMZM2G436593		809			
GRMZM2G084440	49	57				GRMZM2G437683	695	835			
GRMZM2G084463			873			GRMZM2G437711			195	406	
GRMZM2G084587	345	423				GRMZM2G437827				924	
GRMZM2G084791					165	GRMZM2G440259	729	667			
GRMZM2G084825			969			GRMZM2G440831			517	457	
GRMZM2G085825	232	189				GRMZM2G440866			516	456	
GRMZM2G085836			315	199		GRMZM2G441670	524	867			101
GRMZM2G086190					877	GRMZM2G441722				759	828

GRMZM2G086669	313	257				GRMZM2G441888					960
GRMZM2G086750	385	373			505	GRMZM2G441903					962
GRMZM2G086869	384	372			853	GRMZM2G442129			892		
GRMZM2G086940			725	922		GRMZM2G442523			667		
GRMZM2G087032			724	921		GRMZM2G442546			668		
GRMZM2G087063	719	819				GRMZM2G442551					528
GRMZM2G087117	950					GRMZM2G443111	325	178			153
GRMZM2G087243	721	821				GRMZM2G443332					544
GRMZM2G087323					469	GRMZM2G443340			966		248
GRMZM2G087350			651	906		GRMZM2G443345					13
GRMZM2G087484				856		GRMZM2G443447			177	727	
GRMZM2G087679			830			GRMZM2G443762			502	349	
GRMZM2G088443	754					GRMZM2G443776			501	348	
GRMZM2G088487	95	115				GRMZM2G443785					488
GRMZM2G088511			747	341		GRMZM2G444533	873	582			
GRMZM2G088524	845				85	GRMZM2G444540	872	581			
GRMZM2G088613					90	GRMZM2G444543	875	584			
GRMZM2G088765					622	GRMZM2G444560		974			
GRMZM2G089056			542			GRMZM2G444715	279	236			61
GRMZM2G089086			749	343		GRMZM2G445423	862	432			
GRMZM2G089092			559			GRMZM2G445689	489	723			
GRMZM2G089106			29	59		GRMZM2G446047					402
GRMZM2G089147	460	309				GRMZM2G446189				913	
GRMZM2G089285					593	GRMZM2G446858			976		
GRMZM2G089291					592	GRMZM2G447151					690
GRMZM2G089355	853	778				GRMZM2G447176					692
GRMZM2G089517					737	GRMZM2G447447				901	
GRMZM2G089618			793	382		GRMZM2G447542				898	
GRMZM2G090018	547	517				GRMZM2G447989					157
GRMZM2G090029	127	123				GRMZM2G448456			631	418	
GRMZM2G090034	861	431				GRMZM2G448895			953		
GRMZM2G090087	859	429				GRMZM2G449085	164	195			
GRMZM2G090100	126	122				GRMZM2G449355				728	
GRMZM2G090245	858	428				GRMZM2G449558					936
GRMZM2G090266	380	436				GRMZM2G449779					416
GRMZM2G090675			523			GRMZM2G450717					776
GRMZM2G090904			844	659		GRMZM2G450863					696
GRMZM2G091020			899	550		GRMZM2G451187			735	859	
GRMZM2G091226					603	GRMZM2G451746					110
GRMZM2G091233					602	GRMZM2G451769		903			

GRMZM2G091243	81	112	247	166		GRMZM2G451856					938
GRMZM2G091285					601	GRMZM2G452084					930
GRMZM2G091592					916	GRMZM2G452935					471
GRMZM2G091653			839	470		GRMZM2G453225					516
GRMZM2G091811			92	151	243	GRMZM2G453684	958				
GRMZM2G091995			689	400		GRMZM2G453832	322	282			
GRMZM2G092223	895	556				GRMZM2G455407			534		
GRMZM2G092363	280	271				GRMZM2G456149	697	435			
GRMZM2G092379			622			GRMZM2G456357					925
GRMZM2G092391		596				GRMZM2G456422				760	
GRMZM2G092432				710		GRMZM2G456547	401	532			
GRMZM2G092451			275	228		GRMZM2G456578	295	239			19
GRMZM2G092525					612	GRMZM2G456603	344	299			21
GRMZM2G092535			274	227		GRMZM2G456618	397	528			
GRMZM2G092550					611	GRMZM2G456626	422	363			
GRMZM2G092590	158	173			207	GRMZM2G456644	801				
GRMZM2G092804			430	554	311	GRMZM2G456756			317	164	
GRMZM2G092895		948				GRMZM2G456757			318	165	
GRMZM2G092916			913			GRMZM2G456835			983	625	
GRMZM2G092923			914			GRMZM2G457147	190	141			3
GRMZM2G092968			338	360		GRMZM2G457193			418	237	
GRMZM2G093020					606	GRMZM2G457357	367	265			
GRMZM2G093121			911	445		GRMZM2G457370	366	264			
GRMZM2G093291					285	GRMZM2G457415					299
GRMZM2G093312					284	GRMZM2G458084			754	474	
GRMZM2G093441	586	637				GRMZM2G458200			813	452	
GRMZM2G093503	585	636				GRMZM2G458448					981
GRMZM2G093524					398	GRMZM2G458538					977
GRMZM2G093580			883			GRMZM2G458659				928	
GRMZM2G093598			623			GRMZM2G458718	194	311			472
GRMZM2G093731	656	564				GRMZM2G458824					451
GRMZM2G093838	657	565				GRMZM2G459230			50	56	
GRMZM2G093900				959		GRMZM2G459370	938	874			
GRMZM2G093962	780					GRMZM2G459474				768	
GRMZM2G094077		932				GRMZM2G459645					665
GRMZM2G094083				811		GRMZM2G459828				636	
GRMZM2G094168	206	126				GRMZM2G459861					359
GRMZM2G094241					641	GRMZM2G460422	27	48			52
GRMZM2G094390	112	98				GRMZM2G460684			114	184	
GRMZM2G094452	9	11				GRMZM2G462080	132	107			507

GRMZM2G094526			640	393		GRMZM2G462422			496	658	759
GRMZM2G094639					253	GRMZM2G464469	517				
GRMZM2G094781	51	33				GRMZM2G464539			54	13	
GRMZM2G094871			89	50		GRMZM2G464709	464	923			365
GRMZM2G094884	583	634				GRMZM2G464794	948	804			174
GRMZM2G094892	582	633				GRMZM2G464846					598
GRMZM2G094990	559	772				GRMZM2G464885	284	352			103
GRMZM2G095020	976					GRMZM2G465771	688	730			
GRMZM2G095141			827			GRMZM2G466428			872	361	
GRMZM2G095144	541	698				GRMZM2G466517					566
GRMZM2G095326			828			GRMZM2G466519					662
GRMZM2G095384	453	507			62	GRMZM2G466743		933			
GRMZM2G095397			840	359		GRMZM2G466769			546	679	
GRMZM2G095411	278	320			63	GRMZM2G466931			874		
GRMZM2G095631	40	54	43	51		GRMZM2G467123			365	264	
GRMZM2G095643			142	323	147	GRMZM2G467134			366	265	
GRMZM2G096171			30	52		GRMZM2G467147			369	268	
GRMZM2G096365			387	862		GRMZM2G467212		977			
GRMZM2G096367			659	808		GRMZM2G468929				699	
GRMZM2G096422			723			GRMZM2G469371	827	920			
GRMZM2G096553					552	GRMZM2G469414	534	577			
GRMZM2G096600					551	GRMZM2G469697	58	69			
GRMZM2G096680				914		GRMZM2G469795	411	335			
GRMZM2G096996			739	281		GRMZM2G469839			702		
GRMZM2G097040				673		GRMZM2G469873			138	70	
GRMZM2G097068	817	853				GRMZM2G469898					749
GRMZM2G097089			950			GRMZM2G470723			400	684	
GRMZM2G097122			949			GRMZM2G471051	449	537			
GRMZM2G097170			939			GRMZM2G471080					210
GRMZM2G097277	835	703				GRMZM2G471348			958		
GRMZM2G097316					322	GRMZM2G471357					386
GRMZM2G097340					323	GRMZM2G471525			954	390	
GRMZM2G097404					76	GRMZM2G471529	241	109		867	33
GRMZM2G097468				789		GRMZM2G471635			178	124	
GRMZM2G097568	832	700				GRMZM2G472234	502	385			945
GRMZM2G097926					883	GRMZM2G472453				834	
GRMZM2G097995	503	505			860	GRMZM2G472903	667				
GRMZM2G098042		928				GRMZM2G474113			816	706	
GRMZM2G098331	753	461				GRMZM2G474211	871				
GRMZM2G098397		880				GRMZM2G474658					537

GRMZM2G098434		879				GRMZM2G474726			40	179	58
GRMZM2G098900	482	825				GRMZM2G475647			783	796	
GRMZM2G099109				690		GRMZM2G475678			933		
GRMZM2G099328	423	518			535	GRMZM2G476515				556	361
GRMZM2G100005	568	572				GRMZM2G476902	692	630			
GRMZM2G100121			945			GRMZM2G477161				881	
GRMZM2G100225			848			GRMZM2G477200	846	939			
GRMZM2G100328			348	222		GRMZM2G477238			881	540	
GRMZM2G100381					949	GRMZM2G477741					374
GRMZM2G100457			507	377		GRMZM2G477847			198	130	
GRMZM2G100492			470	314		GRMZM2G478212	219	175			5
GRMZM2G100583		956	428	331		GRMZM2G479340	149	136			
GRMZM2G101053					728	GRMZM2G479804					483
GRMZM2G101069			206	542		GRMZM2G479987				605	262
GRMZM2G101080					729	GRMZM2G480106			121	436	589
GRMZM2G101511			658	395		GRMZM2G480439	292	355			
GRMZM2G101745	4	5				GRMZM2G481163	141	230			36
GRMZM2G101875					899	GRMZM2G481222	382	284			
GRMZM2G101900	210	186				GRMZM2G481311	330	273			
GRMZM2G101926					898	GRMZM2G481531			26	150	
GRMZM2G101932					897	GRMZM2G481605					804
GRMZM2G102021			677	484		GRMZM2G481904	609	397			
GRMZM2G102156	414	487				GRMZM2G482256					543
GRMZM2G102161					533	GRMZM2G483622					715
GRMZM2G102218					534	GRMZM2G483624					714
GRMZM2G102322			649	904		GRMZM2G486868	723	739			
GRMZM2G102447					370	GRMZM2G487359					833
GRMZM2G102580			260	839		GRMZM2G489288			450		
GRMZM2G102674				977	190	GRMZM2G490278			766		
GRMZM2G102720	264	232				GRMZM2G493946	358	791			
GRMZM2G102915				966		GRMZM2G495234			432	240	
GRMZM2G102923				965		GRMZM2G496319	894	872			
GRMZM2G103050			427	892		GRMZM2G496821			281	440	914
GRMZM2G103166			901			GRMZM2G502467			775	469	
GRMZM2G103186			991	434		GRMZM2G507533					652
GRMZM2G103245			887	379		GRMZM2G508530			301	781	
GRMZM2G103266				963		GRMZM2G512617					638
GRMZM2G103721			347	374		GRMZM2G513273	413	290			
GRMZM2G103847			675			GRMZM2G515456				952	
GRMZM2G103939			204	193		GRMZM2G519736				581	

GRMZM2G104074			224	368	302	GRMZM2G522468				961	
GRMZM2G104078				736		GRMZM2G523621			286	586	
GRMZM2G104092			223	367	301	GRMZM2G524711	64	87			
GRMZM2G104176					870	GRMZM2G524978			690	401	
GRMZM2G104262			254	75		GRMZM2G525605			388		
GRMZM2G104353					568	GRMZM2G526668			909	666	
GRMZM2G104632	905	754				GRMZM2G527250			918	537	
GRMZM2G104739			463	201		GRMZM2G528010				894	
GRMZM2G104741			464	202		GRMZM2G533031			836	802	
GRMZM2G104768				968		GRMZM2G534888			406	357	
GRMZM2G104843			278	697		GRMZM2G537407			340	157	890
GRMZM2G104907			319	210	503	GRMZM2G538829	416	360			
GRMZM2G104942			320	211	841	GRMZM2G539790	781				
GRMZM2G105229			602			GRMZM2G540732	674	852			
GRMZM2G105542			279	698		GRMZM2G541103	504	506			554
GRMZM2G105571	855	913			572	GRMZM2G542847			799	571	
GRMZM2G105579			482	304		GRMZM2G548261			422	559	
GRMZM2G105787					788	GRMZM2G549274				949	
GRMZM2G105834					787	GRMZM2G556522			709	880	
GRMZM2G105869			938			GRMZM2G558078				870	
GRMZM2G105901					555	GRMZM2G559088			291	278	
GRMZM2G105933			244	234		GRMZM2G562569			436	492	
GRMZM2G105935	419	388				GRMZM2G563662	161	162			
GRMZM2G105954			580	353		GRMZM2G567739			921		
GRMZM2G106026	418	387				GRMZM2G568636				816	
GRMZM2G106099			686			GRMZM2G568912					810
GRMZM2G106108			687			GRMZM2G570109					765
GRMZM2G106143			865	417		GRMZM2G570989					586
GRMZM2G106165			466			GRMZM2G574858			497	261	
GRMZM2G106618			572	300		GRMZM2G575809	39	39			
GRMZM2G106650			811			GRMZM2G578411			831		
GRMZM2G106741			137	116		GRMZM2G580536			22	105	
GRMZM2G106752			136	115		GRMZM2G587543					596
GRMZM2G106781				918		GRMZM2G589696					757
GRMZM2G106792				917		GRMZM2G590450			120	435	588
GRMZM2G106834	156	133				GRMZM2G590792	239	111		826	763
GRMZM2G106836				920		GRMZM2G700034	764	838			
GRMZM2G107021				798		GRMZM2G700239	930	952			
GRMZM2G107121			292	419		GRMZM2G700275		829			
GRMZM2G107129			824			GRMZM2G700655					392

GRMZM2G107199				939		GRMZM2G700659	974				
GRMZM2G107289			20	103		GRMZM2G700662			589	609	113
GRMZM2G107336		917				GRMZM2G700664			588	608	112
GRMZM2G107388			297	541		GRMZM2G700665			76	190	99
GRMZM2G107395			133	411		GRMZM2G700673				950	
GRMZM2G107406					352	GRMZM2G700828		794			
GRMZM2G107437	890	862				GRMZM2G701349					773
GRMZM2G107440			132	410		GRMZM2G701352					772
GRMZM2G107473					600	GRMZM2G701585			849		
GRMZM2G107484	516	421				GRMZM2G701704		833			
GRMZM2G107495					500	GRMZM2G701883					289
GRMZM2G107499					427	GRMZM2G702390					456
GRMZM2G107504	740	807				GRMZM2G702397				891	
GRMZM2G107575					536	GRMZM2G702403			486	389	135
GRMZM2G107591					663	GRMZM2G702406			398	426	
GRMZM2G107741					687	GRMZM2G702422			409	140	
GRMZM2G107774			287	587		GRMZM2G702426			213	501	
GRMZM2G107805			288	588		GRMZM2G702476			736	858	
GRMZM2G107815			744	847		GRMZM2G702483			118	91	
GRMZM2G107854			745	848		GRMZM2G702498			352	236	
GRMZM2G107867	949					GRMZM2G702499			843	412	
GRMZM2G107872			746	849		GRMZM2G702502			986	670	
GRMZM2G107896					376	GRMZM2G702516					846
GRMZM2G108016			475	369		GRMZM2G702522				729	
GRMZM2G108103			55	159		GRMZM2G702564			77	71	
GRMZM2G108149			56	160		GRMZM2G702602				769	
GRMZM2G108180					650	GRMZM2G703132			490	340	
GRMZM2G108272	498	619				GRMZM2G703281					826
GRMZM2G108309	497	618				GRMZM2G703586					646
GRMZM2G108416	486	559			648	GRMZM2G703624					653
GRMZM2G108501	392	302				GRMZM2G703625		951			417
GRMZM2G108597	707	765				GRMZM2G703634					874
GRMZM2G108619	830	728				GRMZM2G703658	720	820			
GRMZM2G108668	829	727				GRMZM2G703670					178
GRMZM2G108780				910		GRMZM2G703671					343
GRMZM2G108861					332	GRMZM2G703677	526	719			241
GRMZM2G108904	665	616				GRMZM2G703688	383	371			852
GRMZM2G109271	885	602				GRMZM2G703691	44	26			
GRMZM2G109354			714	988		GRMZM2G703697	260	250			
GRMZM2G109425	831	699				GRMZM2G703701		724			

GRMZM2G109830				627		GRMZM2G703704	501	404			
GRMZM2G110027					237	GRMZM2G703706	814				239
GRMZM2G110067					896	GRMZM2G703749					431
GRMZM2G110076					895	GRMZM2G703755					614
GRMZM2G110369	2	2				GRMZM2G703769					231
GRMZM2G110381					436	GRMZM2G703783					912
GRMZM2G110553	671	629	376	735		GRMZM2G703790					324
GRMZM2G110582					474	GRMZM2G703846				943	
GRMZM2G110584	670	628	375	734		GRMZM2G703995	713	780			
GRMZM2G110681			620	720		GRMZM2G704212					350
GRMZM2G110932				555		GRMZM2G704220			908		
GRMZM2G111015			855	601		GRMZM2G704222			717		
GRMZM2G111028			685	376		GRMZM2G704224			741		
GRMZM2G111045			363	87		GRMZM2G704265	469	365			
GRMZM2G111117				750		GRMZM2G704270					455
GRMZM2G111208	70	84			900	GRMZM2G704277	628	658	312	643	780
GRMZM2G111228			988			GRMZM2G704345					630
GRMZM2G111247	685	466	903	610		GRMZM2G704475					246
GRMZM2G111261			989			GRMZM5G800429					298
GRMZM2G111300			990			GRMZM5G800751				496	
GRMZM2G111306					410	GRMZM5G801241	115	181	10	2	
GRMZM2G111324			326	681		GRMZM5G801875			833	640	
GRMZM2G111529				987		GRMZM5G802384	915				
GRMZM2G112154	473	763			247	GRMZM5G802941			825	315	
GRMZM2G112228					836	GRMZM5G803275			249	370	
GRMZM2G112278	22	31				GRMZM5G803381					158
GRMZM2G112769	347	254				GRMZM5G803949	921	770			
GRMZM2G113098					229	GRMZM5G804816				957	
GRMZM2G113139	116	108				GRMZM5G805675			94	153	
GRMZM2G113203					703	GRMZM5G806839	13	30			55
GRMZM2G113415					717	GRMZM5G806947	30	35			
GRMZM2G113761	237	216			595	GRMZM5G807619	645	690			
GRMZM2G113790			52	36		GRMZM5G807791			232	117	
GRMZM2G113883			216	363		GRMZM5G808652				831	
GRMZM2G113990		720				GRMZM5G809546	912	930			
GRMZM2G114119	851					GRMZM5G809663	213	261			
GRMZM2G114356				794		GRMZM5G811022	379	344			
GRMZM2G114667	99	102			89	GRMZM5G811034					647
GRMZM2G114672	924					GRMZM5G811095	433	554			447
GRMZM2G114702			755	718		GRMZM5G811373	270	414			238

GRMZM2G114706	406	670				GRMZM5G812126			814	714	422
GRMZM2G114707	98	101			118	GRMZM5G812228			511		
GRMZM2G114772	97	100			117	GRMZM5G812265	29	34			
GRMZM2G114788	923					GRMZM5G812326			895	628	
GRMZM2G114893	791	827			97	GRMZM5G812891	224	252			
GRMZM2G114906	792	828			98	GRMZM5G812923					597
GRMZM2G115049	857	427				GRMZM5G813343					295
GRMZM2G115077			86	68		GRMZM5G813556			414	136	
GRMZM2G115260	274	247				GRMZM5G813670	870				
GRMZM2G115424	480	447	975			GRMZM5G813892					230
GRMZM2G115766					83	GRMZM5G814035			818		
GRMZM2G115828					84	GRMZM5G814904				775	
GRMZM2G115855	776	707				GRMZM5G815736					532
GRMZM2G115875	202	155				GRMZM5G815757			85	67	
GRMZM2G115909			772	823		GRMZM5G816304					411
GRMZM2G115960	203	156				GRMZM5G816582			541	807	
GRMZM2G115961			773	824		GRMZM5G817035	680	502			
GRMZM2G116010	341	648				GRMZM5G817255					145
GRMZM2G116133			970			GRMZM5G817301			924		
GRMZM2G116204	71	42				GRMZM5G817651					822
GRMZM2G116292			166	853		GRMZM5G817777	255	368			
GRMZM2G116517	143	127				GRMZM5G817839	319	324			249
GRMZM2G116626					933	GRMZM5G817854					800
GRMZM2G116638					439	GRMZM5G818126	250	297			245
GRMZM2G116689	227	199				GRMZM5G818776			682	284	
GRMZM2G117007					375	GRMZM5G819500					813
GRMZM2G117153					892	GRMZM5G819965	361	402			550
GRMZM2G117222	396	549			86	GRMZM5G820650			519		
GRMZM2G117329			21	104		GRMZM5G821611	114	180	9	1	
GRMZM2G117360					676	GRMZM5G821787			335	508	
GRMZM2G117513	129	159			618	GRMZM5G822137				829	
GRMZM2G117755			236	144		GRMZM5G822237		866			
GRMZM2G117776	443	279				GRMZM5G823135					369
GRMZM2G117804	454	495			143	GRMZM5G824236					911
GRMZM2G117822			238	146		GRMZM5G824597				494	
GRMZM2G118063	218	237			778	GRMZM5G825987			931	577	
GRMZM2G118409	574	708				GRMZM5G826416					631
GRMZM2G118453	928	945				GRMZM5G826984					124
GRMZM2G118462	951					GRMZM5G827042					454
GRMZM2G118495			165	852		GRMZM5G827611		824			

GRMZM2G118497			164	851		GRMZM5G828284	46	89			
GRMZM2G118717			552	218		GRMZM5G828488	62	85			
GRMZM2G118766			553	219		GRMZM5G828630					857
GRMZM2G118957				692		GRMZM5G828945			379	311	850
GRMZM2G119483			591	317	741	GRMZM5G829396	208	248			
GRMZM2G119496			154			GRMZM5G830085			99	92	
GRMZM2G119527			590	316	740	GRMZM5G830457				799	
GRMZM2G119696			24	17		GRMZM5G830983					708
GRMZM2G119717			25	18		GRMZM5G831135			221	596	
GRMZM2G119766	542	521				GRMZM5G831408					358
GRMZM2G120069					432	GRMZM5G831486	340	647			
GRMZM2G120079		817			433	GRMZM5G831712		894			
GRMZM2G120132			337			GRMZM5G832381	466	925			367
GRMZM2G120246				717		GRMZM5G833253					574
GRMZM2G120373				428		GRMZM5G833563			896	629	
GRMZM2G120414	230	183				GRMZM5G833760	84	63			
GRMZM2G120563	229	182				GRMZM5G834260					144
GRMZM2G120587	67	68				GRMZM5G834605	172	147			
GRMZM2G120592			832	424		GRMZM5G835704					271
GRMZM2G120619					907	GRMZM5G836190			309	386	
GRMZM2G120839			234	112		GRMZM5G837058	420	390			
GRMZM2G120975			440			GRMZM5G837538	154	185			
GRMZM2G120987			439			GRMZM5G837869			767		
GRMZM2G121083			713	664		GRMZM5G838152			289	276	
GRMZM2G121117					983	GRMZM5G838196			756	509	
GRMZM2G121128	826	886			512	GRMZM5G838226			960		
GRMZM2G121143			815	715	423	GRMZM5G838352	299	319			610
GRMZM2G121354				935		GRMZM5G838414			513		
GRMZM2G121374				934		GRMZM5G838435			920		437
GRMZM2G121418					480	GRMZM5G838496					79
GRMZM2G121494			57	73		GRMZM5G839014			308	385	
GRMZM2G121510					814	GRMZM5G840762	10	19			
GRMZM2G121631			928	529		GRMZM5G840885			271	429	
GRMZM2G122076			173	204		GRMZM5G841594					139
GRMZM2G122231	678	594				GRMZM5G841743					829
GRMZM2G122284	679	595				GRMZM5G842103	276	327			
GRMZM2G122290					674	GRMZM5G842306			537	909	
GRMZM2G122330	481	448				GRMZM5G842556				830	
GRMZM2G122965	856	840				GRMZM5G842686		899			
GRMZM2G123140	372	331				GRMZM5G842766	552	381			

GRMZM2G123305	43	55		800		GRMZM5G842855					726
GRMZM2G123418					316	GRMZM5G843237	876				
GRMZM2G123585	275	337			11	GRMZM5G843446					77
GRMZM2G123607	228	330				GRMZM5G843972	703	716			
GRMZM2G123625	77	125				GRMZM5G844309				562	177
GRMZM2G123776	555	764			485	GRMZM5G845080			898	631	
GRMZM2G123872					452	GRMZM5G845296	38	45			235
GRMZM2G123987					354	GRMZM5G846071			565		681
GRMZM2G124059					355	GRMZM5G846082					372
GRMZM2G124115				835		GRMZM5G846140	440	546			
GRMZM2G124313					467	GRMZM5G847100					798
GRMZM2G124421					453	GRMZM5G847159			456	327	634
GRMZM2G124466			144			GRMZM5G847387					965
GRMZM2G124557			170	638		GRMZM5G847923				843	
GRMZM2G124638			171	639		GRMZM5G848365			471	661	
GRMZM2G125001					903	GRMZM5G848435			382	251	
GRMZM2G125149			146			GRMZM5G848692			250	371	
GRMZM2G125162			412	134		GRMZM5G848739			834	641	
GRMZM2G125201	784	469				GRMZM5G848942					958
GRMZM2G125314	710	895				GRMZM5G849471	877				
GRMZM2G125436			385	860		GRMZM5G850848	373	759			
GRMZM2G125476			386	861		GRMZM5G850924					140
GRMZM2G125529			210	128		GRMZM5G851617			963		
GRMZM2G125556			209	127		GRMZM5G851815			101	208	
GRMZM2G125668				827		GRMZM5G852170	6	8			
GRMZM2G125777			964	489		GRMZM5G852177	562	625			
GRMZM2G125853	101	187			625	GRMZM5G852338			556		
GRMZM2G125867	315	491				GRMZM5G852396			512		
GRMZM2G125923					293	GRMZM5G852533	359	792			
GRMZM2G125943					697	GRMZM5G853202					274
GRMZM2G125954	800					GRMZM5G854901				941	
GRMZM2G125991					698	GRMZM5G855094			93	152	242
GRMZM2G126128			41	19		GRMZM5G855629		909			
GRMZM2G126264			46	40		GRMZM5G855672				767	142
GRMZM2G126447			942	528		GRMZM5G855853		955	100	102	
GRMZM2G126603			665			GRMZM5G856067					702
GRMZM2G126900					556	GRMZM5G856180				490	
GRMZM2G127029			666	358		GRMZM5G856583			109	82	
GRMZM2G127067	144	165			654	GRMZM5G856653	850	943			579
GRMZM2G127087	23	28				GRMZM5G856795				896	

GRMZM2G127139					942	GRMZM5G857770	378	343			
GRMZM2G127168	145	166			655	GRMZM5G858165					357
GRMZM2G127173					941	GRMZM5G858653			193	296	
GRMZM2G127184					940	GRMZM5G859786	212	260			
GRMZM2G127266			323	635		GRMZM5G860072	256	369			
GRMZM2G127336	866	805				GRMZM5G860176					78
GRMZM2G127720				885		GRMZM5G860399			842	388	
GRMZM2G127945					790	GRMZM5G861617	331	323			
GRMZM2G128219			592	232		GRMZM5G861659					821
GRMZM2G128319					287	GRMZM5G861678	940	915			
GRMZM2G128399					208	GRMZM5G862016			615	258	
GRMZM2G128445			984	535		GRMZM5G862109			956	472	
GRMZM2G128504			585	260		GRMZM5G862276	455	527			
GRMZM2G128641	386	276				GRMZM5G862467	271	415			707
GRMZM2G128648	121	130				GRMZM5G862565	779	786			331
GRMZM2G128809	576	884				GRMZM5G863229				711	
GRMZM2G129019	593	456				GRMZM5G863420	860	430			
GRMZM2G129031			890		762	GRMZM5G864266	109	207			
GRMZM2G129090			429	332		GRMZM5G865550			108	81	
GRMZM2G129108	594	457			133	GRMZM5G865943	922	771			
GRMZM2G129154	595	458			107	GRMZM5G866758	451	267			
GRMZM2G129157	596	459			108	GRMZM5G866843	461	310			
GRMZM2G129234	750	694				GRMZM5G867642			639	392	
GRMZM2G129247	749	693				GRMZM5G867768	223	259			
GRMZM2G129288	748	692				GRMZM5G868296			392		162
GRMZM2G129304			757	545		GRMZM5G868423			479	623	
GRMZM2G129344	747	691				GRMZM5G868757	130	95			168
GRMZM2G129361					927	GRMZM5G869321	797				
GRMZM2G129444					929	GRMZM5G869572				751	
GRMZM2G129543			867	515		GRMZM5G869587			893		
GRMZM2G129585			324	262		GRMZM5G870572			396	273	524
GRMZM2G129713			642	447		GRMZM5G870932			961		
GRMZM2G129815					489	GRMZM5G871297	110	208			
GRMZM2G129872	642	747				GRMZM5G871463			780	765	
GRMZM2G129973	880	683				GRMZM5G871995	439	295			
GRMZM2G130018			90	48		GRMZM5G872264			350	979	
GRMZM2G130232					154	GRMZM5G872568					844
GRMZM2G130305					743	GRMZM5G873519	111	77			542
GRMZM2G130356					744	GRMZM5G874112			235	113	
GRMZM2G130375					745	GRMZM5G876022					267

GRMZM2G130586	329	266				GRMZM5G876434					314
GRMZM2G131155			721	974		GRMZM5G876450			265	95	
GRMZM2G131177			720	973		GRMZM5G877110	782	906			
GRMZM2G131202			719	972		GRMZM5G877316	836				
GRMZM2G131309	152	104			608	GRMZM5G877590			364	88	
GRMZM2G131329		970	627	404		GRMZM5G877941	253	366			
GRMZM2G131378		969	626	403		GRMZM5G878257					661
GRMZM2G131473			877	633		GRMZM5G878322					400
GRMZM2G131595			782	634		GRMZM5G878561	708	540	11	9	116
GRMZM2G132144	507	488			40	GRMZM5G878732					223
GRMZM2G132607			789			GRMZM5G878913	899	844			
GRMZM2G132623			788			GRMZM5G878970					835
GRMZM2G132704			787			GRMZM5G879116			535	907	
GRMZM2G132748			786			GRMZM5G879527				942	
GRMZM2G132751			14	34	684	GRMZM5G879872			765		
GRMZM2G132857			671	444		GRMZM5G880508			349	978	
GRMZM2G132880				985		GRMZM5G881088	778	832			
GRMZM2G132898	428	322				GRMZM5G881779	852				218
GRMZM2G132958				869		GRMZM5G881961			617		
GRMZM2G133050					686	GRMZM5G883760	150	137			
GRMZM2G133394			42	25		GRMZM5G885706				495	
GRMZM2G133428	458	349				GRMZM5G886835	170	145			
GRMZM2G133444				754		GRMZM5G886952			707	810	
GRMZM2G133483	459	350				GRMZM5G887922					531
GRMZM2G133563			660	299		GRMZM5G889013			242		
GRMZM2G133568				414		GRMZM5G889372					546
GRMZM2G133652					924	GRMZM5G889418					281
GRMZM2G133718		860				GRMZM5G890473			252	398	
GRMZM2G133806			446			GRMZM5G890820			455	326	633
GRMZM2G133941			850	503		GRMZM5G890938					613
GRMZM2G133969	952					GRMZM5G891098					498
GRMZM2G134049			152	295		GRMZM5G891196	942				
GRMZM2G134402			752			GRMZM5G891247					799
GRMZM2G134563			38	45		GRMZM5G891739			536	908	
GRMZM2G134752				582		GRMZM5G892094			636		
GRMZM2G135045					166	GRMZM5G892991	5	7			
GRMZM2G135300					967	GRMZM5G893851			932	578	
GRMZM2G135314	787					GRMZM5G893936	525	868			102
GRMZM2G135320					404	GRMZM5G894031					434
GRMZM2G135354	786					GRMZM5G894109	370	497			

GRMZM2G135366					405	GRMZM5G894916			764		
GRMZM2G135378			910			GRMZM5G895150	956				440
GRMZM2G135387	839					GRMZM5G895573			293	713	
GRMZM2G135396	840					GRMZM5G895899	962				
GRMZM2G135410	841					GRMZM5G896252					421
GRMZM2G135530					396	GRMZM5G896756	700	622			
GRMZM2G135588			554	591		GRMZM5G896883	20	49			180
GRMZM2G135770			885	467		GRMZM5G896901	220	169			
GRMZM2G136300					834	GRMZM5G897825			616	259	
GRMZM2G136395		810				GRMZM5G898259			648	453	
GRMZM2G136443		808				GRMZM5G898740					464
GRMZM2G136765	868	769				GRMZM5G899123			310	599	
GRMZM2G136769	394	422				GRMZM5G899428	204	240			
GRMZM2G136998				990		GRMZM5G899506	435	486			346
GRMZM2G137151	391	562				GRMZM6G174411			676	871	
GRMZM2G137293					273	GRMZM6G178175	654	679			755
GRMZM2G137326	767					GRMZM6G445841			708	879	
GRMZM2G137435			419	547		GRMZM6G508461	427	550			
GRMZM2G137715	913					GRMZM6G577626		812			
GRMZM2G138259			548	345		GRMZM6G617209	701	714			
GRMZM2G138370			472	662		GRMZM6G685663	189	140			594
GRMZM2G138419					573	GRMZM6G790713	898	843			
GRMZM2G138710	8	6			782	zma-MIR1432					855
GRMZM2G138809			397	425		zma-MIR156g	102	188			626
GRMZM2G138829	45	29				zma-MIR159d				886	
GRMZM2G139324					583	zma-MIR166c			433	241	
GRMZM2G139391	745	768			379	zma-MIR169d			246	290	
GRMZM2G139431			929	530		zma-MIR169e			295	322	735
GRMZM2G139434			751		188	zma-MIR169h			294	321	734
GRMZM2G139513			480	624		zma-MIR169i			44	61	
GRMZM2G139517			49	44		zma-MIR172b			770		
GRMZM2G139550			360	243	522	zma-MIR2118c			505	352	
GRMZM2G139574	137	218			520	zma-MIR394b					296
GRMZM2G139670			926	626		zma-MIR399a			566		682
GRMZM2G139680			697	422		zma-MIR399g			567		683
GRMZM2G139878					486	zma-MIR399h			646	450	
GRMZM2G139973					880	zma-MIR399i			647	451	
GRMZM2G140016			3	15							

Table S3. Complete list of non-redundant predictors using in machine learning

Label	Class	Subclass	Pop
Fst: Ames v. CNNAM	Fst	popFst	NA
Fst: Ames v. EUNAM-Dent	Fst	popFst	NA
Fst: Ames v. EUNAM-Flint	Fst	popFst	NA
Fst: Ames v. USNAM	Fst	popFst	NA
Fst: CNNAM v. EUNAM-Dent	Fst	popFst	NA
Fst: CNNAM v. EUNAM-Flint	Fst	popFst	NA
Fst: CNNAM v. USNAM	Fst	popFst	NA
Fst: EUNAM-Dent v. EUNAM-Flint	Fst	popFst	NA
Fst: EUNAM-Dent v. EUNAM-Flint	Fst	popFst	NA
Fst: EUNAM-Dent v. USNAM	Fst	popFst	NA
Fst: Temperate v. N. Flint	Fst	tempTropFst	NA
Fst: Tropical v. N. Flint	Fst	tempTropFst	NA
Fst: Tropical v. Temperate	Fst	tempTropFst	NA
Ames BRHM DTA bon-corr p-value	Mapping	BRHM	Ames
Ames BRHM DTA median h2	Mapping	BRHM	Ames
Ames BRHM DTS bon-corr p-value	Mapping	BRHM	Ames
Ames BRHM DTS median h2	Mapping	BRHM	Ames
Ames GWAS DTA add. p-value (-log10) - 5PCs	Mapping	QonlyGLM	Ames
Ames GWAS DTA add. p-value (-log10) - No covariates	Mapping	BLUPGLM	Ames
Ames GWAS DTS add. p-value (-log10) - 5PCs	Mapping	QonlyGLM	Ames
Ames GWAS DTS add. p-value (-log10) - No covariates	Mapping	BLUPGLM	Ames
Ames RHM DTA BH-corr p-value	Mapping	RHM	Ames
Ames RHM DTA region h2	Mapping	RHM	Ames
Ames RHM DTS BH-corr p-value	Mapping	RHM	Ames
Ames RHM DTS region h2	Mapping	RHM	Ames
Ames RMIP DTA Density	Mapping	RMIP	Ames
Ames RMIP DTA Prop Models	Mapping	RMIP	Ames
Ames RMIP DTS Density	Mapping	RMIP	Ames
Ames RMIP DTS Prop Models	Mapping	RMIP	Ames
CNNAM BRHM DTA bon-corr p-value	Mapping	BRHM	CNNAM
CNNAM BRHM DTA median h2	Mapping	BRHM	CNNAM
CNNAM BRHM DTS bon-corr p-value	Mapping	BRHM	CNNAM
CNNAM BRHM DTS median h2	Mapping	BRHM	CNNAM
CNNAM GWAS DTA add. p-value (-log10) - Family term	Mapping	QonlyGLM	CNNAM
CNNAM GWAS DTA add. p-value (-log10) - No covariates	Mapping	BLUPGLM	CNNAM
CNNAM GWAS DTS add. p-value (-log10) - Family term	Mapping	QonlyGLM	CNNAM

CNNAM GWAS DTS add. p-value (-log10) - No covariates	Mapping	BLUPGLM	CNNAM
CNNAM RHM DTA BH-corr p-value	Mapping	RHM	CNNAM
CNNAM RHM DTA region h2	Mapping	RHM	CNNAM
CNNAM RHM DTS BH-corr p-value	Mapping	RHM	CNNAM
CNNAM RHM DTS region h2	Mapping	RHM	CNNAM
CNNAM RMIP DTA Density	Mapping	RMIP	CNNAM
CNNAM RMIP DTA Prop Models	Mapping	RMIP	CNNAM
CNNAM RMIP DTS Density	Mapping	RMIP	CNNAM
CNNAM RMIP DTS Prop Models	Mapping	RMIP	CNNAM
EUNAM-Dent BRHM DTA bon-corr p-value	Mapping	BRHM	EUNAM-Dent
EUNAM-Dent BRHM DTA median h2	Mapping	BRHM	EUNAM-Dent
EUNAM-Dent BRHM DTS bon-corr p-value	Mapping	BRHM	EUNAM-Dent
EUNAM-Dent BRHM DTS median h2	Mapping	BRHM	EUNAM-Dent
EUNAM-Dent GWAS DTA add. p-value (-log10) - Family term	Mapping	QonlyGLM	EUNAM-Dent
EUNAM-Dent GWAS DTA add. p-value (-log10) - No covariates	Mapping	BLUPGLM	EUNAM-Dent
EUNAM-Dent GWAS DTS add. p-value (-log10) - Family term	Mapping	QonlyGLM	EUNAM-Dent
EUNAM-Dent GWAS DTS add. p-value (-log10) - No covariates	Mapping	BLUPGLM	EUNAM-Dent
EUNAM-Dent RHM DTA BH-corr p-value	Mapping	RHM	EUNAM-Dent
EUNAM-Dent RHM DTA region h2	Mapping	RHM	EUNAM-Dent
EUNAM-Dent RHM DTS BH-corr p-value	Mapping	RHM	EUNAM-Dent
EUNAM-Dent RHM DTS region h2	Mapping	RHM	EUNAM-Dent
EUNAM-Dent RMIP DTA Density	Mapping	RMIP	EUNAM-Dent
EUNAM-Dent RMIP DTA Prop Models	Mapping	RMIP	EUNAM-Dent
EUNAM-Dent RMIP DTS Density	Mapping	RMIP	EUNAM-Dent
EUNAM-Dent RMIP DTS Prop Models	Mapping	RMIP	EUNAM-Dent
EUNAM-Flint BRHM DTA bon-corr p-value	Mapping	BRHM	EUNAM-

			Flint
EUNAM-Flint BRHM DTA median h2	Mapping	BRHM	EUNAM-Flint
EUNAM-Flint BRHM DTS bon-corr p-value	Mapping	BRHM	EUNAM-Flint
EUNAM-Flint BRHM DTS median h2	Mapping	BRHM	EUNAM-Flint
EUNAM-Flint GWAS DTA add. p-value (-log10) - Family term	Mapping	QonlyGLM	EUNAM-Flint
EUNAM-Flint GWAS DTA add. p-value (-log10) - No covariates	Mapping	BLUPGLM	EUNAM-Flint
EUNAM-Flint GWAS DTS add. p-value (-log10) - Family term	Mapping	QonlyGLM	EUNAM-Flint
EUNAM-Flint GWAS DTS add. p-value (-log10) - No covariates	Mapping	BLUPGLM	EUNAM-Flint
EUNAM-Flint RHM DTA BH-corr p-value	Mapping	RHM	EUNAM-Flint
EUNAM-Flint RHM DTA region h2	Mapping	RHM	EUNAM-Flint
EUNAM-Flint RHM DTS BH-corr p-value	Mapping	RHM	EUNAM-Flint
EUNAM-Flint RHM DTS region h2	Mapping	RHM	EUNAM-Flint
EUNAM-Flint RMIP DTA Density	Mapping	RMIP	EUNAM-Flint
EUNAM-Flint RMIP DTA Prop Models	Mapping	RMIP	EUNAM-Flint
EUNAM-Flint RMIP DTS Density	Mapping	RMIP	EUNAM-Flint
EUNAM-Flint RMIP DTS Prop Models	Mapping	RMIP	EUNAM-Flint
Meta RMIP DTA Density	Mapping	RMIP	NA
Meta RMIP DTA Prop Models	Mapping	RMIP	NA
Meta RMIP DTS Density	Mapping	RMIP	NA
Meta RMIP DTS Prop Models	Mapping	RMIP	NA
USNAM BRHM DTA bon-corr p-value	Mapping	BRHM	USNAM
USNAM BRHM DTA median h2	Mapping	BRHM	USNAM
USNAM BRHM DTS bon-corr p-value	Mapping	BRHM	USNAM
USNAM BRHM DTS median h2	Mapping	BRHM	USNAM
USNAM GWAS DTA add. p-value (-log10) - Family term	Mapping	QonlyGLM	USNAM
USNAM GWAS DTA add. p-value (-log10) - No covariates	Mapping	BLUPGLM	USNAM

USNAM GWAS DTS add. p-value (-log10) - Family term	Mapping	QonlyGLM	USNAM
USNAM GWAS DTS add. p-value (-log10) - No covariates	Mapping	BLUPGLM	USNAM
USNAM RHM DTA BH-corr p-value	Mapping	RHM	USNAM
USNAM RHM DTA region h2	Mapping	RHM	USNAM
USNAM RHM DTS BH-corr p-value	Mapping	RHM	USNAM
USNAM RHM DTS region h2	Mapping	RHM	USNAM
USNAM RMIP DTA Density	Mapping	RMIP	USNAM
USNAM RMIP DTA Prop Models	Mapping	RMIP	USNAM
USNAM RMIP DTS Density	Mapping	RMIP	USNAM
USNAM RMIP DTS Prop Models	Mapping	RMIP	USNAM
Ames Major Allele	Power	Pops	Ames
Ames Major Allele Frequency	Power	Pops	Ames
Ames Num Maj Allele	Power	Pops	Ames
Ames Site Coverage	Power	Pops	Ames
CNNAM Major Allele	Power	Pops	CNNAM
CNNAM Major Allele Frequency	Power	Pops	CNNAM
CNNAM Num Maj Allele	Power	Pops	CNNAM
CNNAM Site Coverage	Power	Pops	CNNAM
EUNAM-Dent Major Allele	Power	Pops	EUNAM-Dent
EUNAM-Dent Major Allele Frequency	Power	Pops	EUNAM-Dent
EUNAM-Dent Num Maj Allele	Power	Pops	EUNAM-Dent
EUNAM-Dent Site Coverage	Power	Pops	EUNAM-Dent
EUNAM-Flint Major Allele	Power	Pops	EUNAM-Flint
EUNAM-Flint Major Allele Frequency	Power	Pops	EUNAM-Flint
EUNAM-Flint Num Maj Allele	Power	Pops	EUNAM-Flint
EUNAM-Flint Site Coverage	Power	Pops	EUNAM-Flint
Hapmap 3.21 Landraces Major Allele	Power	General	NA
Hapmap 3.21 Landraces Major Allele Frequency	Power	General	NA
Hapmap 3.21 MAF	Power	General	NA
Hapmap 3.21 Major Allele	Power	General	NA
Hapmap 3.21 Major Allele Frequency	Power	General	NA
Hapmap 3.21 Teosinte Major Allele	Power	General	NA

Hapmap 3.21 Teosinte Major Allele Frequency	Power	General	NA
NAM Cross-overs	Power	General	NA
NAM Major Allele	Power	Pops	USNAM
NAM Major Allele Frequency	Power	Pops	USNAM
USNAM Num Maj Allele	Power	Pops	USNAM
USNAM Site Coverage	Power	Pops	USNAM

Table S4. Hapmap 3.21 taxa included in the North American Fst subsets.

N. Flint	**Temperate**	**Tropical**
282set_IA2132Goodman-Buckler	ZEAxppRBSDIAAPEI-4	CML69
german_D152	B73	KI3
german_UH009	ZEAxppRBHDIAAPEI-1	CML103
german_UH006	ZEAxppRBFDIAAPEI-3	CML52
german_UH007	CAU478	CML333
282set_Il101	ZEAxujRALDIBAPE	MAIdgiRABDIAAPEI-2
282set_Il14H	ye8112	Tzi8
T24	CAUZHENG58	MAIdgiRAXDIAAPEI-6
282set_IDS28	ZEAxppRCPDIAAPEI-10	CML277
282set_i1677a	ZEAxppRBRDIAAPEI-3	MAIdgiRCKDIAAPEI-9
german_EP1	W64A	Ki11
german_F03802	Mo17	MAIdgiRAYDIAAPEI-7
P39	ZEAxujRAHDIBAPE	MAIdgiRASDIAAPEI-2
282set_F7	ZEAxujRAPDIAAPE	MAIdgiRAPDIAAPEI-12
IA5125	ZEAxppRBCDIAAPEI-2	MAIdgiRAWDIAAPEI-5
282set_CM37	La2-4	Tx601
282set_CO255	zheng32	MAIdgiRCCDIAAPEI-10
XOP2	D340	MAIdgiRAMDIAAPEI-10
german_F2	LH74	MAIdgiRAVDIAAPEI-4
282set_L317	T24	MAIdgiRARDIAAPEI-1
MM-1A	DF24	MAIdgiRAKDIAAPEI-9
german_F283	GY3	NC350
H99	624	MAIdgiRAGDIAAPEI-5
282set_MEF156-55-2	DF20	MAIdgiRACDIAAPEI-3
german_DK105	PHW17	MAIdgiRAADIAAPEI-1
282set_4722	BC4B	MAIdgiRAIDIAAPEI-6
Il14H	ZEAxppRBHDIAAPEI-5	ZEAxppRBYDIAAPEI-2
german_FF0721H-7	H114	NC358
german_CH10	OQ603	MAIdgiRAEDIAAPEI-4
282set_Ia5125	ys06	ZEAxppRCJDIAAPEI-7

Table S5. GO term classification for top 1,000 gene models by BRHM

GO_acc	Term	Class	Ames	CNNAM	EUNAM-Dent	EUNAM-Flint	USNAM
GO:0009408	response_to_heat	AbioticResponse	0	9	0	0	6
GO:0009267	cellular_response_to_starvation	AbioticResponse	0	6	0	6	6
GO:0051716	cellular_response_to_stimulus	AbioticResponse	21	26	20	21	26
GO:0033554	cellular_response_to_stress	AbioticResponse	8	16	8	11	16
GO:0009628	response_to_abiotic_stimulus	AbioticResponse	35	30	30	33	48
GO:0046686	response_to_cadmium_ion	AbioticResponse	8	8	13	11	11
GO:0009409	response_to_cold	AbioticResponse	9	10	6	8	12
GO:0010035	response_to_inorganic_substance	AbioticResponse	10	14	15	12	13
GO:0009416	response_to_light_stimulus	AbioticResponse	13	9	9	14	19
GO:0010038	response_to_metal_ion	AbioticResponse	8	12	14	11	11
GO:0031667	response_to_nutrient_levels	AbioticResponse	5	6	0	6	6
GO:0010033	response_to_organic_substance	AbioticResponse	30	22	27	25	38
GO:0006970	response_to_osmotic_stress	AbioticResponse	13	11	9	15	17
GO:0006979	response_to_oxidative_stress	AbioticResponse	10	7	6	7	9
GO:0009314	response_to_radiation	AbioticResponse	14	9	9	14	19
GO:0009651	response_to_salt_stress	AbioticResponse	11	11	9	13	13
GO:0006950	response_to_stress	AbioticResponse	49	66	51	51	59
GO:0009266	response_to_temperature_stimulus	AbioticResponse	10	18	8	10	15
GO:0009415	response_to_water	AbioticResponse	8	8	6	8	9
GO:0009414	response_to_water_deprivation	AbioticResponse	8	8	6	8	9
GO:0009639	response_to_red_or_far_red_light	AbioticResponse	5	0	0	0	7
GO:0031669	cellular_response_to_nutrient_levels	AbioticResponse	0	6	0	6	6
GO:0009629	response_to_gravity	AbioticResponse	5	0	0	0	0
GO:0042594	response_to_starvation	AbioticResponse	5	6	0	6	6
GO:0009606	Tropism	AbioticResponse	5	0	0	0	0
GO:0016036	cellular_response_to_phosphate_starvation	AbioticResponse	0	0	0	0	5
GO:0071214	cellular_response_to_abiotic_stimulus	AbioticResponse	0	0	0	0	5
GO:0009411	response_to_UV	AbioticResponse	0	0	0	0	5
GO:0009743	response_to_carbohydrate_stimulus	AbioticResponse	0	0	0	0	7
GO:0010646	regulation_of_cell_communication	BioticResponse	5	0	5	5	0

GO:0009966	regulation_of_signal_transduction	BioticResponse	0	0	5	5	0
GO:0023051	regulation_of_signaling_process	BioticResponse	0	0	5	5	0
GO:0009733	response_to_auxin_stimulus	BioticResponse	8	0	7	9	0
GO:0048583	regulation_of_response_to_stimulus	BioticResponse	0	0	5	0	5
GO:0051704	multi-organism_process	BioticResponse	16	30	23	28	21
GO:0048523	negative_regulation_of_cellular_process	BioticResponse	6	7	10	0	7
GO:0042742	defense_response_to_bacterium	BioticResponse	0	16	6	6	0
GO:0050832	defense_response_to_fungus	BioticResponse	0	7	5	6	0
GO:0009611	response_to_wounding	BioticResponse	5	0	0	13	0
GO:0070887	cellular_response_to_chemical_stimulus	BioticResponse	15	15	13	12	15
GO:0071496	cellular_response_to_external_stimulus	BioticResponse	0	8	0	7	8
GO:0031668	cellular_response_to_extracellular_stimulus	BioticResponse	0	8	0	7	8
GO:0071310	cellular_response_to_organic_substance	BioticResponse	12	8	11	9	10
GO:0006952	defense_response	BioticResponse	8	30	19	14	13
GO:0009737	response_to_abscisic_acid_stimulus	BioticResponse	11	12	10	10	12
GO:0009617	response_to_bacterium	BioticResponse	5	20	6	8	6
GO:0009607	response_to_biotic_stimulus	BioticResponse	12	29	19	18	21
GO:0042221	response_to_chemical_stimulus	BioticResponse	51	43	53	47	62
GO:0009719	response_to_endogenous_stimulus	BioticResponse	26	19	24	22	25
GO:0009605	response_to_external_stimulus	BioticResponse	16	12	11	23	18
GO:0009991	response_to_extracellular_stimulus	BioticResponse	5	8	0	7	8
GO:0009620	response_to_fungus	BioticResponse	6	9	8	7	5
GO:0009725	response_to_hormone_stimulus	BioticResponse	23	17	23	22	22
GO:0051707	response_to_other_organism	BioticResponse	11	28	17	17	16
GO:0009753	response_to_jasmonic_acid_stimulus	BioticResponse	5	0	0	0	6
GO:0050896	response_to_stimulus	BioticResponse	85	90	87	83	98
GO:0071365	cellular_response_to_auxin_stimulus	BioticResponse	8	0	0	5	0
GO:0071495	cellular_response_to_endogenous_stimulus	BioticResponse	12	7	10	9	6
GO:0032870	cellular_response_to_hormone_stimulus	BioticResponse	11	6	10	9	0
GO:0009755	hormone-mediated_signaling_pathway	BioticResponse	11	6	9	8	0

GO:0002376	immune_system_process	BioticResponse	0	5	0	0	5
GO:0006955	immune_response	BioticResponse	0	5	0	0	5
GO:0009624	response_to_nematode	BioticResponse	0	0	0	0	5
GO:0009814	defense_response,_incompatible_interaction	BioticResponse	0	0	0	0	5
GO:0045087	innate_immune_response	BioticResponse	0	0	0	0	5
GO:0009751	response_to_salicylic_acid_stimulus	BioticResponse	0	0	0	0	7
GO:0010150	leaf_senescence	Development	0	6	0	0	0
GO:0010260	organ_senescence	Development	0	6	0	0	0
GO:0010149	Senescence	Development	0	6	0	0	0
GO:0008219	cell_death	Development	5	6	5	0	0
GO:0016265	Death	Development	5	6	5	0	0
GO:0048869	cellular_developmental_process	Development	15	11	12	16	21
GO:0009793	embryonic_development_ending_in_seed_dormancy	Development	7	0	0	0	9
GO:0048507	meristem_development	Development	5	0	0	0	8
GO:0009909	regulation_of_flower_development	Development	0	5	5	0	7
GO:0010016	shoot_morphogenesis	Development	0	0	5	0	7
GO:0051093	negative_regulation_of_developmental_process	Development	0	5	0	0	5
GO:0010015	root_morphogenesis	Development	5	9	0	0	9
GO:0060560	developmental_growth_involved_in_morphogenesis	Development	8	0	8	8	7
GO:0000902	cell_morphogenesis	Development	10	0	9	9	8
GO:0048366	leaf_development	Development	0	0	5	0	5
GO:0009790	embryonic_development	Development	8	7	7	0	11
GO:0009887	organ_morphogenesis	Development	0	5	6	0	7
GO:0048580	regulation_of_post-embryonic_development	Development	0	10	7	0	10
GO:0016049	cell_growth	Development	13	0	8	13	15
GO:0000904	cell_morphogenesis_involved_in_differentiation	Development	0	0	6	7	0
GO:0009932	cell_tip_growth	Development	0	0	5	8	0
GO:0048589	developmental_growth	Development	11	0	9	9	12
GO:0009826	unidimensional_cell_growth	Development	8	0	8	8	7
GO:0032502	developmental_process	Development	40	49	44	48	58
GO:0040007	Growth	Development	14	5	11	14	17
GO:0007568	Aging	Development	0	6	0	0	0
GO:0048646	anatomical_structure_formation_involved_in_morphogenesis	Development	0	8	0	0	0

GO:0009886	post-embryonic_morphogenesis	Development	0	6	0	0	0
GO:0048468	cell_development	Development	6	5	8	8	7
GO:0030154	cell_differentiation	Development	7	7	10	13	15
GO:0032989	cellular_component_morphogenesis	Development	10	5	9	11	9
GO:0007275	multicellular_organismal_development	Development	37	45	43	43	51
GO:0009791	post-embryonic_development	Development	23	29	22	19	28
GO:0048569	post-embryonic_organ_development	Development	5	5	0	7	0
GO:0022622	root_system_development	Development	8	11	0	8	11
GO:0048367	shoot_development	Development	8	6	9	5	9
GO:0048731	system_development	Development	19	23	19	21	25
GO:0009888	tissue_development	Development	10	9	7	6	16
GO:0048588	developmental_cell_growth	Development	0	0	5	8	0
GO:0048364	root_development	Development	8	11	0	8	11
GO:0048856	anatomical_structure_development	Development	36	37	34	37	43
GO:0009653	anatomical_structure_morphogenesis	Development	17	16	15	16	23
GO:0007010	cytoskeleton_organization	Development	5	0	0	0	0
GO:0048513	organ_development	Development	19	23	19	21	25
GO:0007389	pattern_specification_process	Development	5	0	0	0	0
GO:0048827	phyllome_development	Development	0	0	5	0	6
GO:0071669	plant-type_cell_wall_organization_or_biogenesis	Development	6	0	0	0	0
GO:0022621	shoot_system_development	Development	8	6	9	5	9
GO:0001708	cell_fate_specification	Development	0	0	0	0	6
GO:0045165	cell_fate_commitment	Development	0	0	0	0	6
GO:0048765	root_hair_cell_differentiation	Development	0	0	0	0	5
GO:0048764	trichoblast_maturation	Development	0	0	0	0	5
GO:0010054	trichoblast_differentiation	Development	0	0	0	0	5
GO:0021700	developmental_maturation	Development	0	0	0	0	5
GO:0010053	root_epidermal_cell_differentiation	Development	0	0	0	0	5
GO:0009913	epidermal_cell_differentiation	Development	0	0	0	0	7
GO:0008544	epidermis_development	Development	0	0	0	0	7
GO:0007398	ectoderm_development	Development	0	0	0	0	7
GO:0009965	leaf_morphogenesis	Development	0	0	0	0	5
GO:0051301	cell_division	Housekeeping	7	0	0	0	0
GO:0007049	cell_cycle	Housekeeping	7	10	5	6	7

GO:0048469	cell_maturation	Housekeeping	0	0	0	0	5
GO:0042546	cell_wall_biogenesis	Housekeeping	0	0	5	0	0
GO:0071555	cell_wall_organization	Housekeeping	7	0	0	0	6
GO:0071554	cell_wall_organization_or_biogenesis	Housekeeping	11	9	6	0	8
GO:0051726	regulation_of_cell_cycle	Housekeeping	0	5	0	0	0
GO:0009987	cellular_process	Housekeeping	259	282	294	257	291
GO:0051179	localization	Housekeeping	56	56	65	50	63
GO:0015031	protein_transport	Housekeeping	15	12	16	13	12
GO:0006605	protein_targeting	Housekeeping	7	5	0	6	5
GO:0055114	oxidation_reduction	Housekeeping	34	49	30	41	42
GO:0015979	photosynthesis	Housekeeping	0	0	9	7	7
GO:0019684	photosynthesis,_light_reaction	Housekeeping	0	0	0	7	0
GO:0009657	plastid_organization	Housekeeping	6	0	7	0	7
GO:0022607	cellular_component_assembly	Housekeeping	8	11	12	16	13
GO:0044085	cellular_component_biogenesis	Housekeeping	12	18	22	17	17
GO:0016043	cellular_component_organization	Housekeeping	36	25	37	41	35
GO:0065003	macromolecular_complex_assembly	Housekeeping	5	7	11	14	11
GO:0043933	macromolecular_complex_subunit_organization	Housekeeping	5	9	14	15	11
GO:0006996	organelle_organization	Housekeeping	22	13	18	21	19
GO:0006461	protein_complex_assembly	Housekeeping	0	6	9	13	8
GO:0070271	protein_complex_biogenesis	Housekeeping	0	6	9	13	8
GO:0034613	cellular_protein_localization	Housekeeping	11	10	12	10	7
GO:0015672	monovalent_inorganic_cation_transport	Housekeeping	7	0	5	0	0
GO:0006812	cation_transport	Housekeeping	9	13	16	11	15
GO:0051641	cellular_localization	Housekeeping	20	19	15	17	11
GO:0070727	cellular_macromolecule_localization	Housekeeping	11	12	12	10	8
GO:0051234	establishment_of_localization	Housekeeping	56	55	62	50	62
GO:0051649	establishment_of_localization_in_cell	Housekeeping	20	17	13	17	10
GO:0046907	intracellular_transport	Housekeeping	17	15	13	13	9
GO:0006811	ion_transport	Housekeeping	12	15	20	12	17
GO:0033036	macromolecule_localization	Housekeeping	17	15	20	16	16
GO:0030001	metal_ion_transport	Housekeeping	0	12	9	7	12
GO:0008104	protein_localization	Housekeeping	15	12	18	14	13
GO:0055085	transmembrane_transport	Housekeeping	20	20	24	12	15
GO:0006810	transport	Housekeeping	56	55	62	50	62

GO:0016192	vesicle-mediated_transport	Housekeeping	9	7	7	10	5
GO:0019725	cellular_homeostasis	Housekeeping	6	7	0	6	14
GO:0048518	positive_regulation_of_biological_process	Housekeeping	5	0	5	8	6
GO:0090066	regulation_of_anatomical_structure_size	Housekeeping	13	0	9	14	15
GO:0008361	regulation_of_cell_size	Housekeeping	13	0	9	14	15
GO:0032535	regulation_of_cellular_component_size	Housekeeping	13	0	9	14	15
GO:0010556	regulation_of_macromolecule_biosynthetic_process	Housekeeping	35	46	47	34	35
GO:0061024	membrane_organization	Housekeeping	0	0	0	5	0
GO:0043085	positive_regulation_of_catalytic_activity	Housekeeping	0	0	0	6	0
GO:0048522	positive_regulation_of_cellular_process	Housekeeping	0	0	0	6	0
GO:0044093	positive_regulation_of_molecular_function	Housekeeping	0	0	0	6	0
GO:0031326	regulation_of_cellular_biosynthetic_process	Housekeeping	35	46	48	36	35
GO:0045454	cell_redox_homeostasis	Housekeeping	5	6	0	5	11
GO:0042592	homeostatic_process	Housekeeping	7	9	5	9	15
GO:0048519	negative_regulation_of_biological_process	Housekeeping	7	11	12	7	12
GO:0050789	regulation_of_biological_process	Housekeeping	68	86	76	73	85
GO:0065008	regulation_of_biological_quality	Housekeeping	19	17	17	24	33
GO:0009889	regulation_of_biosynthetic_process	Housekeeping	35	47	48	36	35
GO:0050790	regulation_of_catalytic_activity	Housekeeping	6	10	7	9	8
GO:0031323	regulation_of_cellular_metabolic_process	Housekeeping	38	50	49	39	40
GO:0050794	regulation_of_cellular_process	Housekeeping	66	75	70	65	75
GO:0010468	regulation_of_gene_expression	Housekeeping	34	49	49	32	38
GO:0060255	regulation_of_macromolecule_metabolic_process	Housekeeping	35	51	49	36	38
GO:0019222	regulation_of_metabolic_process	Housekeeping	38	55	51	41	43
GO:0065009	regulation_of_molecular_function	Housekeeping	6	10	7	9	8
GO:0051239	regulation_of_multicellular_organismal_process	Housekeeping	5	10	7	6	11
GO:0051171	regulation_of_nitrogen_compound_metabolic_process	Housekeeping	35	44	45	34	35
GO:0019219	regulation_of_nucleobase,_nucleoside,_nucleotide_and_nucleic_acid_metabolic_process	Housekeeping	35	44	44	32	34

GO:0080090	regulation_of_primary_metabolic_process	Housekeeping	35	48	47	36	36
GO:0006200	ATP_catabolic_process	Housekeeping	0	6	0	0	0
GO:0046034	ATP_metabolic_process	Housekeeping	0	6	5	0	0
GO:0044106	cellular_amine_metabolic_process	Housekeeping	10	15	10	16	15
GO:0044257	cellular_protein_catabolic_process	Housekeeping	0	7	0	6	6
GO:0006778	porphyrin_metabolic_process	Housekeeping	0	5	0	0	0
GO:0033013	tetrapyrrole_metabolic_process	Housekeeping	0	5	0	0	0
GO:0006818	hydrogen_transport	Housekeeping	6	0	0	0	0
GO:0034220	ion_transmembrane_transport	Housekeeping	5	0	0	0	0
GO:0015992	proton_transport	Housekeeping	6	0	0	0	0
GO:0022900	electron_transport_chain	Housekeeping	0	0	0	0	7
GO:0022402	cell_cycle_process	Housekeeping	0	7	0	0	6
GO:0006886	intracellular_protein_transport	Housekeeping	11	10	11	10	7
GO:0009165	nucleotide_biosynthetic_process	MacroMolecules	5	0	0	0	7
GO:0044264	cellular_polysaccharide_metabolic_process	MacroMolecules	6	0	11	6	8
GO:0019438	aromatic_compound_biosynthetic_process	MacroMolecules	0	5	0	9	8
GO:0006260	DNA_replication	MacroMolecules	0	5	0	0	5
GO:0043632	modification-dependent_macromolecule_catabolic_process	MacroMolecules	0	0	0	5	5
GO:0005996	monosaccharide_metabolic_process	MacroMolecules	5	7	0	0	6
GO:0006457	protein_folding	MacroMolecules	7	5	0	0	5
GO:0051603	proteolysis_involved_in_cellular_protein_catabolic_process	MacroMolecules	0	6	0	6	6
GO:0009309	amine_biosynthetic_process	MacroMolecules	5	5	5	8	10
GO:0051188	cofactor_biosynthetic_process	MacroMolecules	0	0	5	0	7
GO:0065007	biological_regulation	MacroMolecules	77	96	88	87	99
GO:0009058	biosynthetic_process	MacroMolecules	98	103	111	95	107
GO:0005975	carbohydrate_metabolic_process	MacroMolecules	20	27	28	29	27
GO:0043170	macromolecule_metabolic_process	MacroMolecules	151	157	171	153	157
GO:0009056	catabolic_process	MacroMolecules	17	37	25	21	18
GO:0044255	cellular_lipid_metabolic_process	MacroMolecules	11	22	14	14	18
GO:0006631	fatty_acid_metabolic_process	MacroMolecules	7	10	0	6	7
GO:0044275	cellular_carbohydrate_catabolic_process	MacroMolecules	0	0	6	0	0
GO:0016052	carbohydrate_catabolic_process	MacroMolecules	0	6	6	0	0

GO:0044249	cellular_biosynthetic_process	MacroMolecules	94	99	108	90	103
GO:0044260	cellular_macromolecule_metabolic_process	MacroMolecules	132	139	164	138	139
GO:0006259	DNA_metabolic_process	MacroMolecules	6	12	8	9	10
GO:0009059	macromolecule_biosynthetic_process	MacroMolecules	65	72	82	67	67
GO:0043412	macromolecule_modification	MacroMolecules	50	43	65	51	54
GO:0034660	ncRNA_metabolic_process	MacroMolecules	0	5	5	5	5
GO:0019538	protein_metabolic_process	MacroMolecules	95	86	96	89	99
GO:0016070	RNA_metabolic_process	MacroMolecules	43	58	58	51	49
GO:0034637	cellular_carbohydrate_biosynthetic_process	MacroMolecules	8	0	7	7	5
GO:0006073	cellular_glucan_metabolic_process	MacroMolecules	6	0	7	5	7
GO:0044042	glucan_metabolic_process	MacroMolecules	6	0	7	5	7
GO:0044237	cellular_metabolic_process	MacroMolecules	186	216	230	193	215
GO:0051186	cofactor_metabolic_process	MacroMolecules	5	9	7	6	11
GO:0006793	phosphorus_metabolic_process	MacroMolecules	52	42	60	42	49
GO:0006725	cellular_aromatic_compound_metabolic_process	MacroMolecules	6	12	0	10	8
GO:0046483	heterocycle_metabolic_process	MacroMolecules	13	23	20	10	17
GO:0006091	generation_of_precursor_metabolites_and_energy	MacroMolecules	7	7	11	9	14
GO:0044036	cell_wall_macromolecule_metabolic_process	MacroMolecules	0	5	0	0	0
GO:0006006	glucose_metabolic_process	MacroMolecules	0	5	0	0	0
GO:0043414	macromolecule_methylation	MacroMolecules	0	6	0	6	0
GO:0016051	carbohydrate_biosynthetic_process	MacroMolecules	8	5	7	8	5
GO:0044262	cellular_carbohydrate_metabolic_process	MacroMolecules	16	11	17	15	20
GO:0000271	polysaccharide_biosynthetic_process	MacroMolecules	5	5	7	6	0
GO:0005976	polysaccharide_metabolic_process	MacroMolecules	8	6	11	7	8
GO:0044271	cellular_nitrogen_compound_biosynthetic_process	MacroMolecules	8	8	10	12	13
GO:0044267	cellular_protein_metabolic_process	MacroMolecules	78	69	90	76	83
GO:0043687	post-translational_protein_modification	MacroMolecules	44	38	57	44	45

GO:0006468	protein_amino_acid_phosphorylation	MacroMolecules	38	29	45	34	34
GO:0006464	protein_modification_process	MacroMolecules	48	38	62	46	49
GO:0030163	protein_catabolic_process	MacroMolecules	0	8	0	7	6
GO:0008152	metabolic_process	MacroMolecules	247	295	283	274	274
GO:0006807	nitrogen_compound_metabolic_process	MacroMolecules	71	101	94	82	88
GO:0006576	cellular_biogenic_amine_metabolic_process	MacroMolecules	0	0	0	5	0
GO:0005984	disaccharide_metabolic_process	MacroMolecules	0	0	0	5	0
GO:0009311	oligosaccharide_metabolic_process	MacroMolecules	0	0	0	7	0
GO:0006066	alcohol_metabolic_process	MacroMolecules	6	9	8	9	9
GO:0042180	cellular_ketone_metabolic_process	MacroMolecules	21	28	18	24	30
GO:0034641	cellular_nitrogen_compound_metabolic_process	MacroMolecules	15	23	17	21	23
GO:0046486	glycerolipid_metabolic_process	MacroMolecules	0	0	5	7	7
GO:0006650	glycerophospholipid_metabolic_process	MacroMolecules	0	0	5	7	6
GO:0006629	lipid_metabolic_process	MacroMolecules	17	29	20	19	23
GO:0006139	nucleobase,_nucleoside,_nucleotide_and_nucleic_acid_metabolic_process	MacroMolecules	56	83	79	65	69
GO:0006730	one-carbon_metabolic_process	MacroMolecules	6	12	7	12	6
GO:0006644	phospholipid_metabolic_process	MacroMolecules	0	5	7	8	6
GO:0044283	small_molecule_biosynthetic_process	MacroMolecules	15	16	12	20	23
GO:0044281	small_molecule_metabolic_process	MacroMolecules	43	64	47	52	56
GO:0008652	cellular_amino_acid_biosynthetic_process	MacroMolecules	5	5	0	6	8
GO:0006575	cellular_amino_acid_derivative_metabolic_process	MacroMolecules	0	6	0	12	7
GO:0046394	carboxylic_acid_biosynthetic_process	MacroMolecules	10	11	7	12	14
GO:0006720	isoprenoid_metabolic_process	MacroMolecules	0	0	6	0	0
GO:0008610	lipid_biosynthetic_process	MacroMolecules	9	14	13	8	12
GO:0008654	phospholipid_biosynthetic_process	MacroMolecules	0	0	6	0	0
GO:0046777	protein_amino_acid_autophosphorylation	MacroMolecules	0	0	5	0	0
GO:0009207	purine_ribonucleoside_triphospha	MacroMolecules	0	10	7	0	0

	te_catabolic_process						
GO:0009205	purine_ribonucleoside_triphospha te_metabolic_process	MacroMolecules	6	10	9	0	0
GO:0019748	secondary_metabolic_process	MacroMolecules	5	7	5	0	8
GO:0009308	amine_metabolic_process	MacroMolecules	13	18	10	16	16
GO:0019752	carboxylic_acid_metabolic_proces s	MacroMolecules	21	27	17	23	29
GO:0006520	cellular_amino_acid_metabolic_pr ocess	MacroMolecules	10	14	9	11	12
GO:0044248	cellular_catabolic_process	MacroMolecules	16	34	25	19	18
GO:0044265	cellular_macromolecule_catabolic _process	MacroMolecules	7	11	11	10	9
GO:0046700	heterocycle_catabolic_process	MacroMolecules	5	13	10	0	0
GO:0009057	macromolecule_catabolic_process	MacroMolecules	8	14	11	11	9
GO:0032787	monocarboxylic_acid_metabolic_ process	MacroMolecules	10	12	5	10	14
GO:0055086	nucleobase,_nucleoside_and_nucl eotide_metabolic_process	MacroMolecules	9	15	14	6	9
GO:0006753	nucleoside_phosphate_metabolic _process	MacroMolecules	7	14	12	5	9
GO:0009141	nucleoside_triphosphate_metabol ic_process	MacroMolecules	6	10	9	0	0
GO:0009117	nucleotide_metabolic_process	MacroMolecules	7	14	12	5	9
GO:0016053	organic_acid_biosynthetic_proces s	MacroMolecules	10	11	7	12	14
GO:0006082	organic_acid_metabolic_process	MacroMolecules	21	27	17	23	30
GO:0019637	organophosphate_metabolic_proc ess	MacroMolecules	0	5	7	8	6
GO:0043436	oxoacid_metabolic_process	MacroMolecules	21	27	17	23	29
GO:0006796	phosphate_metabolic_process	MacroMolecules	52	42	60	42	49
GO:0016310	phosphorylation	MacroMolecules	50	38	56	39	47
GO:0006163	purine_nucleotide_metabolic_pro cess	MacroMolecules	6	10	11	0	6
GO:0009150	purine_ribonucleotide_metabolic_ process	MacroMolecules	6	10	10	0	5
GO:0044282	small_molecule_catabolic_process	MacroMolecules	5	9	8	0	0
GO:0022613	ribonucleoprotein_complex_bioge nesis	MacroMolecules	0	0	5	0	0
GO:0042254	ribosome_biogenesis	MacroMolecules	0	0	5	0	0
GO:0044238	primary_metabolic_process	MacroMolecules	192	227	224	207	212
GO:0006519	cellular_amino_acid_and_derivati ve_metabolic_process	MacroMolecules	14	17	12	22	18
GO:0070882	cellular_cell_wall_organization_or	MacroMolecules	5	0	5	0	0

	_biogenesis						
GO:0034622	cellular_macromolecular_complex_assembly	MacroMolecules	0	5	8	10	8
GO:0034621	cellular_macromolecular_complex_subunit_organization	MacroMolecules	0	7	11	11	8
GO:0016044	cellular_membrane_organization	MacroMolecules	0	0	0	5	0
GO:0033692	cellular_polysaccharide_biosynthetic_process	MacroMolecules	5	0	7	5	0
GO:0043623	cellular_protein_complex_assembly	MacroMolecules	0	0	6	9	5
GO:0045184	establishment_of_protein_localization	MacroMolecules	15	12	16	13	12
GO:0009250	glucan_biosynthetic_process	MacroMolecules	5	0	0	0	0
GO:0019318	hexose_metabolic_process	MacroMolecules	5	6	0	0	0
GO:0019941	modification-dependent_protein_catabolic_process	MacroMolecules	0	0	0	5	5
GO:0009143	nucleoside_triphosphate_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009166	nucleotide_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009146	purine_nucleoside_triphosphate_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009144	purine_nucleoside_triphosphate_metabolic_process	MacroMolecules	6	10	9	0	0
GO:0006195	purine_nucleotide_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009154	purine_ribonucleotide_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009203	ribonucleoside_triphosphate_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009199	ribonucleoside_triphosphate_metabolic_process	MacroMolecules	6	10	9	0	0
GO:0009261	ribonucleotide_catabolic_process	MacroMolecules	0	10	7	0	0
GO:0009259	ribonucleotide_metabolic_process	MacroMolecules	6	10	10	0	5
GO:0044272	sulfur_compound_biosynthetic_process	MacroMolecules	5	0	0	0	0
GO:0006790	sulfur_metabolic_process	MacroMolecules	6	0	0	0	0
GO:0006511	ubiquitin-dependent_protein_catabolic_process	MacroMolecules	0	0	0	5	5
GO:0006721	terpenoid_metabolic_process	MacroMolecules	0	0	5	0	0
GO:0043648	dicarboxylic_acid_metabolic_process	MacroMolecules	0	0	0	0	5
GO:0006164	purine_nucleotide_biosynthetic_process	MacroMolecules	0	0	0	0	5

GO:0016567	protein_ubiquitination	MacroMolecules	0	0	0	0	6
GO:0032446	protein_modification_by_small_protein_conjugation	MacroMolecules	0	0	0	0	6
GO:0018130	heterocycle_biosynthetic_process	MacroMolecules	0	0	0	0	6
GO:0070647	protein_modification_by_small_protein_conjugation_or_removal	MacroMolecules	0	0	0	0	6
GO:0006732	coenzyme_metabolic_process	MacroMolecules	0	0	0	0	7
GO:0006508	proteolysis	MacroMolecules	20	23	10	18	23
GO:0034645	cellular_macromolecule_biosynthetic_process	MacroMolecules	65	72	81	66	67
GO:0006633	fatty_acid_biosynthetic_process	MacroMolecules	0	6	0	0	5
GO:0042398	cellular_amino_acid_derivative_biosynthetic_process	MacroMolecules	0	0	0	7	6
GO:0009451	RNA_modification	Regulation	0	0	0	5	5
GO:0034470	ncRNA_processing	Regulation	0	0	5	0	0
GO:0010467	gene_expression	Regulation	61	79	78	69	66
GO:0006396	RNA_processing	Regulation	6	11	11	10	10
GO:0071103	DNA_conformation_change	Regulation	5	0	0	5	0
GO:0006413	translational_initiation	Regulation	5	0	5	0	0
GO:0006350	transcription	Regulation	35	44	48	35	34
GO:0006325	chromatin_organization	Regulation	0	5	6	5	6
GO:0051276	chromosome_organization	Regulation	0	5	7	6	7
GO:0006412	translation	Regulation	19	20	21	23	22
GO:0032259	methylation	Regulation	6	12	7	11	5
GO:0032501	multicellular_organismal_process	Regulation	41	45	44	44	56
GO:0051252	regulation_of_RNA_metabolic_process	Regulation	34	42	44	30	34
GO:0006355	regulation_of_transcription,_DNA-dependent	Regulation	34	42	44	30	33
GO:0032774	RNA_biosynthetic_process	Regulation	35	44	48	36	34
GO:0006351	transcription,_DNA-dependent	Regulation	35	44	48	35	34
GO:0050793	regulation_of_developmental_process	Regulation	6	12	8	6	12
GO:0045449	regulation_of_transcription	Regulation	34	42	44	30	33
GO:0019953	sexual_reproduction	Reproduction	6	0	0	0	6
GO:0048608	reproductive_structure_development	Reproduction	18	22	17	14	22
GO:0048316	seed_development	Reproduction	8	6	5	0	10
GO:0009856	pollination	Reproduction	6	0	7	10	0
GO:0003006	reproductive_developmental_process	Reproduction	20	23	21	21	22
GO:0048610	reproductive_cellular_process	Reproduction	0	0	0	7	0

GO:0010154	fruit_development	Reproduction	9	8	6	5	12
GO:0048229	gametophyte_development	Reproduction	6	9	5	6	6
GO:0009555	pollen_development	Reproduction	6	8	5	5	5
GO:0048438	floral_whorl_development	Reproduction	6	5	5	7	0
GO:0048868	pollen_tube_development	Reproduction	0	0	0	9	0
GO:0035295	tube_development	Reproduction	0	0	0	9	0
GO:0000003	reproduction	Reproduction	31	27	28	26	33
GO:0022414	reproductive_process	Reproduction	27	26	24	26	29
GO:0048437	floral_organ_development	Reproduction	5	5	0	7	0
GO:0009908	flower_development	Reproduction	9	11	9	10	10
GO:0048467	gynoecium_development	Reproduction	5	0	0	0	0
GO:0009860	pollen_tube_growth	Reproduction	0	0	0	7	0
GO:0010817	regulation_of_hormone_levels	Signalling	0	0	0	0	5
GO:0007154	cell_communication	Signalling	9	11	8	12	17
GO:0007165	signal_transduction	Signalling	24	21	20	17	20
GO:0007264	small_GTPase_mediated_signal_transduction	Signalling	7	5	8	7	0
GO:0009734	auxin_mediated_signaling_pathway	Signalling	8	0	0	5	0
GO:0023060	signal_transmission	Signalling	24	21	21	17	24
GO:0023052	signaling	Signalling	30	29	24	25	38
GO:0007242	intracellular_signaling_cascade	Signalling	20	16	19	13	17
GO:0023033	signaling_pathway	Signalling	9	0	0	8	6
GO:0023046	signaling_process	Signalling	24	21	21	17	24

Table S6. Proportion of candidate genes in significant regions relative to random random genes within significant regions. Significance based on Bonferroni corrected significance for BRHM method, Benjamini-Hochberg for RHM, and for SNPs selected in at least a certain number of models for the RMIP method. Random genes in significant regions are the average of 100 samples. BRHM results duplicated from Table 5.

	DTA			**DTS**		
RMIP	3	2	1	3	2	1
Ames	1.953 (5)	1.759 (7)	1.497 (14)	1.24 (3)	0.75 (3)	1.086 (11)
CNNAM	1.404 (4)	1.931 (9)	1.304 (17)	2.214 (6)	1.709 (8)	1.239 (17)
EUNAM-Dent	5.556 (4)	4.167 (5)	2.353 (8)	2.02 (2)	2 (2)	2.141 (7)
EUNAM-Flint	0 (0)	0 (0)	0.943 (3)	0 (0)	0 (0)	0.356 (1)
NAM	1.339 (3)	1.016 (5)	0.84 (10)	1.688 (4)	0.899 (4)	1.227 (14)
RHM	0.001	0.01	0.05	0.001	0.01	0.05
Ames	12 (3)	6.25 (3)	2.013 (3)	2.548 (8)	1.224 (10)	1.364 (16)
CNNAM	0 (0)	5.882 (4)	2.844 (6)	0 (0)	1.235 (1)	2.315 (5)
EUNAM.Dent	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
EUNAM.Flint	2 (2)	1.653 (2)	1.117 (2)	1.471 (1)	1.19 (1)	1.55 (2)
NAM	3.704 (3)	2.5 (6)	1.621 (16)	1.432 (6)	1.332 (23)	1.166 (35)
BRHM	0.001	0.01	0.05	0.001	0.01	0.05
Ames	0.968 (6)	1.316 (10)	1.22 (11)	1.297 (8)	1.282 (9)	1.424 (13)
CNNAM	1.579 (6)	1.777 (11)	1.31 (11)	2.174 (8)	2.609 (15)	1.724 (16)
EUNAM-Dent	1.019 (7)	1.046 (8)	1.012 (12)	1.372 (9)	1.171 (10)	1.033 (14)
EUNAM-Flint	1.156 (8)	0.937 (8)	0.915 (13)	1.441 (10)	1.202 (12)	0.904 (13)
NAM	1.146 (28)	1.095 (32)	1.044 (38)	1.139 (29)	1.16 (36)	1.164 (42)

*Candidates in significant regions/Average over 100 samples of non-candidate genes from the maize v3 gene models in significant regions (Number of candidates in significant regions)

Table S7. Significance for downsampled Bonferroni corrected BRHM regions containing candidate gene midpoint

Locus	Ames	CNNAM	EUNAM-Dent	EUNAM-Flint	NAM
CONZ1	1/0.031	1/1	1/1	0.269/0.45	0.28/1
D8	1/1	1/1	0.278/1	0.474/1	0.952/0.366
D9	1/1	0.005/0.002	1/1	1/1	1/1
DLF1	1/1	0.494/0.007	1/1	0.034/0.073	0.014/0.009
FCA	1/1	1/1	1/1	0.144/1	0.019/1
GA2ox1	1/1	1/1	0.076/0.205	0.137/4.62E-05	0.113/0.035
GA2ox1	1/1	0.415/0.044	1/1.10E-04	0.335/0.481	0.07/1.00E-04
GA2ox1	1/1	0.228/1	1/1	1/0.002	0.059/1.02E-04
GIGZ1A	1/0.038	1/1	1/1	0.237/1	1/1
GIGZ1B	1/1	1/1	1/1	1/1	7.15E-05/0.023
GL15	1/1	0.823/1	0.423/1	1/1	1.10E-04/0.045
ID1	1/1	0.996/1	1/0.036	0.091/0.161	0.081/9.94E-04
KN1	1/1	1/1	4.32E-05/1	1/1	1.29E-04/7.15E-05
LDL1	1/1	1/0.099	1/1	1/1	1.67E-04/6.39E-04
LDL1	0.219/0.078	1/1	0.959/1	4.62E-05/9.85E-05	0.094/7.15E-05
LDL1	1/1	1/0.069	0.403/1	0.482/1	0.05/0.132
LUX	6.36E-05/1	7.38E-05/1.46E-04	0.025/1	0.182/0.063	7.15E-05/0.503
miR156	1/1	1/1	1/1	1/1	0.153/1.14E-04
miR172	1.06E-04/0.293	0.892/1	0.632/1	0.247/0.368	0.62/1
MITE	1/6.36E-05	1/1.38E-04	1/1	1/0.464	1/7.15E-05
PhyA1	1/1	1/1	0.026/1	1/1	8.37E-04/7.15E-05

PhyA2	1/1	0.39/0.001	0.126/1	1/1	1/1
PhyB1	1/1	1/1	0.21/0.965	1/1	7.15E-05/1.02E-04
PhyB2	0.001/0.023	1/1	0.075/0.045	0.04/4.62E-05	7.15E-05/7.15E-05
PhyC1	0.766/0.015	1/1	0.765/1	1/1	7.15E-05/7.15E-05
PhyC2	1/1	1/1	9.08E-04/1	1/1	0.005/0.276
ZAP1	0.004/8.42E-04	1/1	1/1	1.94E-04/4.62E-05	0.045/1
ZAP1b	1/1	1/1	1/2.74E-04	1/1	0.317/1
ZCN1	9.12E-05/1	1/1	1/1	1/1	1/1
ZCN10	0.086/6.36E-05	1/1	1/1	1/1	0.163/0.116
ZCN10	1/1	1/1	0.113/0.026	1/1	0.341/1
ZCN11	1/1	1/1	1/1	0.075/1	1/1
ZCN12	1/1	1/1	0.138/0.163	1/0.096	0.005/7.15E-05
ZCN13	1/1	1/1	0.288/1	1/1	0.089/0.034
ZCN14	1/1	1/1	1/0.107	1/1	1/1
ZCN15	1/1	1/1	0.322/0.466	1/1	0.416/1
ZCN16	1/1	1/1	0.144/0.065	1/1	0.199/1
ZCN17	1/1	1/1	1/1	1/1	1/1
ZCN18	1/1	1/1	1/0.059	0.345/1	1.81E-04/0.002
ZCN19	1/1	0.363/0.206	6.34E-05/0.006	5.50E-04/5.63E-04	0.058/7.15E-05
ZCN2	1/1	1/1	1/1	1/0.075	0.118/0.06
ZCN20	1/1	1/1	1/1	1/1	1.36E-04/7.15E-05
ZCN21	1/1	1/1	1/1	0.256/1	0.039/0.002
ZCN24	1/1	0.004/0.002	0.108/0.077	0.019/0.002	1/0.266
ZCN25	0.271/1	1/1	1/0.181	0.066/0.114	1.48E-04/7.15E-

					05
ZCN26	1/1	0.002/9.18E-05	0.571/0.829	1/1	0.037/1.53E-04
ZCN3	1/1	4.69E-04/1	0.301/1	1/1	0.301/0.811
ZCN4	1/1	1/1	1/1	1/1	1/1
ZCN5	1/0.307	1/0.005	2.10E-04/4.32E-05	0.993/1	0.004/0.244
ZCN6	1/0.428	1/1	1/1	1/1	1.29E-04/1.05E-04
ZCN7	1/1	1/0.06	1/1.27E-04	0.741/1	1.07E-04/9.53E-05
ZCN8	6.36E-05/6.36E-05	0.003/4.92E-05	0.033/0.161	2.17E-04/0.259	7.15E-05/1.95E-04
ZCN9	1/1	1/1	1/1	1/1	1/1
ZFL1	1/1	1/1	1/1	1/0.3	0.061/0.005
ZFL2	1/1	1/5.46E-04	1/1	1/1	1.76E-04/9.06E-05
ZmCCA1	1/1	1/1	1/0.487	0.131/0.672	9.15E-04/0.004
ZmCCT	0.021/6.36E-05	7.71E-05/0.003	1/1	4.62E-05/4.62E-05	1.33E-04/7.15E-05
ZmFKF1a	1/1	1/1	1/1	1/1	0.272/1
ZmFKF1b	0.479/0.583	0.366/0.898	1/1	9.39E-05/2.02E-04	9.53E-05/0.184
ZmHD6	1/1	1/1	0.091/6.79E-04	1/1	9.29E-05/1.36E-04
ZmHD6	0.004/8.67E-04	1/1	1/1	1.28E-04/4.62E-05	0.054/1
ZmHD6	1/1	0.125/0.139	0.002/0.129	1/1	1.17E-04/0.136
ZmHD6	1/1	1/1	1/1.23E-04	1/1	0.435/1
ZmHy2	1/1	4.92E-05/0.356	1/1	1/0.077	1/5.70E-04
ZmLD	1/1	0.154/4.92E-05	4.32E-05/0.052	0.133/0.307	1/0.003

ZmLHY1	1/1	0.373/0.084	1/1	1/0.117	0.73/1
ZmLHY2	1/1	1/1	1/1	1/0.877	0.419/1
ZMM4	0.002/0.001	1/1	0.636/1	1/1	7.15E-05/0.042
ZMM5	1/1	0.002/0.005	1/1.34E-04	4.62E-05/8.47E-05	1.72E-04/0.013
ZmPRR37	1/1	1/0.364	1/1	0.017/0.055	0.007/0.002
ZmPRR37.1	1/1	4.92E-05/2.59E-04	0.891/1	0.149/0.034	7.15E-05/7.15E-05
ZmPRR59	0.162/1	1/0.204	1/0.196	1/1	0.336/1
ZmPRR73	1/1	1/1	4.32E-05/2.31E-04	0.022/4.62E-05	7.15E-05/7.15E-05
ZmRAP2.7	6.36E-05/6.36E-05	7.05E-05/2.79E-04	0.041/1	0.164/0.641	7.15E-05/7.15E-05
ZmTOC1.1	1/1	1/1	5.91E-05/0.021	0.418/1	1.02E-04/7.15E-05
ZmTOC1.2	1/1	1/1	1/1	0.094/1	0.028/1
ZmTOC1.3	1/0.127	1/1	0.324/4.34E-04	1/1	1.38E-04/0.055
ZmTOC1.4	1/1	1/1	1/1	1/1	0.077/1.24E-04
ZmTOC1.5	6.36E-05/6.36E-05	1/1	1/1	1/1	1/0.082

REFERENCES

1. M. Shirali *et al*., Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. *Heredity*. **116**, 333–338 (2016).
2. Y. Nagamine *et al*., Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. *PLoS ONE*. **7**, e46501 (2012).
3. F. Tian *et al*., Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet*. **43**, 159–162 (2011).
4. P. J. Brown *et al*., Distinct Genetic Architectures for Male and Female Inflorescence Traits of Maize. *PLoS Genet*. **7** (2011), doi:10.1371/journal.pgen.1002383.
5. E. S. Buckler *et al*., The Genetic Architecture of Maize Flowering Time. *Science*. **325**, 714–718 (2009).
6. S. Salvi *et al*., Conserved Noncoding Genomic Sequences Associated with a Flowering-Time Quantitative Trait Locus in Maize. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11376–11381 (2007).
7. S. Salvi *et al*., Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol*. **48**, 601–613 (2002).
8. S. (1967) Salvi, S. Castelletti, R. Tuberosa, An updated consensus map for flowering time QTLS in maize. *MAYDICA*. **54**, 501–512 (2009).
9. J. M. Thornsberry *et al*., Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet*. **28**, 286 (2001).
10. H.-Y. Hung *et al*., ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc. Natl. Acad. Sci*. **109**, 11068–11069 (2012).
11. M. C. Romay *et al*., Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. **14**, R55 (2013).
12. F. Chardon *et al*., Genetic Architecture of Flowering Time in Maize As Inferred From Quantitative Trait Loci Meta-analysis and Synteny Conservation With the Rice Genome. *Genetics*. **168**, 2169–2185 (2004).
13. N. D. Coles, M. D. McMullen, P. J. Balint-Kurti, R. C. Pratt, J. B. Holland, Genetic Control of Photoperiod Sensitivity in Maize Revealed by Joint Multiple Population Analysis. *Genetics*. **184**, 799–812 (2010).
14. C. Wang *et al*., Mapping QTL Associated with Photoperiod Sensitivity and Assessing the Importance of QTL×Environment Interaction for Flowering Time in Maize. *PLoS ONE*. **5**, e14068 (2010).
15. J. Xu *et al*., The Genetic Architecture of Flowering Time and Photoperiod Sensitivity in Maize as Revealed by QTL Review and Meta Analysis. *J. Integr. Plant Biol*. **54**, 358–373 (2012).

16. H. Giraud *et al*., Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. *Genetics*. **198**, 1717–1734 (2014).
17. E. Durand *et al*., Flowering Time in Maize: Linkage and Epistasis at a Major Effect Locus. *Genetics*. **190**, 1547–62 (2012).
18. F. Chardon, D. Hourcade, V. Combes, A. Charcosset, Mapping of a spontaneous mutation for early flowering time in maize highlights contrasting allelic series at two-linked QTL on chromosome 8. *Theor. Appl. Genet*. **112**, 1–11 (2005).
19. S. Ducrocq *et al*., Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information. *Genetics*. **178**, 2433–2437 (2008).
20. X. Meng, M. G. Muszynski, O. N. Danilevskaya, The FT-like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. *Plant Cell*. **23**, 942–960 (2011).
21. A. Kozaki, S. Hake, J. Colasanti, The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic Acids Res*. **32**, 1710–1720 (2004).
22. J. Colasanti *et al*., The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genomics*. **7**, 158 (2006).
23. T. A. Miller, E. H. Muslin, J. E. Dorweiler, A maize CONSTANS-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta*. **227**, 1377–1388 (2008).
24. M. G. Muszynski *et al*., delayed flowering1 Encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. *Plant Physiol*. **142**, 1523–1536 (2006).
25. K. Bomblies, J. F. Doebley, Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. *Genetics*. **172**, 519–531 (2006).
26. C. Lehermeier *et al*., Usefulness of Multiparental Populations of Maize (Zea mays L.) for Genome-Based Prediction. *Genetics*. **198**, 3–16 (2014).
27. C. Li *et al*., Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. *BMC Biol*. **13** (2015), doi:10.1186/s12915-015-0187-4.
28. K. M. Robbirt, D. L. Roberts, M. J. Hutchings, A. J. Davy, Potential Disruption of Pollination in a Sexually Deceptive Orchid by Climatic Change. *Curr. Biol*. **24**, 2845–2849 (2014).
29. P. Q. Craufurd, T. R. Wheeler, Climate change and the flowering time of annual crops. *J. Exp. Bot*. **60**, 2529–2539 (2009).
30. J. R. Porter, M. A. Semenov, Crop responses to climatic variation. *Philos. Trans. R. Soc. B Biol. Sci*. **360**, 2021–2035 (2005).
31. M. A. Semenov, P. Stratonovitch, F. Alghabari, M. J. Gooding, Adapting wheat in Europe for climate change. *J. Cereal Sci*. **59**, 245–256 (2014).

32. Y. Li *et al*., Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J*. **86**, 391–402 (2016).
33. E. S. Mace, C. H. Hunt, D. R. Jordan, Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet*. **126**, 1377–1395 (2013).
34. E. Antevs, Geologic-Climatic Dating in the West. *Am. Antiq*. **20**, 317–335 (1955).
35. S. E. Metcalfe, S. L. O'Hara, M. Caballero, S. J. Davies, Records of Late Pleistocene-Holocene climatic change in Mexico - a review. *Quat. Sci. Rev*. **19**, 699–721 (2000).
36. K. Fukunaga *et al*., Genetic Diversity and Population Structure of Teosinte. *Genetics*. **169**, 2241–2254 (2005).
37. T. Pyhäjärvi, M. B. Hufford, S. Mezmouk, J. Ross-Ibarra, Complex patterns of local adaptation in teosinte. *arXiv:1208.0634* (2012) (available at http://arxiv.org/abs/1208.0634).
38. M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, N. C. Ellstrand, J. Ross-Ibarra, in *Poster Session* (Portland, OR, 2012).
39. D. R. Piperno, K. V. Flannery, 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**, 2101–2103 (2001).
40. D. R. Piperno, A. J. Ranere, I. Holst, J. Iriarte, R. Dickau, Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci*. **106**, 5019–5024 (2009).
41. L. Perry *et al*., Early maize agriculture and interzonal interaction in southern Peru. *Nature*. **440**, 76–79 (2006).
42. D. G. Smith, G. W. Crawford, in *Northeast subsistence-settlement change, A.D. 700-A.D. 1300*, J. P. Hart, C. B. Rieth, Eds. (New York State Museum/New York State Education Department, Albany, 2002), *New York State Museum bulletin*.
43. C. Rebourg *et al*., Maize introduction into Europe: the history reviewed in the light of molecular data. *TAG Theor. Appl. Genet. Theor. Angew. Genet*. **106**, 895–903 (2003).
44. P.-T. Ho, The Introduction of American Food Plants into China. *Am. Anthropol*. **57**, 191–201 (1955).
45. E. N. Anderson, *The Food of China* (Yale University Press, 1988).
46. P. Dubreuil, M. Warburton, M. Chastanet, D. Hoisington, A. Charcosset, More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements (2006) (available at http://repository.cimmyt.org/xmlui/handle/10883/3026).
47. P. Revilla, P. Soengas, M. E. Cartea, R. A. Malvar, A. Ordas, Isozyme variability among European maize populations and the introduction of maize in Europe. *Maydica*. **48**, 141–152 (2003).
48. Y. Barrière *et al*., Past and prospects of forage maize breeding in Europe. II. History, germplasm evolution and correlative agronomic changes (2006) (available at http://digital.csic.es/handle/10261/42855).

49. J. L. Bennetzen, S. C. Hake, *Handbook of Maize: Its Biology* (Springer Science & Business Media, 2008).
50. E. Evangelou, J. P. A. Ioannidis, Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).
51. A. P. Morris, Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
52. C. A. Reynolds, D. Finkel, A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychol. Rev.* **25**, 97–112 (2015).
53. X. Chen *et al.*, Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. *Am. J. Hum. Genet.* **97**, 708–714 (2015).
54. S. Shifman, J. Kuypers, M. Kokoris, B. Yakir, A. Darvasi, Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* **12**, 771–776 (2003).
55. K. A. Frazer *et al.*, A second generation human haplotype map of over 3.1 million SNPs. *Nature.* **449**, 851–861 (2007).
56. A. Keinan, A. G. Clark, Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants. *Science.* **336**, 740–743 (2012).
57. A. N. Famoso *et al.*, Genetic Architecture of Aluminum Tolerance in Rice (Oryza sativa) Determined through Genome-Wide Association Analysis and QTL Mapping. *PLoS Genet.* **7**, e1002221 (2011).
58. A. Hund *et al.*, QTL controlling root and shoot traits of maize seedlings under cold stress. *Theor. Appl. Genet.* **109**, 618–629 (2004).
59. M. D. McMullen *et al.*, Genetic Properties of the Maize Nested Association Mapping Population. *Science.* **325**, 737–740 (2009).
60. M. W. Ganal *et al.*, A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. *PLoS ONE.* **6**, e28334 (2011).
61. R. J. Elshire *et al.*, A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE.* **6**, e19379 (2011).
62. R. Bukowski *et al.*, Construction of the third generation Zea mays haplotype map. *bioRxiv*, 26963 (2015).
63. D. Money *et al.*, LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3 Bethesda Md.* **5**, 2383–2390 (2015).
64. K. Swarts *et al.*, Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. *Plant Genome.* **7**, 0 (2014).
65. P. J. Bradbury *et al.*, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* **23**, 2633–2635 (2007).
66. S. Takuno *et al.*, Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics.* **200**, 1297–1312 (2015).
67. P. M. VanRaden, Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
68. J. B. Endelman, J.-L. Jannink, Shrinkage estimation of the realized relationship matrix. *G3 Bethesda Md.* **2**, 1405–1413 (2012).

69. E. Rodgers-Melnick, D. L. Vera, H. W. Bass, E. S. Buckler, Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci.* **113**, E3177–E3184 (2016).
70. W. Valdar, C. C. Holmes, R. Mott, J. Flint, Mapping in structured populations by resample model averaging. *Genetics*. **182**, 1263–1277 (2009).
71. E. Rodgers-Melnick *et al.*, Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci.* **112**, 3823–3828 (2015).
72. D. Speed, G. Hemani, M. R. Johnson, D. J. Balding, Improved Heritability Estimation from Genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).
73. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
74. R. A. Fisher, *Statistical methods for research workers*. (Oliver and Boyd, Edinburgh, 1925).
75. M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, in *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing* (USENIX Association, Berkeley, CA, USA, 2010; http://dl.acm.org/citation.cfm?id=1863103.1863113), *HotCloud'10*, pp. 10–10.
76. Z. Dong *et al.*, A Gene Regulatory Network Model for Floral Transition of the Shoot Apex in Maize and Its Dynamic Modeling. *PLOS ONE*. **7**, e43450 (2012).
77. O. N. Danilevskaya, X. Meng, Z. Hou, E. V. Ananiev, C. R. Simmons, A Genomic and Expression Compendium of the Expanded PEBP Gene Family from Maize. *Plant Physiol.* **146**, 250–264 (2008).
78. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
79. Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70 (2010).
80. P. Gauthier *et al.*, RFLP diversity and relationships among traditional European maize populations. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **105**, 91–99 (2002).
81. A. F. Troyer, L. G. Hendrickson, Background and Importance of "Minnesota 13" Corn. *Crop Sci.* **47**, 905 (2007).
82. F. Technow, T. A. Schrag, W. Schipprack, A. E. Melchinger, Identification of key ancestors of modern germplasm in a breeding program of maize. *Theor. Appl. Genet.* **127**, 2545–2553 (2014).
83. W. L. Schurz, *The Manila galleon* (E.P. Dutton & Company, Inc., New York, 1939; //catalog.hathitrust.org/Record/006555178).
84. E. Bauer *et al.*, Intraspecific variation of recombination rate in maize. *Genome Biol.* **14**, R103 (2013).
85. J. Doebley, J. D. Wendel, J. S. C. Smith, C. W. Stuber, M. Goodman, The Origin of Cornbelt Maize: The Isozyme Evidence. *Econ. Bot.* **42**, 120–131 (1988).
86. Y. Lu *et al.*, Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor. Appl. Genet.* **120**, 93–115 (2009).

87. W. Beavis, in *Molecular Dissection of Complex Traits* (CRC Press, Boca Raton).

CHAPTER 4

WHOLE GENOME PREDICTION OF FLOWERING TIME IN ARCHAEOLOGICAL MAIZE INDICATES ANCESTRAL PUEBLOANS ADAPTED MAIZE TO TEMPERATE NORTH AMERICA 2,000 YEARS AGO

Author List

Kelly Swarts (1), Rafal Gutaker (2), Bruce Benz (3), Michael Blake (4), Robert Bukowski (5), James Holland (6), Johannes Krause (7), Melissa Kruse-Peeples (8), Nick Lepak (9), Lynda Prim (8), Cinta Romay(10), Jeffrey Ross-Ibarra (11), Jesus Sanchez (12), Chris Schmidt (8), Verenna Schuenemann (2), Evan Sofro (8), Detlef Weigel (2), Ed Buckler (1,9), Hernan Burbano (2)

1 Dep. of Plant Breeding and Genetics, 175 Biotechnology Bldg., Cornell Univ., Ithaca, NY 14853 USA
2 Max Planck Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, GERMANY
3 Department of Biology, Schollmaier Science and Technology Rm 109
, Texas Wesleyan University, Forth Worth, TX 76105 USA, Fort Worth, TX 76105
4 Department of Anthropology, Vancouver Campus, 6303 NW Marine Drive, V6T 1Z1, Vancouver, BC, Canada
5 Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, 14853, NY USA
6 Department of Crop Science, 1238 Williams Hall, Campus Box 7620, North Carolina State University, Raleigh, 27695-7620, NC, USA
7 Max-Planck-Institut für Menschheitsgeschichte, Kahlaische Strasse 10, 07745 Jena, GERMANY
8 Native Seeds/SEARCH 3584 E. River Rd., Tucson, 85718, AZ USA
9 USDA-ARS, 409 Bradfield Hall, Cornell Univ., Ithaca, 14853, NY, USA
10 Genomic Diversity Facility, Institute of Biotechnology, Cornell University, Ithaca, 14853, NY USA
11 Dept. of Plant Sciences, 262 Robbins Hall, Mail Stop 4, University of California, One Shields Ave, Davis, 95616, CA, USA
12 Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco CP45110, Mexico

Kelly Swarts wrote the manuscript and led the landrace hybrid experiment, most of the analyses and projection of hapmap 3.21 SNPs onto the GBS samples. Rafal Gutaker contributed to the population genetic analyses. Bruce Benz, Michael Blake and R.G. Matson contributed the Turkey Pen samples and archaeological analyses from the

Turkey Pen site. Johannes Krause did the molecular extraction from the Turkey Pen samples, Verenna Schuenemann did the sequencing and Hernan Burbano did the bioinformatics and led the ancient DNA extraction and sequencing. Jeffrey Ross-Ibarra and Jesus Sanchez contributed the pan-American and teosinte samples respectively, and Chris Schmidt and Melissa Kruse-Peeples contributed the Southwestern panel from the Native Seeds/SEARCH core collection. Nick Lepak in New York, Jim Holland in North Carolina and Lynda Prim and Evan Sofro in Arizona oversaw the landrace hybrid experiments. Edward Buckler and and Detlef Weigel

Abstract

Maize (*Zea mays* ssp. *mays*) appears across the southwest US by 4,000 years ago, and people quickly established maize agriculture in the lowland deserts, but full adoption was delayed in the temperate uplands for 1,500-2,000 years. Reduced days to flowering characterizes modern temperate maize and we test the hypothesis that early flowering was a requirement for full agricultural adoption in the uplands. Using diverse modern germplasm to train whole-genome predictions and validated in related modern landraces, we confidently predict that 2,000 year old maize from Turkey Pen Shelter in southeast Utah was already adapted to the uplands. Our results suggest that ancestral southwestern peoples selected for temperate adaptation in situ, primarily from standing variation in the lowland S. Arizona deserts. That prehistoric adaptation in maize relied on even older diversity highlights the importance of standing genetic variation in light of today's unprecedented climate change.

Summary of Results

Archaeological maize from Turkey Pen shelter was adapted to temperate environments by 2,000 years ago

Introduction

The temperate US maize landraces known as Northern Flint are one of two maize germplasm pools that contributed to the development of the agronomically dominant "Cornbelt Dent" of the US (*1, 2*), and were critical for adapting maize to Europe (*3, 4*). Morphological similarity (*5, 6*) and early genetic evidence (*7, 8*) support the origin of the Northern Flints in the southwestern US. The geography of the Americas, constricted through much of the tropics before widening at the environmentally diverse modern border between Mexico and the United States, also points to the Southwest US as a logical context of selection for temperate adaptation in North America by prehispanic agriculturalists. Maize in the Southwest US today is tremendously important culturally to indigenous peoples (*9*), morphologically distinct yet diverse (*5*), and early genetic evidence suggests complex demographic histories, with waves of introduced germplasm (*8*).

The earliest evidence for maize in the archaeological record of the Southwest dates to just before 4,000 BP (before present). This evidence comes from multiple roughly contemporaneous sites situated in the lowland Sonoran desert floodplains of southern Arizona (*10–13*) and also from upland sites (both open air seasonal camps (*14*) and rockshelters (*15*) located in the temperate highlands of the Colorado Plateau, approximately 200 miles from the southern locations. Recent genetic evidence suggests that this early maize took an interior, rather than coastal, route to the Southwest US (*16*). Despite widespread experimentation with maize agriculture in the uplands (*14, 15, 17–21*), maize agriculture was not fully adopted as a primary

subsistence system (inferred from changes in material culture and architectural features and botanical, faunal and bone isotopic analysis) until sometime between 2,000-1,500 BP (*22, 23*). In contrast, people in the southern lowlands rapidly adopted maize agriculture (*11, 13, 24*) and by 3,100 BP peoples in the southern deserts were constructing large, labor intensive irrigation canals (*12, 13*) and terraced gardens (*24*) that imply residential stability, social complexity sufficient to organize a workforce, and a commitment to agricultural production.

The Tucson basin has, on average, 283 frost-free days whereas Blanding, Utah, 450 miles from Tucson and geographically and elevationally proximal to early temperate sites with archaeological maize, has an average of 149 frost-free days, or 2605 average growing-degree days from an April 15th (*25, 26*). In contrast, planting in the Tucson Basin on February 1st generates over 7000 growing-degree says of heat. However, microsite variation is substantial across the Southwest and has been exploited by indigenous farmers since antiquity (*9, 27, 28*). We suspect the delay in full agricultural adoption in the uplands was due to lack of adaptation to temperate environments in the earliest introduced germplasm and that agricultural adoption in the uplands transitioned with the introduction of early-flowering, adapted varieties.

Turkey Pen Shelter is a dry-cave shelter in temperate Southeast Utah (near Blanding) occupied at the cusp of agricultural adoption in the uplands (1,800-2,000 years ago) (*29, 30*). We test the hypothesis of insufficient adaptation in early maize by predicting days to flowering in archaeological Turkey Pen maize from a diverse modern inbred panel, validated in a landrace panel from the southwestern US, and

situate the evolution of Turkey Pen maize in the larger temperate adaptation of maize to the Americas.

Results

Preservation in Turkey Pen maize was excellent, and 14 of the 21 samples tested contained greater than 80% endogenous maize DNA, with an average fragment length of approximately 65bp. In-vitro molecular repair of post-mortem degradation followed by whole genome sequencing the 14 high-endogenous samples to 5-20X coverage allowed us to call SNPs on the 81 million maize SNPs discovered as part of Hapmap3.21 with 10-50% missing data in 14 highly preserved samples.

To situate Turkey Pen in the Southwest US and the Americas, we GBS-genotyped 1,316 individual teosintes and landraces, with a focus on the southwestern US. We clustered these based on geographic location, altitude, and cultural affiliation for subsequent analysis (Table S1). Multidimensional Scaling Analysis (MDS), Admixture, and FST estimates strongly implicate affiliation between Turkey Pen with the temperate Southwest (Puebloan) samples, with a greater measured contribution from teosintes, either *Z. mays* ssp. *parviglumis* (Parviglumis), the wild ancestor of maize, or *Z. mays* ssp. *mexicana*, a closely related upland teosinte. (Figures 1-3).

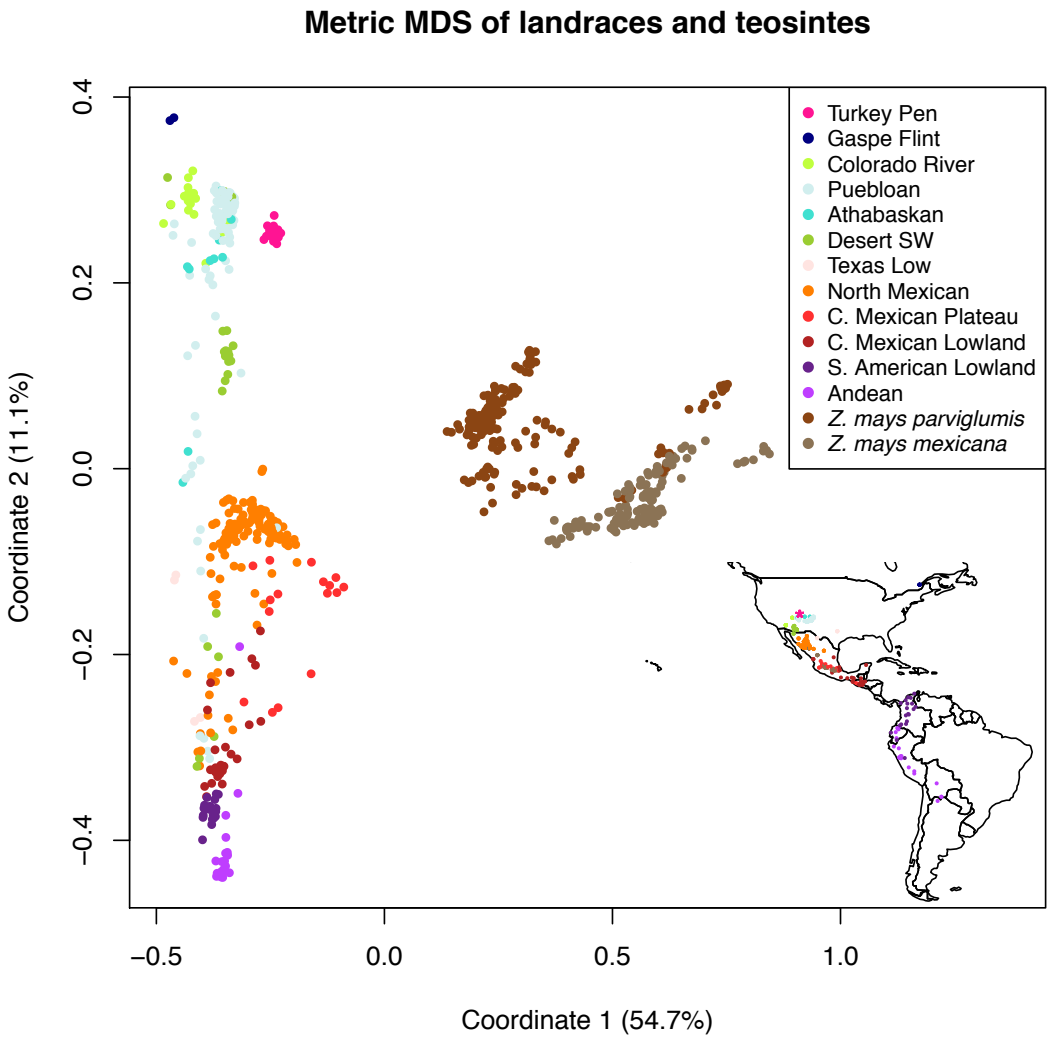


Figure 1. Metric MDS of Turkey Pen and modern landraces and teosinte from 446,999 polymorphic sites shared by GBS and Hapmap 3.21. Groups are defined in Table S1. Geographic/Cultural subsets. The first coordinate separates teosintes from domesticated maize and the second separates domesticated maize geographically across the Americas, with Andean South American maize at one extreme and Gaspe Flint from Quebec, Canada at the other. Turkey Pen maize clusters together, and with modern temperate adapted maize from the Southwest US on the second coordinate, the first coordinate shows that Turkey Pen is pulled towards the modern teosintes, relative to most modern samples with the exception of the Mexican highland samples that contain extensive introgression from *Z. mays* ssp. *mexicana*.

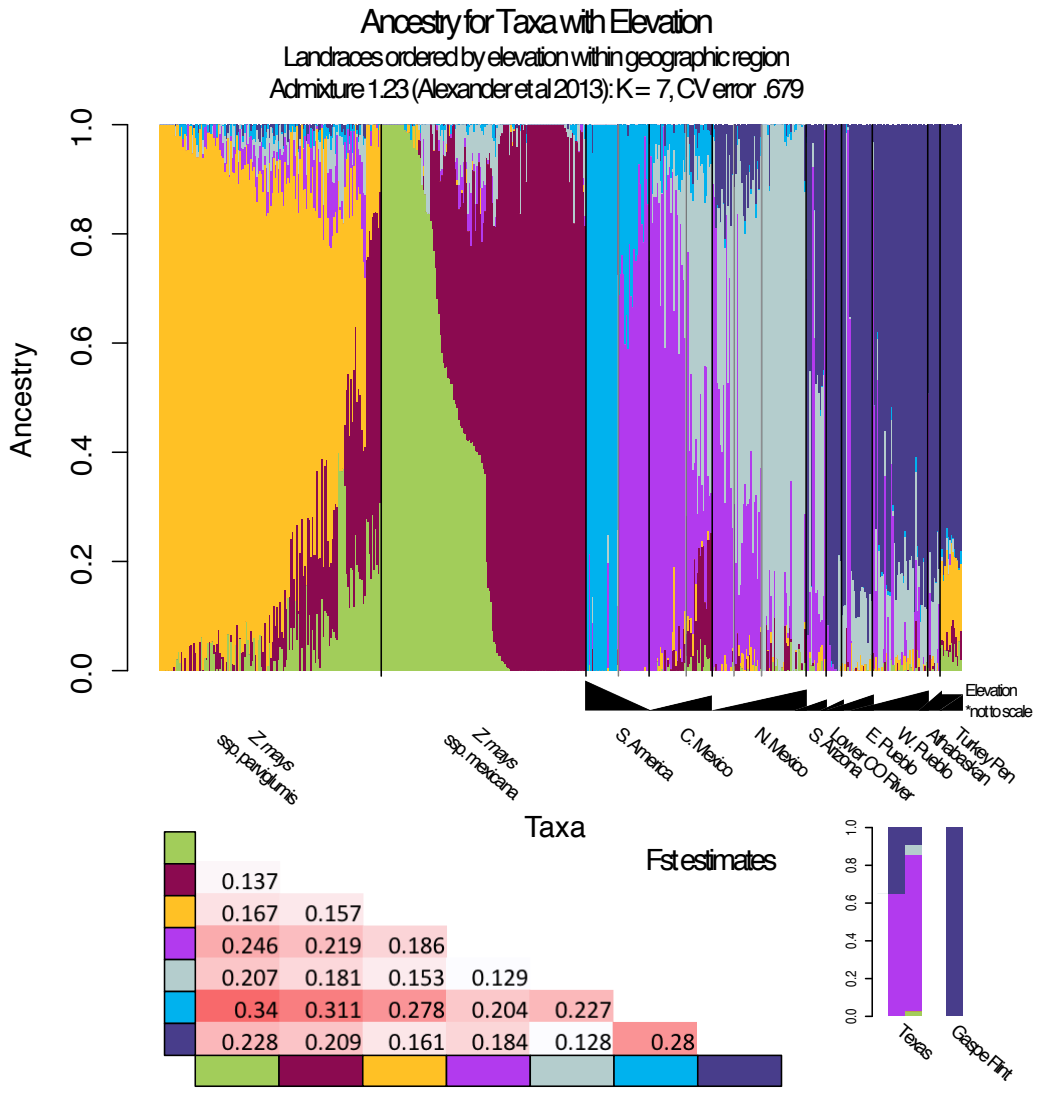


Figure 2. Admixture analysis organized by elevation within groups, K = 7 gave the lowest cross-validated error rate (*31*). Gaspe Flint and Texas samples shown separately for clarity. Turkey Pen is dominated by the dark blue component, similar to the other Puebloan landraces where the dark blue component is dominant. This component is also present as a minor component in North Mexican middle and low elevation samples, but not at high elevation. It is also the dominant component of Gaspe Flint, the Northern Flint landrace from Canada. What is most distinct about the Turkey Pen models with respect to the rest of the landraces is the primarily yellow component, which predominantly maps to Parviglumis, the progenitor teosinte. The light blue component, which is minor in some Mexicana samples, which increases in frequency in the Central Mexican highlands, and is dominant in the North Mexican germplasm, is a minor component in much of the modern Puebloan germplasm, with the exception of the Lower Colorado River, representing primarily fast-cycle sweet corn, and is also less prevalent in the Turkey Pen samples.

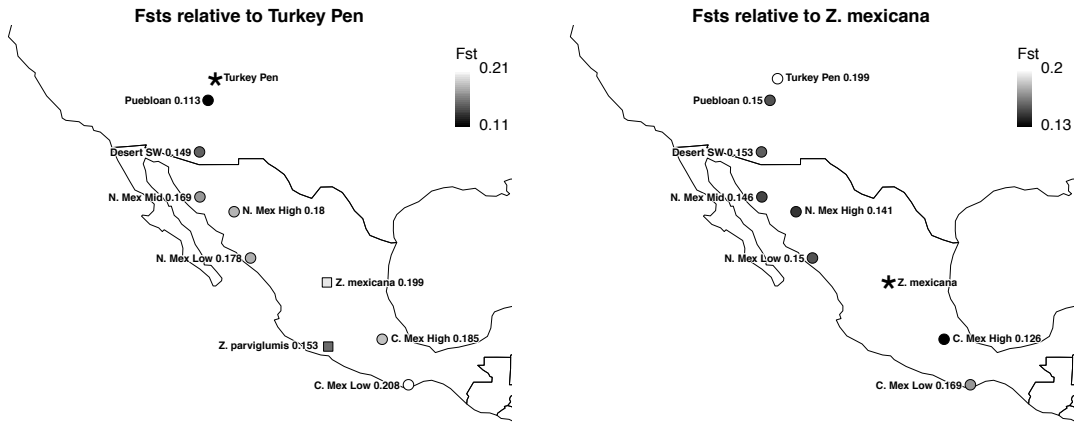


Figure 3. F_{ST} estimates of population differentiation. For the contrast with Mexicana, the Parviglumis samples were excluded to highlight upland teosinte's relatedness with domesticated maize. Turkey Pen is mostly closely related to the Puebloan samples, but, in contrast with modern Puebloan samples, most distant from Mexicana, suggesting the the highland germplasm contributed less to Turkey Pen than to modern Southwestern landraces.

Highland teosinte commonly introgresses into domesticated maize (*32–34*), and chromosomal inversions have been implicated in highland adaptation and early flowering (*35*). With respect to *Z. mays* ssp. *mexicana* (Mexicana), the modern landrace F_{ST}s show a clear pattern where higher elevation populations show reduced differentiation relative to lower elevation samples which is indicative of introgression. (Figure 3) While this pattern holds for modern Puebloan samples, Turkey Pen shows a lowland pattern of relatedness to Mexicana, suggesting that the introgression was not common at Turkey Pen.

The greater contribution from Parviglumis to Turkey Pen but none of the modern landraces cannot be explained by drift, because under drift alone you would expect the least amount of drift near the center of origin, where population sizes have

always been high, and the most in the periphery. Turkey Pen was at the front edge of expansion, and under this model should show the most divergence from Parviglumis.

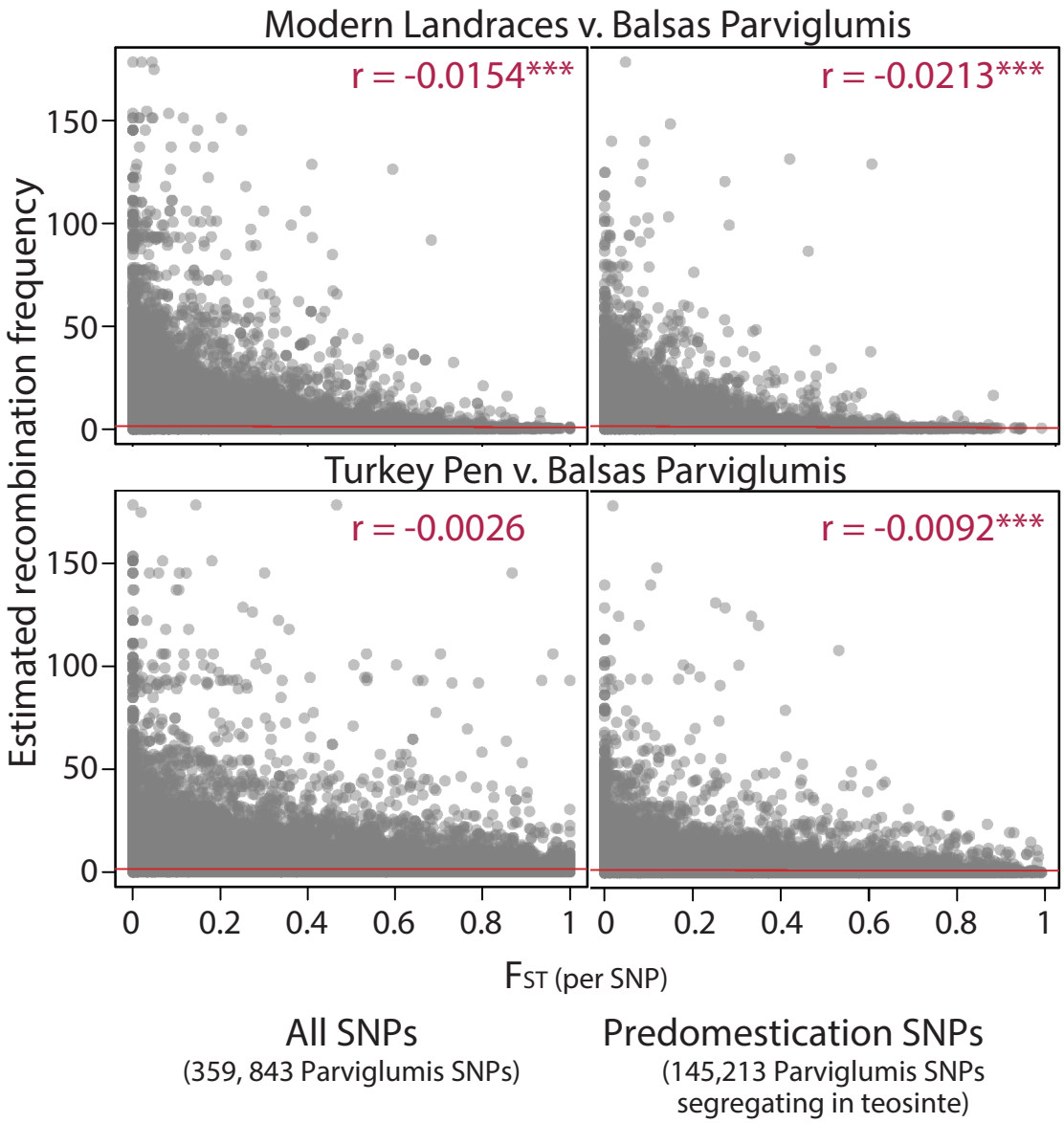


Figure 4. Recombination rate by population differentiation for Balsas Parviglumis vs. all modern landraces and Turkey Pen samples. Recombination rate is the corrected estimates from Rodgers-Melnick et al (2015) (*36*), and F_{ST} is calculated on a per-site basis in vcftools (Weir and Cockerham estimates), where negative values are set to zero (*37*). SNPs contributed by Mexicana (high frequency in Mexicana and modern maize, and low frequency in Parviglumis) excluded to control for differentiation introgression from Mexicana between sample sets.

Purifying selection is efficient in genic regions of maize (*38*), but large pericentromeric regions of low recombination (*36*) limit recombination and Turkey Pen samples would have less time to purge these alleles than all modern samples. We compared F_{ST} between Balsas Parviglumis and both Turkey Pen and all modern samples to recombination rates estimated from the maize NAM population (*36*), and found a small but significant negative correlation for modern samples but not Turkey Pen across all Parviglumis-derived SNPs, suggesting that Turkey Pen samples have purged fewer deleterious SNPs than modern samples (Figure 4). However, both were significant when SNPs private to domesticated maize were excluded, suggesting that most of these mutations arose after domestication, likely during post-domestication demographic expansion, and do not explain the increased relatedness of Turkey Pen samples to Parviglumis.

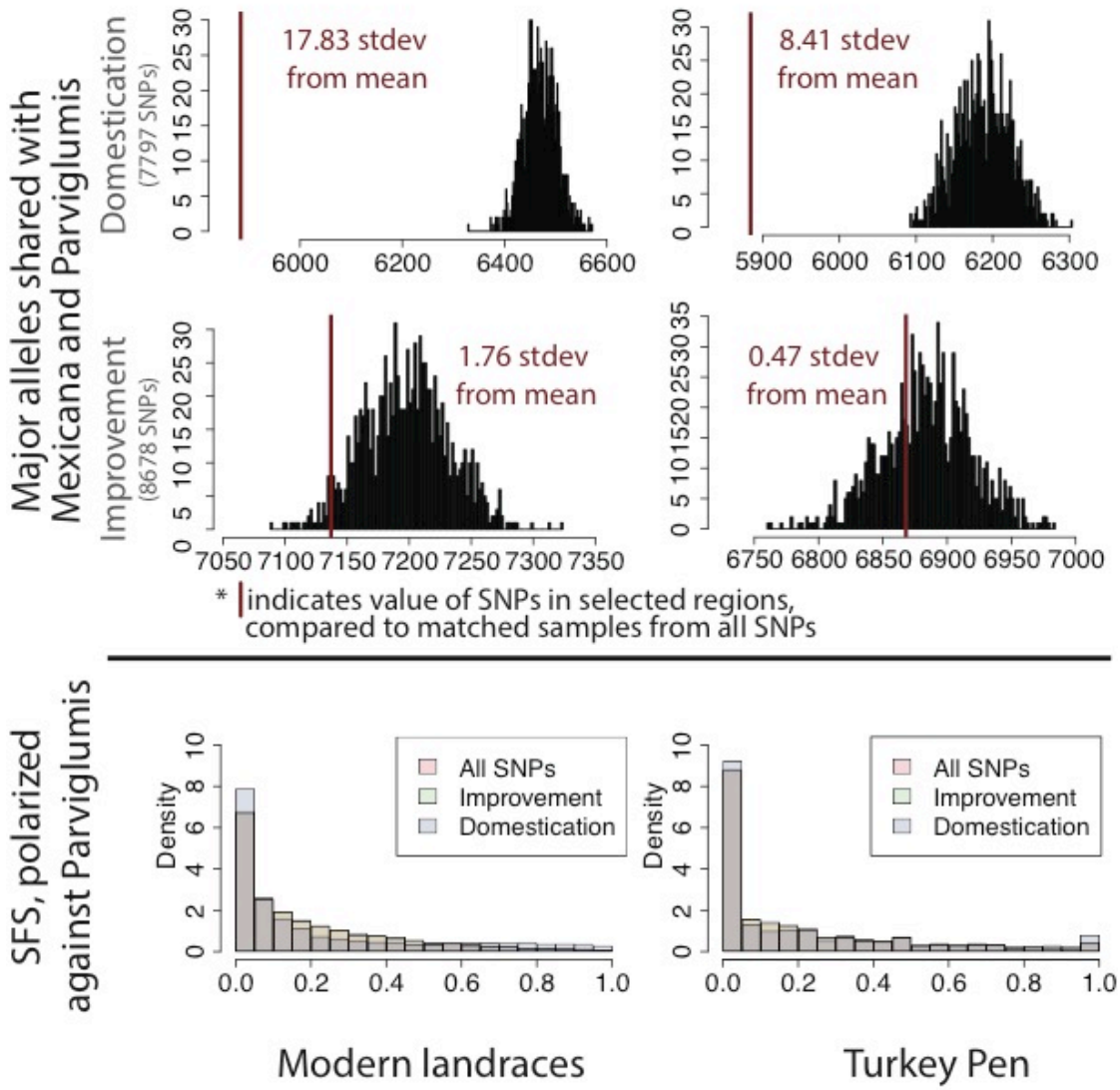


Figure 5. Proportion of shared major alleles with Parviglumis and Mexicana (to control for differential Mexicana introgression) at domestication and improvement loci, relative to matched SNPs iteratively drawn 1000 times with replacement from the whole genome. The test statistic is the distance from the measured value in the selected region to the mean of the sampled distribution, in standard deviations estimated from the sampled distribution. Enrichment is calculated as the difference between Turkey Pen and modern maize for the test statistic. Site frequency spectra were polarized against Balsas Parviglumis.

Patterns in positive selection leading towards modern maize better explain the increased affiliation between Turkey Pen and Parviglumis. We examined the relative abundance of major alleles shared with teosinte at putative improvement and

domestication regions from Hufford et al. (2012) (*39*) and found that modern maize landraces were 2.12-fold enriched for derived SNPs in domestication regions and 3.76-fold enriched for improvement regions, and allele frequencies for improvement loci are higher in modern landraces than in Turkey Pen (Figure 5).

Days to flowering is a complex additively inherited trait (*40*), and critical for temperate adaptation (*40–42*). We generated days to flowering predictions for Turkey Pen and a subset of the landraces representing 80 southwest landrace accessions, phenotyped in nine replicate/environments (Figure S1). Broad-sense heritability for days to silking (DTS) was 0.90 and 0.89 for days to anthesis (DTA) in the phenotyped landrace panel. Predictions from RR-BLUP were generated from the inbred Ames diversity panel (*43*), which was shown to have good cross predictive ability in American germplasm (Figure S2). Ames predicts the landrace panel with a Pearson correlation of 0.68 (prediction accuracy 0.72) for both DTS and DTA between predicted and observed values (Figures 6, S3-4). The predictions for Turkey Pen from Ames, which cannot be evaluated directly for accuracy but should have similar prediction accuracy to the equally diverged and closely related modern landraces, fall near the mean of the Puebloan germplasm, and are equivalent to modern temperate lines, but not as early as the modern Northern Flint inbred lines. Turkey Pen and the Puebloan samples flower within one day of each other and one week earlier than average for southwestern desert samples (Figure S5). The only subgroup that flowers earlier than Turkey Pen is fast-cycle sweet corn from the Lower Colorado River, and the sample of Athabaskan germplasm included in the trial. The mean DTS for Turkey

Pen is 73.6 days as predicted by Ames (103 days in Blanding); assuming 55 days to physiological maturity (*44*), Turkey Pen maize would take 158 days from planting to full dry-down. At 149 average frost-free days, Turkey Pen maize was marginally adapted, and probably would have been harvested and dried off the plant to avoid frost.

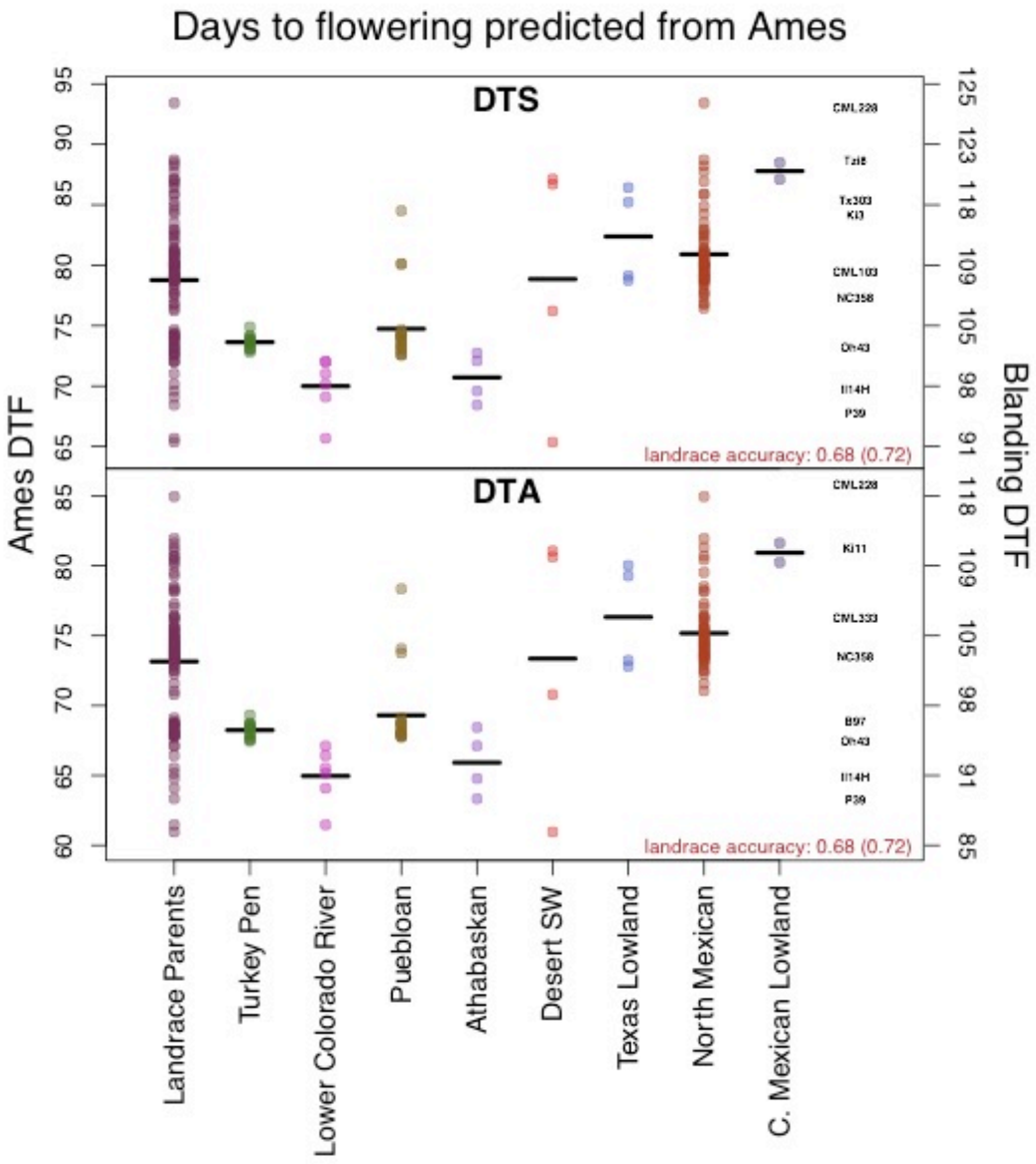


Figure 6. Predicted flowering for Turkey Pen and phenotyped southwestern landraces using the SNPs that overlap between Hapmap 3.21 and GBS. Predictions generated from RR-BLUP in TASSEL (*45*) by masking test phenotypes, and Pearson correlation between observed and masked test phenotypes (landraces only) generated using cor.test() in R. Prediction accuracy in parentheses. Landraces split according to gross group classifications in Table SX. Geographic/Cultural subsets. Turkey Pen flowers at an equivalent time to the temperate NAM founders B97 and Oh43, but not as early as the northern Flints, Il14H and P39. 60-85 days in Ames phenotypes corresponds to 1,284-2,018 growing degree days, and 65-95 days corresponds to 1,320-2,099 growing degree days.

The most selected alleles associated with temperate adaptation should be tagged by the SNPs with the highest F_{ST}s between domesticated maize from the Central Mexican Lowlands and Gaspe Flint, a northern Flint landrace from Quebec. We used the top 1,000 of these high F_{ST} SNPs to understand relatedness between populations, specifically at SNPs associated with temperate adaptation. High F_{ST}s at these sites relative to random matched subsets from background SNPs imply different haplotypes between these populations at temperate loci. (Figure 7) Turkey Pen does not show elevated F_{ST}s against Sonoran desert lowland or the temperate southwestern samples, Puebloan and Athabaskan, but it does against all of the Mexican samples, and also the fast-cycle Lower Colorado River samples, suggesting an *in situ* temperate adaptation in the Southwest US.

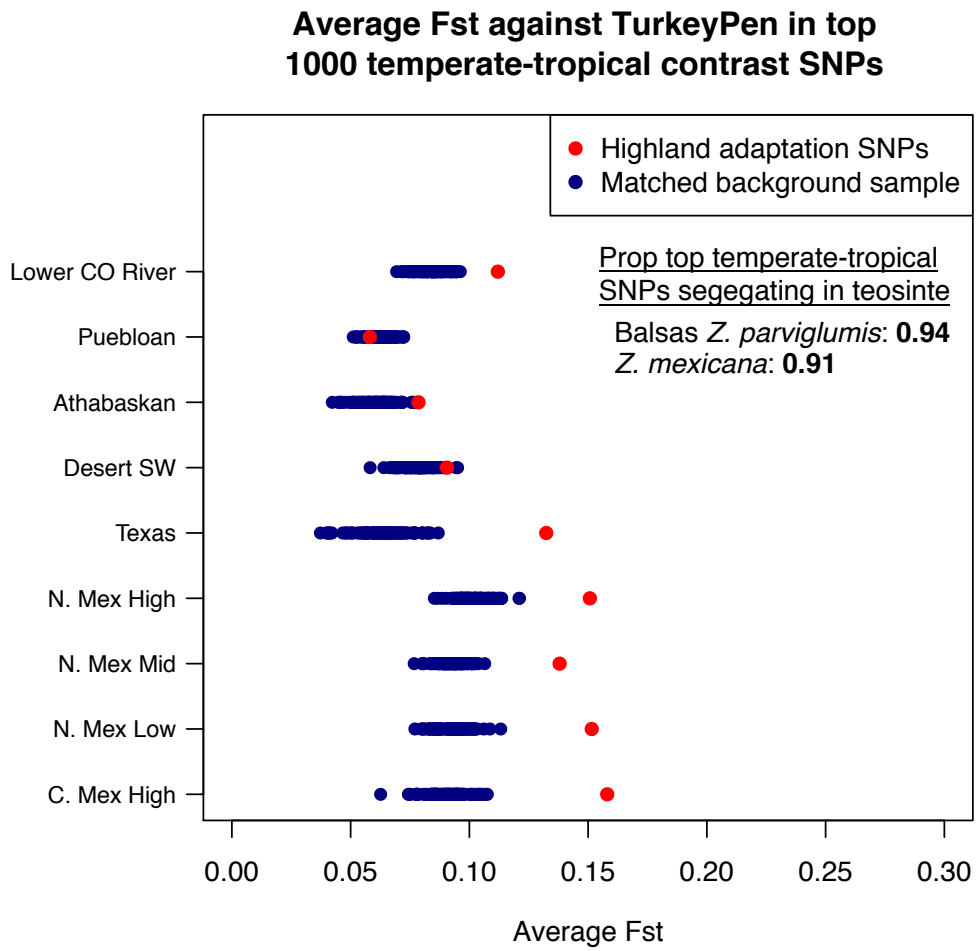


Figure 7. F_{ST}s between Turkey Pen and other populations at the top 1,000 F_{ST} SNPs from the contrast between the Central Mexican Lowlands and Gaspe Flint. These SNPs should carry the greatest signal for temperate adaptation. If F_{ST} at these SNPs falls within the expectation from background SNPs, the population in question shares alleles with Turkey Pen.

F_{ST}s relative to Gaspe Flint show the degree to which the landrace populations are differentiated from the modern Northern Flints (Figure 8). Genomic F_{ST} estimates show a gradual increase in relatedness moving north from Mexico, with an increase in relatedness in the modern southwestern populations. Turkey Pen does not show this pattern, but is one of the most diverged samples, suggesting that Turkey Pen germplasm was not on its own directly ancestral to Gaspe Flint. However, looking at

the top differentiated SNPs for temperate adaptation, Turkey Pen is as equally differentiated as the southwestern desert samples but less so than the modern Puebloan samples.

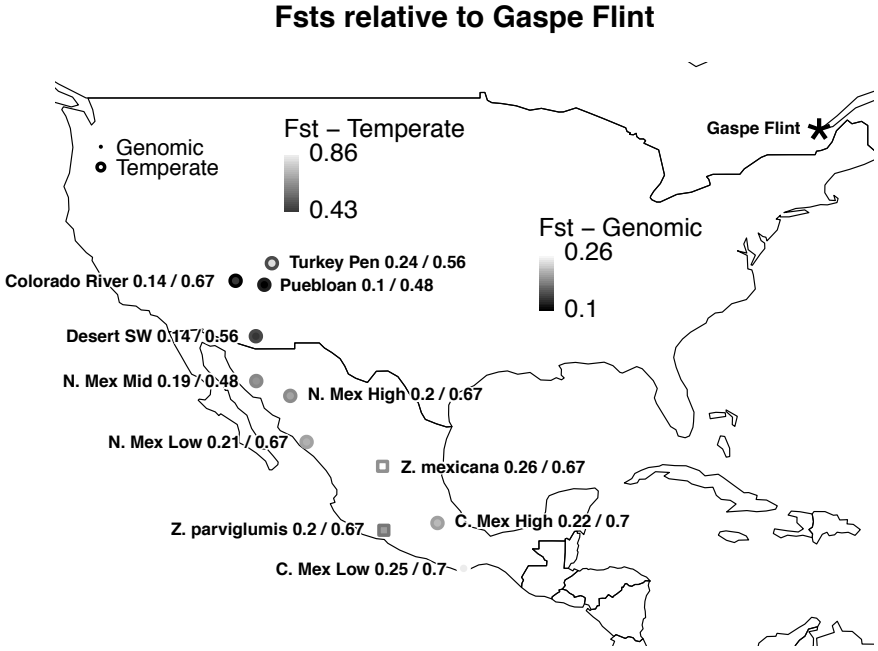


Figure 8. Weighted F_{ST}s relative to Gaspe Flint across all SNPs (circle fill) and at the top 1,000 F_{ST} SNPs from the contrast between the Central American Lowlands and Gaspe Flint (circle outline). None of the southwestern populations are very closely related to Gaspe at the top temperate adaptation SNPs, but the modern Pueblos are the closest. Turkey Pen and the southwestern desert samples are equally distantly related at the top SNPs. At genomic SNPs, all modern southwestern samples are most closely related to Gaspe, and Turkey Pen is one of the most differentiated.

To evaluate the age of high F_{ST} temperate variants, we also looked at the proportion of the top temperate adaptation SNPs that segregate in teosinte and find that 94% segregate in Balsas Parviglumis and 91% in the highland teosinte, Mexicana, suggesting that temperate adaptation was selected from very old standing variation. This result is reinforced by case studies at major flowering and photoperiod loci,

ZmCCT, ZCN8 and the Vgt1 locus (including Rap2.7 and the upstream regulatory MITE insertion) where early flowering variants are segregating in teosinte. (Figure S6, Table S2). Additionally, the genome-wide estimate of nucleotide diversity in Turkey Pen is 0.0044, is comparable to estimates of diverse modern landraces of 0.0049 (*39*), and suggests a minimal population bottleneck, consistent with long-term *in situ* adaptation.

Our evidence suggests that Turkey Pen was adapted with respect to flowering, but we also wanted to investigate if Turkey Pen maize could have played a similar role to modern southwestern maize with respect to culture and nutrition. Maize is foundational to life for modern southwestern cultures, and it is critically important for the proper performance of religious ritual to have the correct types of maize (*9*). It would also have been important to maintain sufficient nutrients as maize increased in the diet and, especially in the more temperate regions of the Southwest, in the winter months when alternative foraged plant foods would have been more limited.

Flour-type maize is culturally important in the modern Southwest (*9*) and *sugary1* is a starch pathway locus that others have suggested may have driven the introduction of novel Mexican germplasm evidence in modern samples but not Turkey Pen into the Southwest US (*8, 16, 46*). We looked at this locus in Turkey Pen and modern landraces and inbred lines for evidence of influence of this germplasm introduction (Figure 9). The low-diversity, selected haplotype contains dent and flours, which are good for tortilla, and includes the modern Puebloan flour variety from Santo Domingo; however, Turkey Pen does not cluster with this pool and probably had flint or pop-type endosperm. The floury haplotype is present as a heterozygote in one

sample, so the presence of this haplotype, previously noted in contemporaneous Southwest US samples (*46*), is confirmed at Turkey Pen, but at low frequency. We also confirmed that Turkey Pen does not have either the North Mexican (N561S) or the Southwestern (W578R) mutation that confers a sweet endosperm type (*47*), but Turkey Pen haplotypes are basal to modern sweet corn.

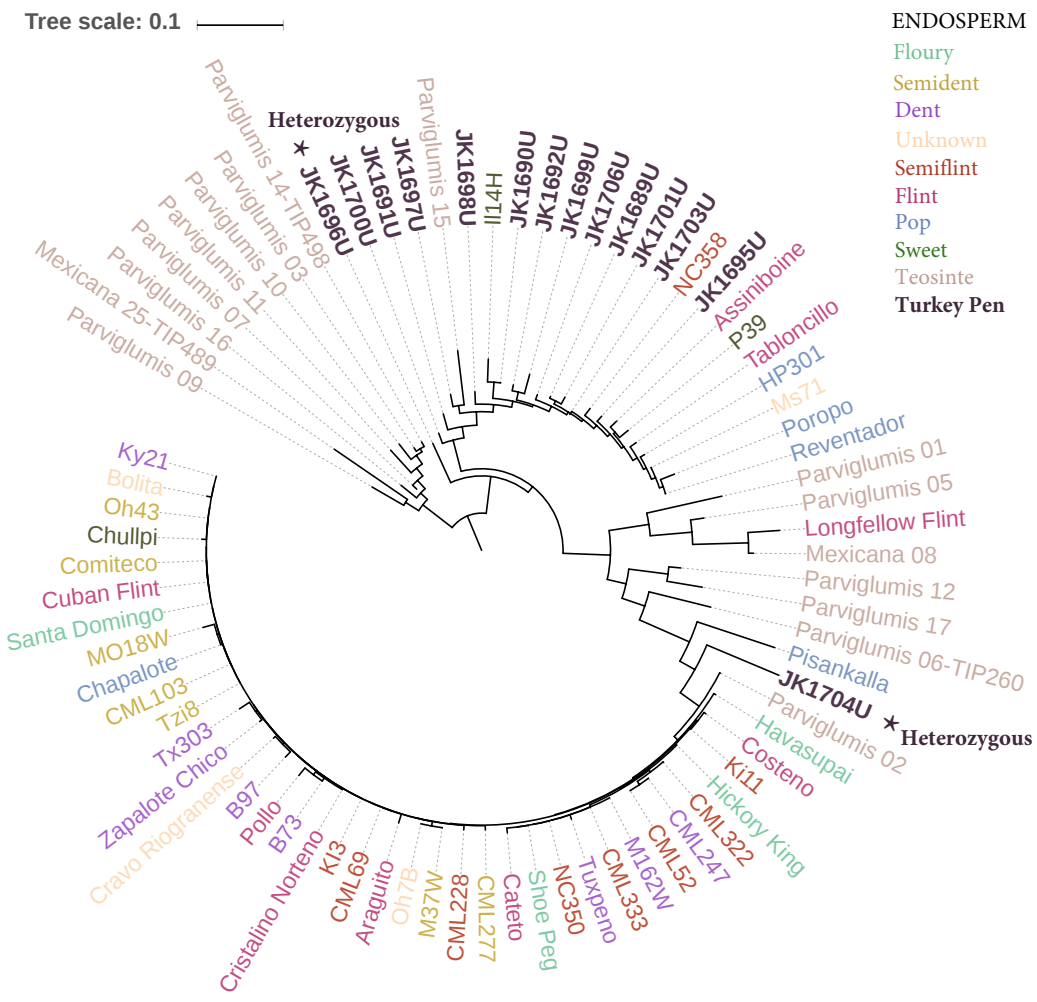


Figure 9. Neighbor-joining tree at the *su1* locus for Turkey Pen, US NAM founders, inbred landraces and teosintes in Hapmap 3.21 based on 227 SNPs(*45, 48*). Most of the Turkey Pen samples fall outside of the selected haplotype, suggesting Turkey Pen maize has a pop or flint endosperm type. One individual is heterozygous for the selected haplotype, and other is heterozygous for a teosinte haplotype so the selected haplotype was present at Turkey Pen, consistent with previous results for contemporaneous sites(*46*), but in Turkey Pen it is segregating at very low frequency. Turkey Pen samples are basal to modern sweet corn represented in the dataset.

Kernel color is very important for indigenous groups in the modern Southwest US, as it is critical for the proper performance of religious rituals(*9*); while there are

insufficient comparative samples in Hapmap3.21 to robustly evaluate flavonoids for blue and red kernel color, we evaluated Turkey Pen haplotypes for the gene *yellow1 (y1)*, *Y1* codes for phytoene synthase and controls flux into the carotenoid pathway; a recent gain of function mutation leads to the accumulation of carotenoids (*49*). Yellow color is caused by the accumulation of carotenoids in the endosperm, which are also vitamin A precursors critical for human development (*50*), making the evolution of yellow corn an important adaptation for human nutrition. Turkey Pen is clearly segregating for the conserved yellow-producing haplotype, suggesting that some of the maize from Turkey Pen was yellow (Figure 10). Additionally, a subset of Turkey Pen samples cluster directly basal to the selected yellow haplotype, suggesting that the mutation for yellow kernel color may have originated in the Southwest US.

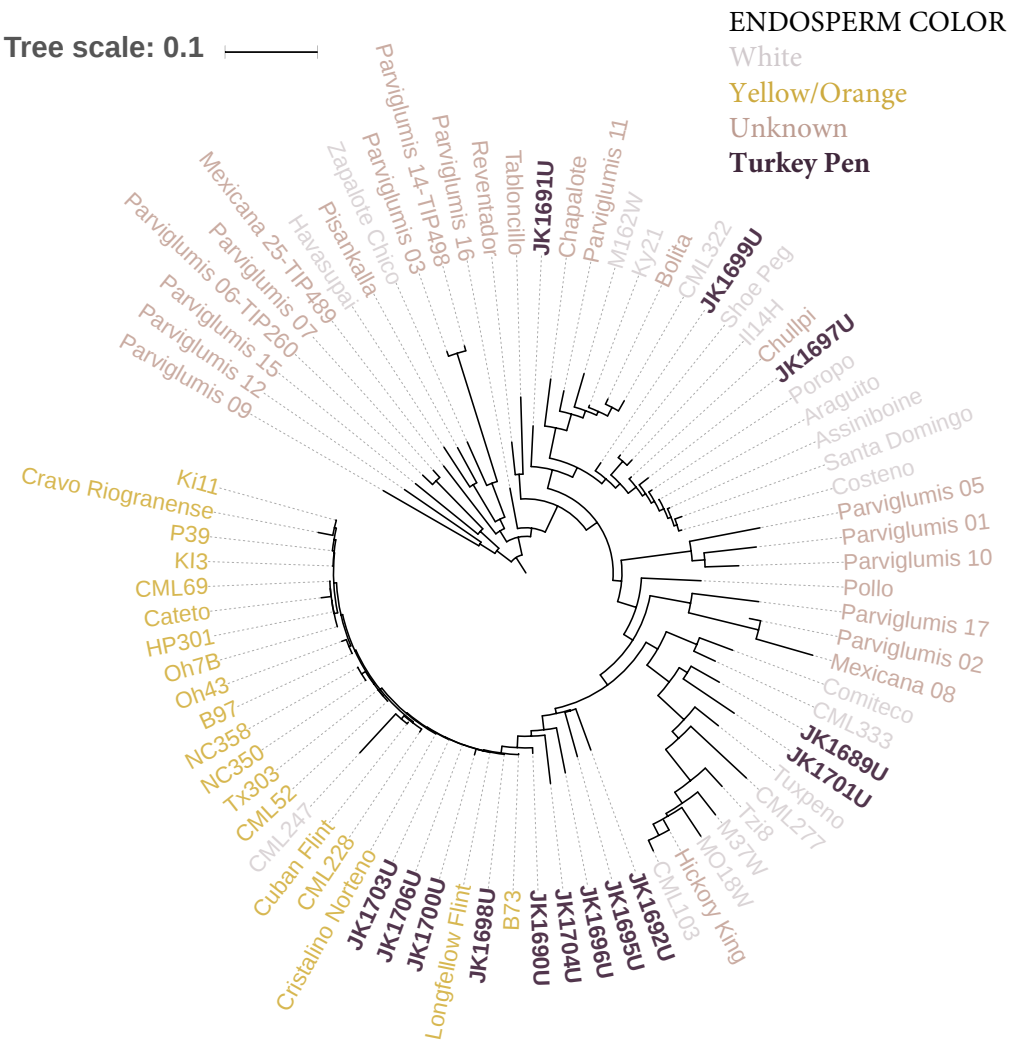


Figure 10. Neighbor-joining tree(*45, 48*) at the *y1* locus, including the upstream region that contains previously implicated transposon insertions (*49*), for Turkey Pen, US NAM founders, inbred landraces and teosintes in Hapmap 3.21 based on 216 SNPs. Four Turkey Pen samples fall within the selected yellow haplotype at the *y1* locus.

Discussion

MDS, admixture, and population statistics confirm that Turkey Pen maize is most closely related to modern Puebloan maize and at least partially ancestral to the agronomically important Northern Flint germplasm, allowing for good whole genome

prediction of flowering in modern southwestern landraces from, and high-confidence prediction of Turkey Pen itself. This is true even though the maize in the southwest has undergone additional selection and demographic shifts not observed in Turkey Pen, as indicated by the lack of a highland Mexican component in Turkey Pen (based on admixture components and FST with Mexicana). Advances in ancient genomics and excellent preservation at Turkey Pen allowed us to predict previously unknowable traits from archaeological samples, from both well-known biosynthetic pathways such as carotenoids and starch but also complex traits such as days to flowering that are critical for understanding adaptation. Widespread archaeological evidence for upland maize in the temperate regions before the adoption of full agriculture suggests that ancestral Puebloan people selected for temperate adapted maize for millennia before assembling the proper combination of alleles. Analysis of SNPs important for temperate adaptation supports the archaeological interpretation that temperate adaptation was selected for in situ in the southwestern US, and, additionally, that temperate adaptation was selected from standing variation present in teosinte. This case study suggests that maize has an outstanding adaptive capacity, which will be important going forward in the face of unprecedented climate change.

Methods

Turkey Pen extraction and sequencing

Isolation of DNA from archaeological specimens of maize was carried out in clean laboratory facilities using stringent precautions typical for working with aDNA: UV-treatment of reagents and equipment, employment of contamination prevention

gear such as masks, full-body suits and laminar flow hoods. Samples were ground in 2.0 ml tubes with metal pestle.

DNA isolation followed the protocol from Yoshida et al (*51*), following a similar protocol from Rohland et al (*52*). Each tube with ground tissue was filled with 1.2 ml of freshly prepared PTB lysis buffer (1% SDS, 10 mM Tris, 10 mM EDTA, 0.4 mg/ml Proteinase K, 5 mM NaCl, 2.5 mM PTB and 50 mM DTT), sealed with parafilm and incubated at 37°C overnight in the vertical rotor (12 rpm). On the next day the lysate was centrifuged (10 min, 14000 rpm), supernatant was mixed with 325 µl of DNEasy® Plant Mini kit P3 buffer in a fresh 2 ml tube and incubated on ice for 5 minutes. After another centrifugation step (5 min, 14000 rpm) the supernatant was transferred to QIAshredder spin column. From that step isolation was carried out according to the QIAgen DNEasy® Plant Mini kit protocol with following modifications:

a) Due to large volume of DNA solution in column binding step, reservoir extension is recommended. To that end, a sterile funnel has been forced onto DNeasy® Mini spin column and placed into 50 ml Falcon tube. The whole volume of DNA isolate was transferred onto the funnel and centrifuged for 10 minutes at 1300 rpm.

b) After washing the columns with AW2 buffer we introduced 2 consecutive steps of 1-minute dry spin (14000 rpm).

c) DNA was eluted from the columns in two steps of centrifugation (1 min, 14000 rpm) with 50 µl of AE buffer each.

Two independent genomic libraries were constructed from 20 µl of DNA extract for each sample; i) libraries without enzymatic removal of cytosine to thymine

(C-to-T) substitutions typical of ancient DNA (aDNA) associated damage. These libraries were used to present positive evidence of the authenticity of the reads derived from archaeological specimens; ii) libraries treated with uracil glycosylase, which remove the excess of C-to-T substitutions. These libraries are devoid of aDNA-associated damage and were used for deep sequencing (*53*). Shotgun libraries were constructed following published protocol (*54*) with modifications suggested in (*55*). Libraries were amplified for 10 cycles with unique combination of two indexing primers (*56*). The quality of libraries was tested in two consecutive RT-qPCR reactions (prior to and after indexing amplification step) and in Bioanalyzer. Non-UDG treated libraries were sequenced on Illumina MiSeq instrument. UDG-treated libraries were modified by addition of USER™ enzyme at blunting step and sequenced.

The 15 samples with >60% endogenous maize were selected for further sequencing, where half of them were sequenced three times in Tuebingen on an Illumina HiSeq 3000 platfor, and the remainder were sequenced at Cornell, once on an Illumina HiSeq 3000, and once on an Illumina NextSeq platform. SNPs were called from the resulting reads against 81 million variants discovered as part of Hapmap 3.21 (*57*).

Modern landrace and teosinte accessions and sequencing

All modern landraces and teosintes were genotyped using GBS at the Genomic Diversity Facility at Cornell University (*58*). Ninety-five landraces from across the Americas were chosen in the laboratory of Dr. Jeffery Ross-Ibarra (UC-Davis), and

were previously published in Takuno et al (*59*). The modern teosintes were contributed by Dr. Jesus Sanchez, and were collected from wild populations. The Northern Flints were contributed by Dr. Jason Wallace, and were genotyped as part of the mini-maize project (*60*). Ninety-five of the landraces from the Southwest US derive from publically available USDA accessions. The remaining 391 landraces from the Southwest US and North Mexico were part of the maize core collection held by Native Seeds/SEARCH in Tucson, Arizona, chosen to represent the diversity of the accessions. 446,999 GBS sites that overlap with Hapmap 3.21 positions and major/minor allele (to merge the GBS and Turkey Pen datasets) were used for population genetic analysis and genomic prediction.

MDS

MDS generated using cmdscale() in R from an IBS distance matrix calculated in TASSEL (*45*) where Hapmap 3.21 matches the major and minor allele with GBS 2.7. Before distance calculations, all heterozygotes were set to either homozygote with a probability of 0.5 to control for uneven taxon coverage. Sites are additionally filtered for a minimum of 10 called genotypes for each site. Taxa are filtered so that only one individual per named accession remains and filtered for 0.2 minimum site coverage after site filtering, resulting in 838 taxa.

Admixture

Ancestry estimates were generated from 2.7 GBS SNPs that overlap positionally with Hapmap 3.1 using the software package Admixture 1.23 (*31*). Admixture

generates similar results to the commonly used software STRUCTURE, but is faster due the use of log-likelihood estimation for ancestry estimation, and additionally provides cross-validated error estimates for each K-populations tested. The dataset begins with all Mexicana and Parviglumis, modern landrace and Turkey Pen genotypes (1,008 taxa); other teosinte genotypes are not included because they fall outside of the coalescent. Before running admixture:

1. Third allele states were removed due to constraints of the admixture software.
2. Only the highest coverage individual from each accession was retained. If possible (in all but two cases), this individual was representative in that 60% of the accession is represented in closest individuals, when ranked by genetic distance.
3. Sites are filtered for Hardy-Weinburg equilibrium using a chi-squared test for equality at alpha = .05, based only on modern landrace genotypes with a read depth of 9-300 to mitigate heterozygote-undercalling.
4. Monomorphic sites and those with a site coverage of 0.5 are removed.
5. Taxa are filtered for taxon coverage of 0.3 (resulting in 1,000 taxa) and the lower coverage individual is removed for any pair of individuals with a genetic distance less than 0.15.

These filters result in 533 taxa and 14,885 sites. K = 7 was used because it had the lowest cross-validated error rate.

Landrace Hybrid design and phenotyping

Landrace hybrids were generated at the Native Seeds/SEARCH farm in Patagonia, AZ in the summer of 2013 by crossing as many individuals as possible from 108 landrace accessions onto the inbred line PHZ51, chosen for good agronomic performance and disease resistance. PHZ51 plantings were staggered once a week for a month to encourage nicking, and the earliest estimated landrace accessions were planted in the second planting. These are the same individuals that were genotyped for population genetic analysis.

In the summer of 2015, the progeny from 111 hybrid individuals representing 80 accessions and 10 PhZ51xB73 checks were grown in an 11x11 alpha lattice incomplete block design (using agricolae() in R), replicated three times each in three separate locations, Aurora, NY, Clayton, NC, and Native Seeds/SEARCH research farm in Patagonia, AZ for a total of nine replicates. Because the landrace parent is heterozygous, the hybrid progeny are segregating for the landrace parent individual's genotype. Thus, we phenotyped days to anthesis and days to flowering on a modified individual rather than a plot basis. Each day, from the day the first individual in a plot flowered to the last day, the number of plants flowering was recorded, giving a plot-level variance for anthesis and silking respectively, but not for anthesis-silking interval. Field-corrected phenotypes were calculated using the lmer() package in R, according to the model;

Days to flowering ~ 1 + (1 | Genotype) + (1 | Location) + (1 | Genotype:Location) + (1 | Rep:Location) + (1 | Block:(Rep:Location))

Broad-sense heritability was calculated as $H = \sigma^2_G/(\sigma^2_G+ \sigma^2_{GE} /e+ \sigma^2_e /re)$. Prediction accuracies for the landrace hybrid panel was calculated as the Pearson correlation, conservatively, and the Pearson correlation divided by the square-root of the broad-sense heritability

Cross-population prediction for days to flowering

Cross population prediction were performed to better understand how populations were related to each other with respect to shared genetic architecture for days to flowering. Cross population prediction was performed using RR-BLUP as implemented in TASSEL (*45*). The training and test populations were combined in a single kinship (similarity) matrix, calculated using Centered-IBS (after Van Raden (*61*), Endelman and Jannink (*62*)), and phenotypes for the test population masked so that the model was trained solely from the training population phenotypes. Predictions from the resulting model for the test phenotypes were then correlated with the true phenotypes (the "prediction accuracy") in R using the Pearson method in cor.test() (stats).

Conversion of predicted days to flowering in Ames to days to flowering in Blanding, Utah was made based on conversion through the Growing Degree Days (GDD) for the Ames population (*43*). For conversion of GDDs to days to anthesis in Blanding, we assume a planting date of April 15th (*63*), and found the "normal" number of days to achieve the Ames GDD using https://www.pioneer.com/home/site/ca/agronomy/tools/gdu/

FSTs

Weighted FSTs were generated in vcftools from 464,432 consensus sites between Hapmap 3.21 and GBS, with a minimum per-site taxa count of 50.

Neighbor-joining trees

Neighbor-joining trees generated on Hapmap 3.21 SNPs in Turkey Pen, US NAM founders, inbred landraces and teosintes for SNPs within the canonical coding sequence. SNPs filtered for polymorphic sites, taxa filtered for minimum 0.3 taxon coverage. Heterozygous genotypes randomly set to either homozygote at probability 0.5 before generation of IBS distance matrix and neighbor-joining tree in TASSEL (*45*) to account for differential coverage.

Nucleotide diversity calculation

Per-site nucleotide diversity (pi) was calculated from the Hapmap 3.21 SNPs in vcftools (*37*), and total number of variant and invariant sites were called from the merged and sorted bam files in samtools. The 0.0044 figure reported is the sum of per-site nucleotide diversity divided by the total number of invariant and variant sites.

Supplemental

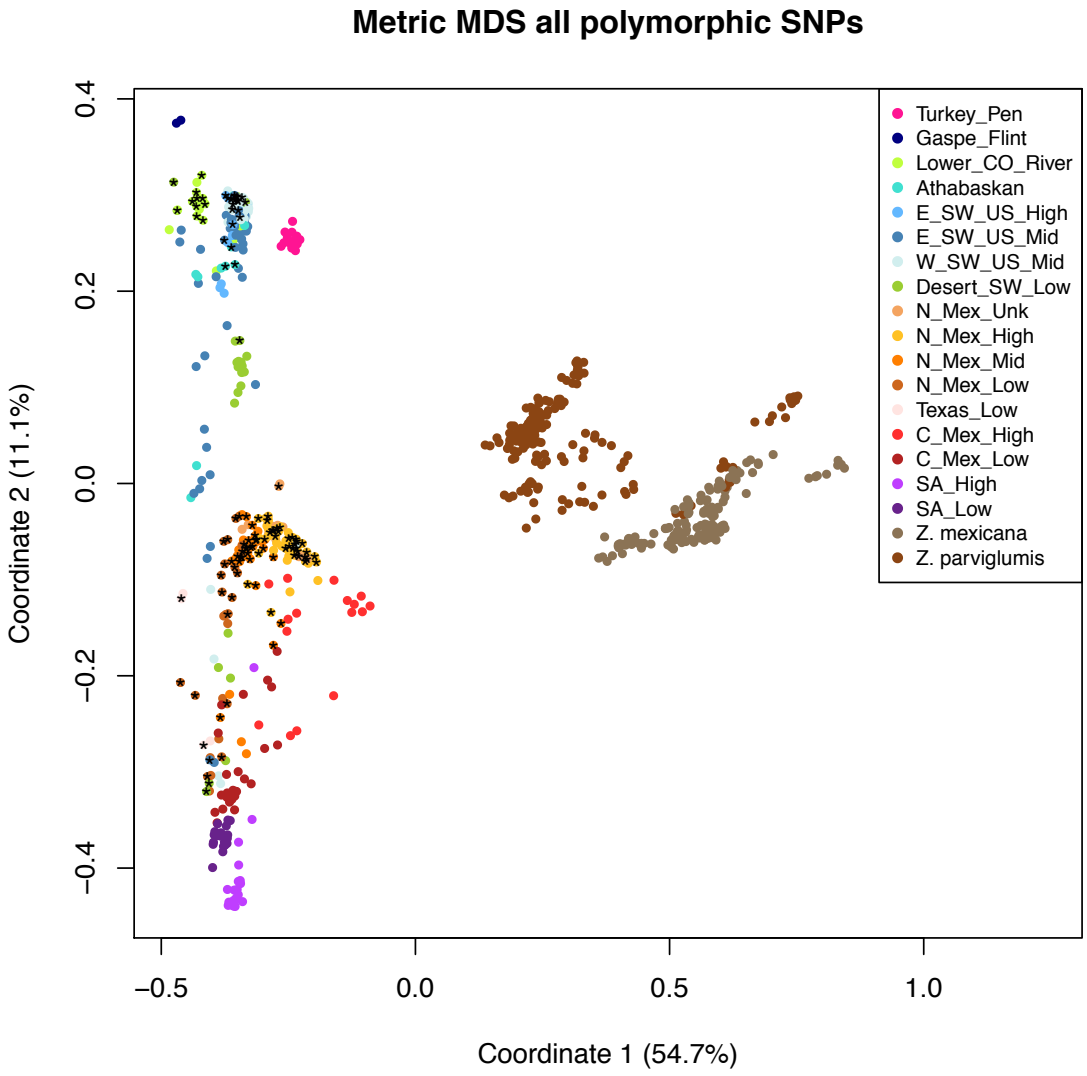


Figure S1. Distribution of landrace parents in black. The landrace parents were chosen to cover the temperate tropical gradient from populations in the Southwest US and North Mexico

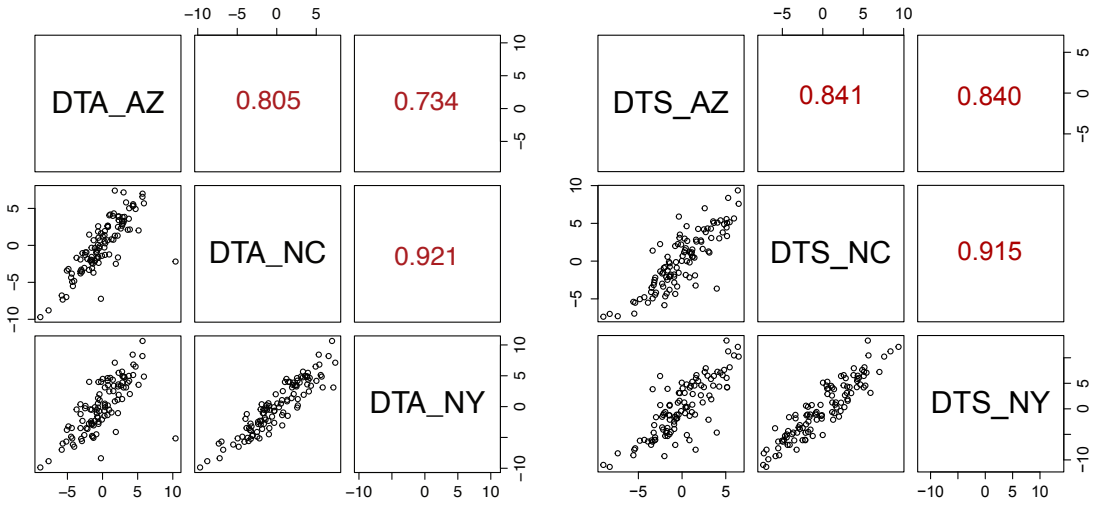


Figure S2. Correlations between environments for days to anthesis and silking in the landrace hybrids

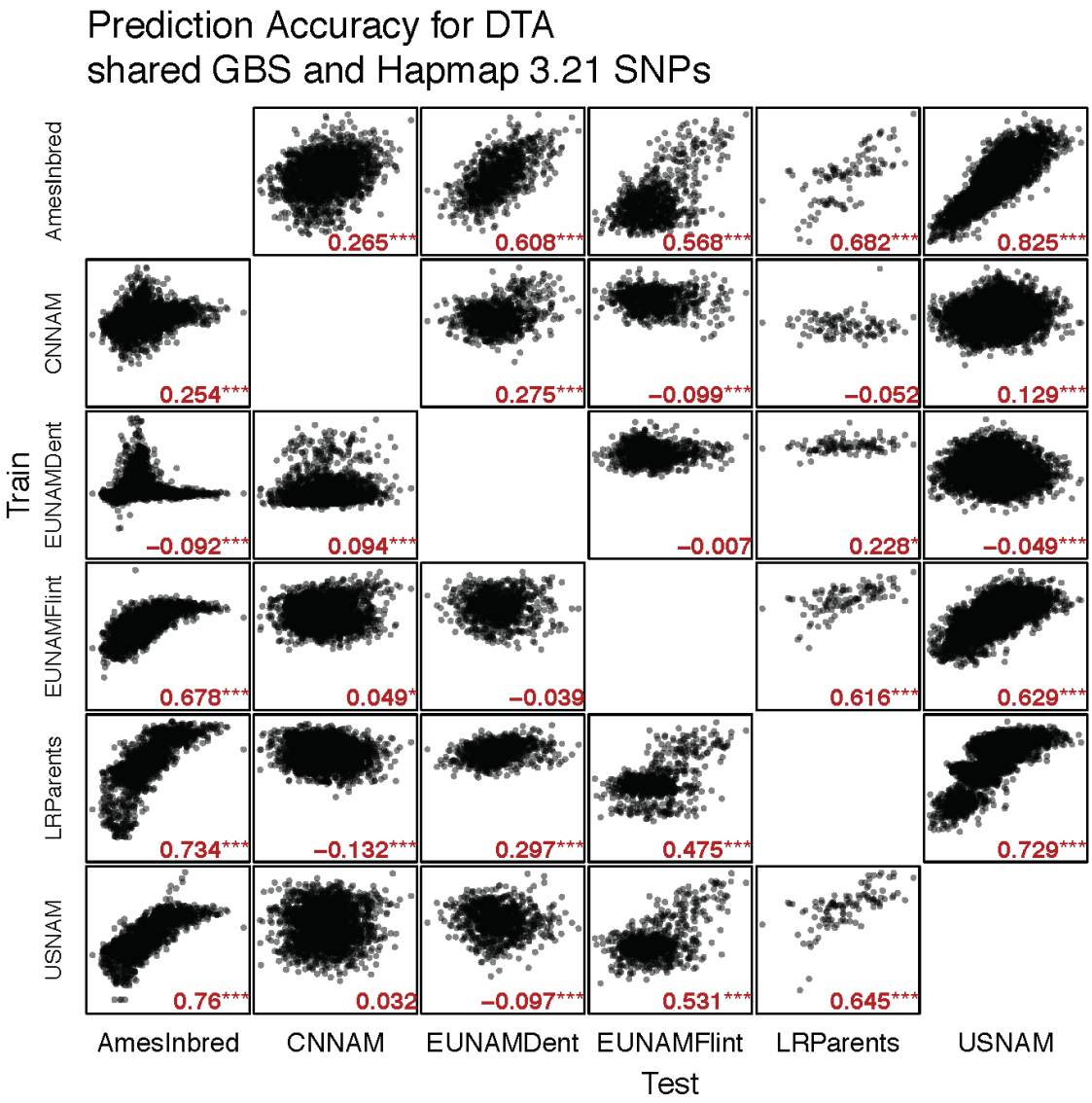


Figure S3. Cross population prediction accuracy for DTA trained and tested on inbred and landrace populations using the SNPs that overlap between Hapmap 3.21 and GBS. Predictions generated from RR-BLUP in TASSEL(*45*) by masking test phenotypes, and Pearson correlation between observed and masked test phenotypes generated using cor.test() in R.

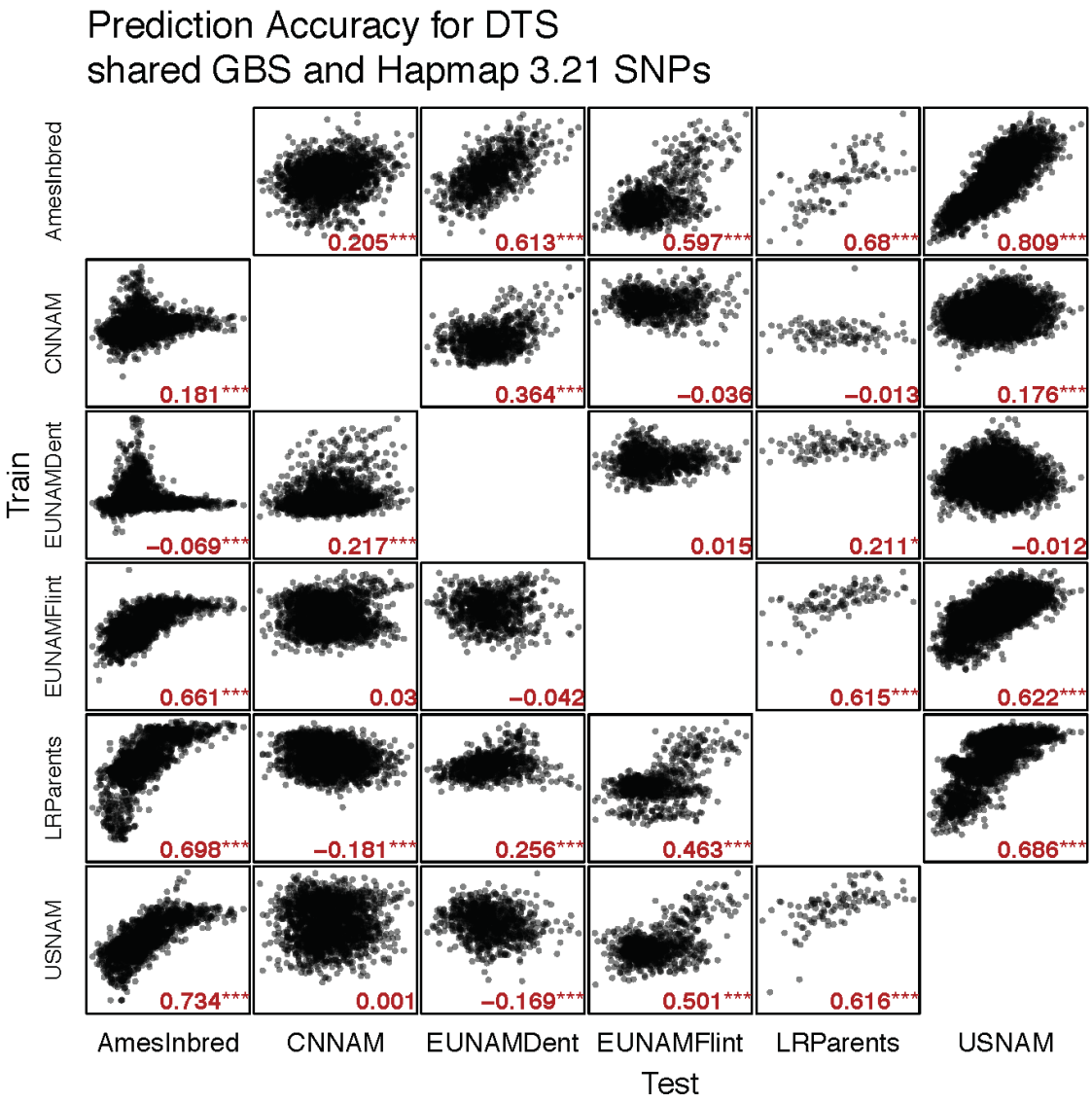


Figure S4. Cross population prediction accuracy for DTS trained and tested on inbred and landrace populations using the SNPs that overlap between Hapmap 3.21 and GBS. Predictions generated from RR-BLUP in TASSEL(*45*) by masking test phenotypes, and Pearson correlation between observed and masked test phenotypes generated using cor.test() in R.

Figure S5. Average days to flowering for landrace hybrid parent lines as predicted by Ames and translated to Blanding, Utah though growing-degree days. Assumes first planting on April 15th.

	Predicted DTA Ames	Predicted DTS Ames	Predicted GDD_DTA	Predicted GDD_DTS	DTA Blanding	DTS Blanding
Lower Colorado River	65	70	1421	1558	91	98
Athabaskan	65.9	70.7	1447	1583	93	99
Puebloan	69.3	74.7	1542	1704	97	105
Turkey Pen	68.2	73.6	1496	1670	95	103
Mean Landraces	73.1	78.8	1649	1815	101	109
Desert SW	73.4	78.9	1663	1874	103	111
North Mexican	75.2	80.9	1700	1882	105	112
Texas	76.3	82.4	1754	1940	106	114
C. Mexican lowland	80.9	87.8	1882	2067	112	120

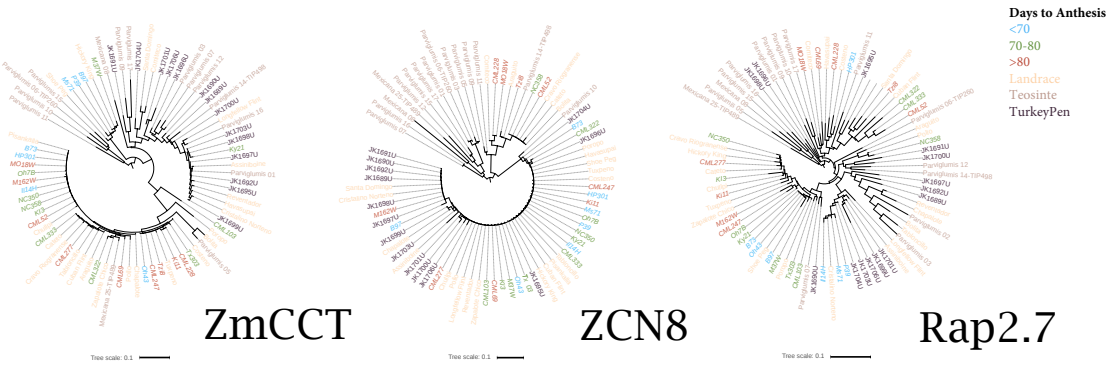


Figure S6. Neighbor-joining trees (*45*, *48*) for cloned flowering loci; ZCN8 (*64*), ZmCCT (*65*), and Rap2.7 of the Vgt1 locus (*66*). Teosinte does not form a monophyletic clade in any of these loci, suggesting that selection at these loci primarily acts on standing variation.

Table S1. Geographic/Cultural subsets

Turkey_Pen	Southeast Utah, Colorado Plateau (RGM/MB)
Gaspe_Flint	Northeastern landrace form Quebec, Canada (JW)
Lower_CO_River	Mostly sweet corn from the western colorado river (NSS)
Athabaskan	Navajo and Apache germplasm from all environments (NSS)
E_SW_US_High	Tiwa, Tewa, Towa or Keresan cultural affiliation above 2,000m (NSS)
E_SW_US_Mid	Tiwa, Tewa, Towa or Keresan cultural affiliation below 2,000m (NSS)
W_SW_US_Mid	Hopi and Zuni cultural affiliation (NSS)
Desert_SW_Low	O'odham culture, South Arizona deserts (NSS)
N_Mex_Unk	Elevation unknown, Northern Mexican origin (NSS/JRI)

N_Mex_High	Northern Mexico, above 2,000m (Sierra Madres) (NSS/JRI)
N_Mex_Mid	Northern Mexico, 1,000-2,000m (Elevation comparable to southern Arizona) (NSS/JRI)
N_Mex_Low	Northern Mexico, below 1,000m (coastal) (NSS/JRI)
Texas_Low	Sub-tropical Texas (NSS)
C_Mex_High	Below 1500m, below 23N (used in Takuno et al) (JRI)
C_Mex_Low	Above 1500m, below 23N (used in Takuno et al) (JRI/NSS)
SA_High	Andean material (used in Takuno et al) (JRI)
SA_Low	Lowland South American (used in Takuno et al) (JRI)
Mexicana	Upland teosinte. Multiple collection areas (JGS)
Parviglumis	Progenitor teosinte. Multiple collections (JGS)

Table S2. Presence of MITE insertion in Hapmap 3.21 samples tested and Turkey Pen. We BLASTed(*67*) the tightly conserved sequence surrounding the MITE insertion at the Vgt1 locus(*66*, *68*), which confers early flowering and is not present in the reference sequence B73. We found that it is segregating in Turkey Pen (71% of tested individuals) and teosinte (18%), and common in inbred landraces from Northern Mexico northward (47% overall, and 71% Northern Mexico northward). It is also present in South America, but Cristalino Norteno from South America is considered to derive from a North American Flint(*69*)) .

Insertion	NoInsertion	NoReads
BKN014(Assiniboine)	BKN009(Chullpi)	BKN019(Cateto)
BKN016(Longfellow Flint)	BKN010(Poropo)	BKN029(Bolita)
BKN022(Reventador)	BKN011(Pollo)	BKN032(Cravo Riogranense)
BKN026(Pisankalla)	BKN015(Havasupai)	BKN040(Hickory King)
BKN027(Cristalino Norteno)	BKN018(Shoe Peg)	JK1690
BKN030(Zapalote Chico)	BKN020(Cuban Flint)	TDD39103
BKN033(Chapalote)	BKN023(Araguito)	TIL02
BKN035(Tabloncillo)	BKN025(Tuxpeno)	TIL04-TIP454
Il14H	BKN031(Comiteco)	TIL15
JK1692	BKN034(Costeno)	
JK1695	Hp301	
JK1704	JK1703	
JK1706	JK1703	
Mo17	JK1705	
Ms71	TIL01	
P39	TIL03	
RIMMA0438.1(Maranon)	TIL05	
TIL06-260(Parviglumis)	TIL07	
TIL06-496(Parviglumis)	TIL08	
BKN017(Santa Domingo)	TIL09	
TIL10(Parviglumis)	TIL11	

	TIL12 TIL14 TIL16 TIL17 TIL25	

REFERENCES

1. J. Doebley, J. D. Wendel, J. S. C. Smith, C. W. Stuber, M. Goodman, The Origin of Cornbelt Maize: The Isozyme Evidence. *Econ. Bot.* **42**, 120–131 (1988).
2. E. Anderson, W. L. Brown, in *Heterosis*, J. W. Gowen, Ed. (Iowa State College Press, Ames, Iowa, 1952), pp. 124–148.
3. C. Rebourg *et al.*, Maize introduction into Europe: the history reviewed in the light of molecular data. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **106**, 895–903 (2003).
4. S. Bouchet *et al.*, Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with a Major Contribution of the Vgt2 (ZCN8) Locus. *PLOS ONE.* **8**, e71377 (2013).
5. E. Anderson, H. C. Cutler, Races of Zea Mays: I. Their Recognition and Classification. *Ann. Mo. Bot. Gard.* **29**, 69–88 (1942).
6. W. C. Galinat, J. H. Gunnerson, SPREAD OF EIGHT-ROWED MAIZE FROM THE PREHISTORIC SOUTHWEST. *Bot. Mus. Leafl. Harv. Univ.* **20**, 117–160 (1963).
7. J. F. Doebley, M. Goodman, C. W. Stuber, Exceptional Genetic Divergence of Northern Flint Corn. *Am. J. Bot.* **73**, 64–69 (1986).
8. J. F. Doebley, M. M. (North C. S. U. R. (USA) D. of S. and G. Goodman, C. W. (Agricultural R. S. Stuber, Isozyme variation in maize from the Southwestern United States: taxonomic and anthropological implications. *Maydica Italy* (1983) (available at http://agris.fao.org/agris-search/search.do?recordID=XE8363476).
9. R. . Ford, The Color of Survival. *Discovery*, 17–30 (1980).
10. J. B. (ed. . Mabry, *Archaeological Investigations of Early Village Sites in the Middle Santa Cruz Valley* (Center for Desert Archaeology, Tucson, 1998), *Anthropological Papers*.
11. J. B. (ed. . Mabry, *Las Capas: Early Irrigation and Sedentism in a Southwestern Floodplain* (Center for Desert Archaeology, Tucson, 2008), *Anthropological Papers*.
12. D. A. (ed. . Gregory, *Excavations in the Santa Cruz River Floodplain: The Early Agricultural Period Component at Los Pozos*. (Center for Desert Archaeology, Tucson, 2001), *Anthropological Papers*.
13. D. A. Gregory, N. M. Stevens, F. L. Nials, M. R. Schurr, M. W. (editors) Diehl, *Excavations in the Santa Cruz River Floodplain: Further Investigations at Los Pozos*. (Center for Desert Archaeology, Tucson, 2007), *Anthropological Papers*.
14. E. K. Huber, C. V. . (editors) West, *Fence Lake Project. Archaeological Data Recovery in the New Mexico Transportation Corridor and First Five-Year Permit Area, Fence Lake Coal Mine Project, Catron County, New Mexico* (Statistical Research Inc., Tempe, 2005), *Technical Series*.
15. F. E. Smiley, The Agricultural Transition in the Northern Southwest: Patterns in the Current Chronometric Data. *Kiva.* **60**, 165–189 (1994).

16. R. R. da Fonseca *et al*., The origin and evolution of maize in the Southwestern United States. *Nat. Plants*. **1**, 14003 (2015).
17. H. W. Dick, *Bat Cave,* (School of American Research, Santa Fe, N.M., 1965).
18. W. H. Wills, *Early prehistoric agriculture in the American Southwest* (School of American Research Press ; Distributed by University of Washington Press, Santa Fe, N.M; [Seattle], 1988).
19. B. J. Vierra, R. . Ford, in *Histories of maize: multidisciplinary approaches to the prehistory, linguistics, biogeography, domestication, and evolution of maize, edited by J. E. Staller, R. H. Tykot and B. F. Benz* (Elsevier Academic Press, Amsterdam; Boston, 2006), pp. 497–510.
20. L. W. Huckell, in *Histories of maize: multidisciplinary approaches to the prehistory, linguistics, biogeography, domestication, and evolution of maize, edited by J. E. Staller, R. H. Tykot and B. F. Benz* (Elsevier Academic Press, Amsterdam; Boston, 2006), pp. 97–107.
21. W. L. Merrill *et al*., The diffusion of maize to the southwestern United States and its impact. *Proc. Natl. Acad. Sci.* **106**, 21019–21026 (2009).
22. L. S. Cordell, M. E. McBrinn, *Archaeology of the Southwest* (Academic Press, San Diego, Calif., Third Edition., 2012).
23. S. Plog, *Ancient peoples of the American Southwest* (Thames and Hudson, New York, N.Y., 1997).
24. B. J. (ed. . Vierra, *The Late Archaic across the Borderlands: from foraging to farming* (University of Texas Press, Austin, 2005).
25. Utah State Extension, Average frost free days in Utah, (available at http://extension.usu.edu/htm/faq/faq_q=153).
26. Frost Dates for Arizona, (available at https://www.victoryseeds.com/frost/az.html).
27. B. A. Bellorado, K. C. Anderson, Early Pueblo Responses to Climate Variability: Farming Traditions, Land Tenure, and Social Power in the Eastern Mesa Verde Region. *KIVA*. **78**, 377–416 (2013).
28. S. E. Ingram, R. C. Hunt, *Traditional Arid Lands Agriculture: Understanding the Past for the Future* (University of Arizona Press, 2015).
29. R. G. Matson, *The Origins of Southwestern Agriculture* (University of Arizona Press, 1991).
30. R. G. Matson, B. Chisholm, Basketmaker II Subsistence: Carbon Isotopes and Other Dietary Indicators from Cedar Mesa, Utah. *Am. Antiq.* **56**, 444–459 (1991).
31. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* (2009), doi:10.1101/gr.094052.109.
32. M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, N. C. Ellstrand, J. Ross-Ibarra, in *Poster Session* (Portland, OR, 2012).
33. M. B. Hufford *et al*., The Genomic Signature of Crop-Wild Introgression in Maize. *PLoS Genet*. **9**, e1003477 (2013).
34. T. Pyhäjärvi, M. B. Hufford, S. Mezmouk, J. Ross-Ibarra, Complex patterns of local adaptation in teosinte. *arXiv:1208.0634* (2012) (available at http://arxiv.org/abs/1208.0634).

35. J. Romero-Navarro *et al*., Tapping a world of genetic variation: surveying flowering time diversity in maize landraces. **Nature Genetics (in review)**.
36. E. Rodgers-Melnick *et al*., Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc. Natl. Acad. Sci.* **112**, 3823–3828 (2015).
37. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
38. T. M. Beissinger *et al*., Recent demography drives changes in linked selection across the maize genome. *Nat. Plants*. **2**, 16084 (2016).
39. M. B. Hufford *et al*., Comparative population genomics of maize domestication and improvement. *Nat. Genet*. **44**, 808–811 (2012).
40. E. S. Buckler *et al*., The Genetic Architecture of Maize Flowering Time. *Science*. **325**, 714–718 (2009).
41. A. F. Troyer, L. G. Hendrickson, Background and Importance of "Minnesota 13" Corn. *Crop Sci*. **47**, 905 (2007).
42. Y. Li *et al*., Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J*. **86**, 391–402 (2016).
43. M. C. Romay *et al*., Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. **14**, R55 (2013).
44. L. . Abendroth, R. W. Elmore, M. J. Boyer, S. R. Marlay, *Corn growth and development* (Iowa State University, University Extension, Ames, Iowa, 2011; http://newcatalog.library.cornell.edu/catalog/7300063), *PMR*.
45. P. J. Bradbury *et al*., TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**, 2633–2635 (2007).
46. V. Jaenicke-Després *et al*., Early Allelic Selection in Maize as Revealed by Ancient DNA. *Science*. **302**, 1206–1208 (2003).
47. W. F. Tracy, S. R. Whitt, E. S. Buckler, *Crop Sci*., in press, doi:10.2135/cropsci2006-03-0149tpg.
48. I. Letunic, P. Bork, *Nucleic Acids Res*., in press, doi:10.1093/nar/gkw290.
49. K. Palaisa, M. Morgante, S. Tingey, A. Rafalski, Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9885–9890 (2004).
50. E. J. Johnson, The role of carotenoids in human health. *Nutr. Clin. Care Off. Publ. Tufts Univ*. **5**, 56–65 (2002).
51. K. Yoshida *et al*., The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *eLife*. **2**, e00731 (2013).
52. N. Rohland, M. Hofreiter, Comparison and optimization of ancient DNA extraction. *BioTechniques*. **42**, 343–352 (2007).
53. A. W. Briggs *et al*., Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res*. **38**, e87 (2010).
54. M. Meyer, M. Kircher, *Cold Spring Harb. Protoc*., in press, doi:10.1101/pdb.prot5448.

55. M. Meyer *et al*., A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*. **338**, 222–226 (2012).
56. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. **40**, e3 (2012).
57. R. Bukowski *et al*., Construction of the third generation Zea mays haplotype map. *bioRxiv*, 26963 (2015).
58. R. J. Elshire *et al*., A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*. **6**, e19379 (2011).
59. S. Takuno *et al*., Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics*. **200**, 1297–1312 (2015).
60. M. McCaw, P. Albert, J. Birchler, (Pheasant Run, Illinois, 2013), vol. 55:P227.
61. P. M. VanRaden, Efficient methods to compute genomic predictions. *J. Dairy Sci*. **91**, 4414–4423 (2008).
62. J. B. Endelman, J.-L. Jannink, Shrinkage estimation of the realized relationship matrix. *G3 Bethesda Md*. **2**, 1405–1413 (2012).
63. National Agricultural Statistics Service, "Usual Planting and Harvesting Dates for U.S. Field Crops, December 1997," *Agricultural Handbook* (628, United States Department of Agriculture), (available at https://www.nass.usda.gov/Publications/Usual_Planting_and_Harvesting_Dates/uph97.pdf).
64. X. Meng, M. G. Muszynski, O. N. Danilevskaya, The FT-like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. *Plant Cell*. **23**, 942–960 (2011).
65. H.-Y. Hung *et al*., ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc. Natl. Acad. Sci*. **109**, 11068–11069 (2012).
66. S. Salvi *et al*., Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol*. **48**, 601–613 (2002).
67. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol*. **215**, 403–410 (1990).
68. S. Ducrocq *et al*., Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information. *Genetics*. **178**, 2433–2437 (2008).
69. G. F. Sprague, J. W. Dudley, M. M. Goodman, W. L. Brown, in *Agronomy Monograph* (American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1988; https://dl.sciencesocieties.org/publications/books/abstracts/agronomymonogra/cornandcornimpr/33).
70. R. A. Swanson-Wagner *et al*., Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res*. **20**, 1689–1699 (2010).

APPENDIX A

FAILED EXPERIMENTS

Long-range PCR for detailed modern comparative haplotypes in diverse modern landraces

A project to which I devoted considerable time, but which did not produce usable results, was in generating de novo haplotypes for carotenoid, tocochromanol, flavonoid, and major, known flowering time and starch loci. I started this project because these traits are important in the Southwest for the regions stated above, and because Buckler Laboratory whole genome sequenced inbred lines contained no blue-kerneled lines and few red kernels. These haplotypes could be compared with Turkey Pen to predict kernel color and other traits in a diverse panel, enriched for Southwestern samples.

Significant structural variation in maize (*1*) makes it challenging to find conserved regions for which to design primers. To increase the chance of placing primers in conserved regions, I designed a java plugin to integrate the Primer3 program with custom blasts to four de novo genomes available at the time. The plugin pulled the gene region in question, with additional flanking sequence, and queried Primer3 for primer targets. Following this, these targets were checked for conservation with the de novo genomes using the BLAST algorithm, and finally a list of the best targets was output for primer ordering.

I used the Q5 (NEB) system, which can amplify regions up to 10kb, and I was by chance successful for the first few regions that I attempted to amplify 56 gene regions in a test set of 8 diverse temperate and tropical inbreds, and two Southwestern landraces. Table 1 shows the reaction conditions, Table 2 the thermocycler conditions, and Table 3 the gene regions tested and the outcomes. Only about a third of the gene

regions successfully amplified in all of the test set, about a third never amplified, and the rest would amplify in some test individuals but not others.

Table 1. Reaction conditions

Q5 (no GC)

	1 rxn	8.2 rxn	33 rxn (4x8)	50 rxn (6x8)	66 rxn (8x8)	100 rxn (12x8)
Q5 buffer	**2**	**16.4**	**66**	**100**	**132**	**200**
10mM dNTP	**0.2**	**1.64**	**6.6**	**10**	**13.2**	**20**
10mM primersFR	**1**	**8.2**	variable	variable	variable	variable
10ng/ul DNA	**1**	variable	variable	variable	variable	variable
Q5 polymerase	**0.1**	**0.82**	**3.3**	**5**	**6.6**	**10**
H20 to 10ul	**5.7**	**46.74**	**188.1**	**285**	**376.2**	**570**

65.6ul master mix to each primer specific master tube
1ul DNA in each rxn

Table 2. Thermocycler conditions

Table 3. Gene regions and success rate for amplicon sequencing

Gene	Name	Chr	v3Start	v3End	Pathway	Success
GRMZM2G179147	ABA8HO	5	200705194	200707991	Carotenoid	yes
GRMZM2G493395	dxs2	7	14086686	14089909	Carotenoid	yes
GRMZM5G837869	lut5	5	215877677	215882183	Carotenoid	yes
GRMZM2G410515	pds	1	17660122	17666235	Carotenoid	yes
GRMZM2G300348	y1	6	82180486	82184217	Carotenoid	yes
GRMZM2G164318	bohase2	2	15877183	15879464	Carotenoid	no
GRMZM2G150363	CCD4	5	200743141	200745544	Carotenoid	no
GRMZM2G152135	crtRB1	10	136081567	136084686	Carotenoid	no
GRMZM2G173641	dxs3	9	20471360	20476373	Carotenoid	no
GRMZM2G137409	HDS	5	182174564	182181190	Carotenoid	no
GRMZM2G012966	lcyE	8	138416903	138424121	Carotenoid	no
GRMZM2G143202	Lut1	1	86848122	86858514	Carotenoid	no
GRMZM5G849107	lycB	5	100737438	100739288	Carotenoid	no
GRMZM2G014392	vp14	1	250953388	250956063	Carotenoid	no
GRMZM2G057243	wc1	9	152353213	152359196	Carotenoid	no
GRMZM2G454952	zds1	7	17481578	17490013	Carotenoid	no
GRMZM2G127139	zep1	2	45068944	45077882	Carotenoid	no
GRMZM2G026930	a1	3	216386230	216387972	flavenoid	yes
GRMZM2G345717	a2	5	66136522	66139360	flavenoid	yes
GRMZM2G172795	b1	2	19041697	19046154	flavenoid	yes

RMZM2G016241	bz2	1	241430827	241431829	flavenoid	yes
GRMZM2G422750	c2	4	192758391	192761788	flavenoid	yes
GRMZM2G058292	pac1	5	196661199	196666224	flavenoid	yes
GRMZM2G701063	pl1	6	108491299	108492380	flavenoid	yes
GRMZM2G031859	BCHO1	1	6347072	6348308	flavenoid	no
GRMZM2G165390	bz1	9	11779648	11781406	flavenoid	no
GRMZM2G005066	c1	9	9746518	9747591	flavenoid	no
GRMZM2G155329	chi1	1	293118119	293119647	flavenoid	no
GRMZM2G042733	in1	7	19360149	19366043	flavenoid	no
GRMZM2G084799	p1	1	48118788	48129338	flavenoid	no
GRMZM2G025832	pr1	5	180084037	180086107	flavenoid	no
GRMZM5G822829	r1	10	138489998	138498818	flavenoid	no
GRMZM2G144744	d8	1	266160101	266163168	FT	yes
GRMZM2G067921	DLF1	7	175583965	175585451	FT	yes
GRMZM2G179264	ZCN8	5	123538197	123539985	FT	yes
GRMZM2G129889	ZmPHYC2	5	7129937	7134953	FT	yes
GRMZM2G160730	GL15	9	96744684	96748027	FT	no
GRMZM2G011357	ID1	1	239667869	239671192	FT	no
	MITE	8	131517973	131520473	FT	no
GRMZM2G381691	ZmCCT	10	94261581	94264581	FT	no
GRMZM2G171365	ZMM5	9	154082657	154102449	FT	no
GRMZM2G057935	ZmPHYC1	1	277059620	277064623	FT	no
GRMZM2G700665	ZmRap2.7	8	131576889	131580316	FT	no
GRMZM2G089713	sh1	9	11500945	11506749	starch	yes
GRMZM2G032628	ae1	5	168492139	168509225	starch	no
GRMZM2G068506	bt2	4	58979526	58985265	starch	no
GRMZM2G429899	sh2	3	216495981	216505345	starch	no
GRMZM2G138060	su1	4	41396390	41405179	starch	no
GRMZM2G024993	waxy1	9	23267684	23271612	starch	no
GRMZM2G173358	HGGT	9	93489726	93493445	Toco	yes
GRMZM2G112728	sps2	1	11107901	11112061	Toco	yes
GRMZM2G082998	VTE3	1	174010201	174012244	Toco	no
GRMZM2G377115	ABC1K1	4	12946774	12961838	Toco	no
GRMZM2G088396	HPPD	5	83895923	83898077	Toco	no
GRMZM2G084942	PrepdHO	5	59318968	59326934	Toco	no
GRMZM2G035213	vte4	5	200419844	200423666	Toco	no

REFERENCES

1. R. A. Swanson-Wagner *et al.*, Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* **20**, 1689–1699 (2010).