

Review: Adaptive Introgression Expanded the Genetic Base of Crops during post-Domestication Spread

Authors: Garrett M. Janzen¹, Li Wang¹, and Matthew B. Hufford^{1,*}

¹Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

*Correspondence: mhufford@iastate.edu (M.B. Hufford)

The process of domestication is often conceptualized as geographically constrained, with crops originating from a wild progenitor within one or more defined centers followed by expansion to the modern-day extent of cultivation [1]. However, archaeological and genetic evidence are beginning to reveal that, in many cases, domestication has been temporally protracted and geographically diffuse [2, 3, 4, 5, 6]. An additional important aspect of the emerging complexity of domestication is beneficial gene flow (*i.e.*, adaptive introgression) from locally adapted wild relatives during crop expansion following initial domestication.

Adaptive introgression has three components: hybridization between two genomes, backcrossing to one of the parents, and selection on different recombinant genotypes with progressively diminished linkage drag [7, 8]. In domesticated species, adaptive introgression would consist of crop/wild hybrids backcrossing to a crop, retention and increase in frequency of adaptive wild haplotypes in the crop, and selection against undesirable wild background. To date, literature on crop-wild gene flow has focused on the risk of transgene introgression from domesticated crops into wild relatives (for a review, [9]) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, [10]). The history of natural and potentially adaptive introgression of wild alleles into domesticated crops over evolutionary timescales has received considerably less attention. However, new tools and methods have recently been employed to detect genome-wide patterns of introgression, granting new insights into the prevalence of adaptive introgression in crop histories. Preliminary results suggest a need to expand our conception of domestication to include the broadening of the genetic base of crops that occurred during post-domestication expansion through gene flow with newly encountered wild relatives. *I think these paragraphs read pretty smooth. I pondered at the last sentence of this paragraph. Suggestions: Preliminary results suggest a need to expand our conception of genetic bases of domestication in crops that inhabited new territory through adaptive gene flow from newly encountered wild relatives during post-domestication expansion.*

In this review, we will: 1) briefly describe recently developed methods for detecting adaptive introgression and provide a summary of how they can be applied to detect crop-wild introgression, 2) present case studies suggesting wild-to-crop introgression has conferred local adaptation, 3) consider how introgression bears upon fundamental questions of domestication, and 4) describe future advances in both basic and applied genetics that can be made through the study of introgression in agroecosystems.

Introgression methods and their application

In this section, I think the overall content is good, but we need to edit to make it more accessible and more explicit about how methods are implemented to detect adaptive introgression

The decreasing cost of genome-wide resequencing and availability of reduced-representation genotyping (*e.g.*, GBS and RAD-Seq), combined with new analytical methods, has facilitated comprehensive study of introgression across a number of species (**Table 1**). *Are there newer resequencing methods to discuss here? DART-Seq, etc? I think it is fine to mention two of these methods, it is not our focus.* High-density marker data can be used with haplotype-based and other methods to assign specific genomic regions to a taxon of origin and identify introgression across taxa [11, 12, 13, 14, 15, 16]. The methods reviewed here do not include those marginally estimating introgression/migration rate as a component of demographic history (*e.g.*, Approximate Bayesian Computation (ABC) [17], diffusion approximations for demographic inference (*δaδi*) [18], isolation with migration models [19], and a series of methods utilizing the sequentially Markovian coalescent (PSMC, MSMC and SMC++) [20, 21, 22]). Rather, we focus on methods that explicitly identify introgressed genomic segments based on the extent of differentiation, on patterns of nucleotide/haplotype sharing, and phylogenetic relationships.

First, introgressed segments are expected to show low differentiation from their source population. The F_{st} and d_{XY} statistics and their derivatives including G_{min} [16] and RND_{min} [15] gauge differentiation. The former two statistics are insensitive to rare migrants and therefore lack power to detect recent introgression, while the latter two overcome this limitation. Additionally, RND_{min} accounts for variable mutation rate, which is detected based on branch length to an outgroup:

$$RND_{min} = \frac{d_{min}}{d_{out}} \quad (1)$$

where d_{min} is the minimum sequence distance between haplotypes in species X and Y and d_{out} equals $(d_{XO} + d_{YO})/2$, the average sequence distance between each species and the outgroup (O).

These statistics have recently been further developed by adding differentiation between both non-admixed (A) and admixed populations (B) and a source population (C) [23]. For example, the $U_{A,B,C(w,x,y)}$ statistic summarizes number of sites where an allele at frequency y in the source population (C) has a frequency higher than x in the admixed population (B) and lower than w in the non-admixed population (A). A similar statistic, $Q95_{A,B,C(w,y)}$, sets a hard cutoff at the 95th percentile of allele frequencies in the admixed population (B) [23]. Further modifications have allowed specification of more than one source population (see details in [23]).

Second, local ancestry deconvolution (also known as chromosome painting) assigns genomic regions to various source populations based on patterns of allele/haplotype sharing [24]. One form of chromosome painting utilizes hidden Markov models to evaluate ancestry across admixed genomes through comparison to reference, non-admixed individuals (*e.g.*, HAPMIX [12]). Another clusters admixed populations with reference samples using a sliding-window approach (*e.g.*, PCAdmix [25] and LAMP [26]). And finally, introgression can be detected through chromosome painting by using a Bayesian model [27] in which deviations from Hardy-Weinberg equilibrium are minimized through creation of genetic groups (*e.g.*, fineSTRUCTURE [13]).

Li, are you familiar with the analytical tools MIGRATE-N and BAYESASS? Rieseberg compares these two to STRUCTURE at some length in the discussion of this paper: <https://biology.unm.edu/Whitney/Whitney> Should we include these two methods, even in passing, in this portion of the paper? Garrett, I donot know these two methods. I am reading the paper you mentioned here.

Third, the ABBA-BABA statistic (also known as the D-statistic) and its derivatives are widely applied to introgression detection. These statistics make inferences regarding introgression based

on genomic patterns of derived variants that are shared between populations or species. Patterns of allele sharing are interpreted in a phylogenetic context and the method is best suited to detection of introgression at the genome level. Elaborations of the D-statistic capable of localizing introgression to specific genomic regions include \hat{f}_d [11] and the five-taxon D-statistic [14]. The former is quite similar to the D-statistic but uses allele frequencies from each population/species, and the latter detects introgression based on the localized phylogenetic pattern and is capable of determining introgression directionality.

This paragraph about application seems like it might not belong in this section. Maybe it should be a part of the introduction? Or, maybe each of these examples could be mentioned directly following the method(s) that was used in the study?

Application of these approaches across a number of plant and animal species suggests introgression can play an adaptive role. For example, introgression from ancient hominins (*e.g.*, Neanderthals and Denisovans) to humans has been detected at loci controlling skin pigmentation, defense against pathogens, and toleration of high altitude (reviewed in [28]); introgression of wing coloration loci has conferred Müllerian mimicry (a type of mimicry in which two unpalatable species resemble one another, compounding the efficacy of their repellent signal) across butterfly species [29]; introgression has spread insecticide resistance across mosquito species [30], and introgression across *Mimulus* (*i.e.*, monkeyflower) species has resulted in adaptation to pollinator preference and contributed to speciation [31].

Crop adaptation through introgression

Genome-wide data from extensive samples of crops and their wild relatives, in combination with the new methods described above, have recently allowed detailed analysis of wild-to-crop introgression in some of the world’s most important crops (**Table 2**). Below we present a summary of findings from maize, barley, and rice, three promising systems in which introgression from wild relatives appear to have played an adaptive role.

1. Maize:

The relationship between maize (*Zea mays* ssp. *mays*) and the teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) offers a prime case study of adaptive wild-to-crop introgression. Maize was domesticated from *Zea mays* ssp. *parviglumis* (hereafter, *parviglumis*) approximately 9,000 BP in the lowlands of the Balsas River Valley in Mexico [32]. From this domestication center, maize spread into the highlands of the Mexican Central Plateau, where it came into sympatry with *mexicana*. Introgression from *mexicana* to maize in the Central Plateau has been reported based on both morphological [33] and molecular [34, 35] data. However, Hufford et al. [36] first localized *mexicana* introgression to chromosomal regions and provided evidence that it was likely adaptive. The authors identified nine genomic regions in several maize populations which consistently showed evidence of *mexicana* introgression based on chromosome painting methods including HAPMIX (Figure 1). These introgressed segments overlapped with QTLs that had previously been found to control anthocyanin content and leaf macrohairs [37], traits known to be adaptive at high elevation. In a growth chamber experiment, the authors demonstrated that maize populations with *mexicana* introgression showed greater plant height (a proxy for fitness) under highland environmental conditions than populations that lacked introgression. Height differences were not detected under lowland conditions.

Populations of *mexicana* cannot be found outside of the highlands of Mexico, yet maize has colonized and adapted to high elevation in a number of other regions. A recent study

[4] employed the ABBA-BABA and \hat{f}_d statistics to evaluate whether maize with *mexicana* introgression was transferred to other highland regions or whether highland adaptation was obtained *de novo* outside of Mexico. Overall, analyses revealed that *mexicana* introgressions were transferred to Meso-American high elevation regions in Guatemala and the southwestern United States, but more distant high elevation regions (*e.g.*, the Andes) showed no *mexicana* ancestry.

Since *mexicana* haplotypes appear to have played an important role historically in adaptation of maize to challenging high-elevation conditions, modern breeding programs may further benefit from drawing on *mexicana* germplasm, particularly in regions like the Andes where *mexicana* alleles are not known to have spread.

2. Barley:

Barley (*Hordeum vulgare* subsp. *vulgare*) was domesticated at least twice roughly 8,000 to 10,000 BP: once from the wild subsp. *spontaneum* in the Fertile Crescent and once from subsp. *spontaneum* var. *agriocrithon* in Tibet [38, 39, 40, 41, 42?]. Presently, the distribution of subsp. *spontaneum* stretches from the eastern Mediterranean through the Middle-East to west-central Asia, spanning clines in temperature, precipitation, soil type, and altitude [43]. Barley-*spontaneum* hybrids are fertile and found spontaneously when these taxa co-occur. Introgression between wild and domesticated barley is frequent [42], at times occurring over distances greater than a kilometer [44].

Poets and co-authors [45] recently investigated the range-wide contribution of wild barley to landraces, assessing both genome-wide and geographical patterns. This study identified several lines of evidence consistent with wild introgression aiding the dispersal and adaptation of domesticated barley. Genomic regions of shared ancestry were detected linking particular landraces to several distinct wild relative populations, suggesting landraces may have received wild introgression on a continual basis during post-domestication expansion. However, barley landraces showed an excess of ancestry from nearby wild relatives, indicating a prevalence of local and potentially adaptive gene flow. Low linkage disequilibrium and small blocks of identity by state indicated even these locally introgressed regions are old, perhaps dating back to the early expansion of barley following domestication. While these results are suggestive, wild barley haplotypes have yet to be definitely linked to specific local adaptations in landraces.

3. Asian Rice:

The story of Asian rice (*Oryza sativa*) domestication is still debated, with hybridization between wild and domesticated rice contributing to the complexity of this crop’s history. Certain genetic and archaeobotanical evidence point toward independent domestications of the two prominent varietal groups *japonica* and *indica* from the wild species *Oryza rufipogon* (*rufipogon* hereafter) in the Yangtze Basin of China and the Indian Ganges plain, respectively [46]. Other studies support a single domestication occurring 8,200-13,500 BP in the Yangtze Basin, with later divergence of *japonica* and *indica* [47]. Huang and colleagues [48] developed a genetic map of rice variation, which they used to measure genetic distance between wild and domesticated rices at and around domestication loci for various geographical locations, finding that *japonica* was likely domesticated in the Pearl River area of Guangxi province, China (just south of the Yangtze Basin), and that *indica* was likely the result of hybridization between *japonica* and local *rufipogon* populations in Southern and South-eastern Asia.

Vaughan and colleagues [49] posit a hypothesis combining elements of both theories. According to this hypothesis, the non-shattering allele arose in a single cultivated rice population and

then introgressed into many diverse cultivated populations (some japonica-like, some indica-like) in the region. This would have facilitated the subsequent spread of other domestication alleles among those populations. As these domesticated populations then spread into new environments, they would have received introgressions from locally-adapted wild relatives, retaining alleles that improved fitness (in agreement with [50]). Contemporary populations of wild and domesticated rice show evidence of this history of local adaptation. The wild relatives *rufipogon* and *nivara* both maintain high genotypic and phenotypic diversity and exhibit locally-adaptive traits (*rufipogon* is adapted to forested wetland environments, whereas *nivara* is adapted to dryer conditions and has life cycle adaptations to survive grazing pressure), suggesting that either offered (and offers) adaptive genetic potential. Likewise, domesticated rice varieties display patterns of local adaptation (for examples, two of the domesticated rice deepwater varieties (*rayada* and *ashwina*) are said to be selected for the environment along the Ganges river, the *japonicas* are split into temperate and tropical subgroups, and the *indicas* are best suited for lowland environments).

To date, research into adaptive introgression in the domestication of rice has been insufficient to detect clearly-supported examples (though [51] emphasizes the necessity of this research). There are perhaps some practical reasons why research has not yet been devoted to this inquiry. As with many other domesticated crops, gene flow between wild and domesticated rices is highly asymmetric (estimates of wild rice admixture in domesticated rice are less than 5 percent [52]). This asymmetry is due in part to the closed floret architecture of the domesticated rice, which hinders outcrossing. Crop-wild hybridization is frequent enough that nuisance weedy hybrids commonly arise amongst cultivated populations and must be managed with difficulty [49]. During early domestication, introgression may have been more prevalent than at present because barriers to crop-to-wild introgression may have been less severe and because the inbreeding reproductive system of rice would not have been as firmly established [49]. Furthermore, the contemporary distribution of wild rice does not capture the range and diversity of wild rice during early domestication and range expansion of rice.

4. Potato

I had saved these notes from an earlier version of the paper. There is plenty of evidence of hybridization and introgression (both farmer-mediated and and natural), and some adaptive traits have been conferred, but I could not find direct evidence of specific cases of adaptive natural introgression from wild relatives to cultivated potatoes. This new paper fills that gap, if we decide that the paper's merits are substantial enough that we can overlook its flaws. If we decide to keep and use the new potato introgression publication, that would be the natural choice for the potato reference, whether or not we include the write-up of the potato case study.

Modern potato (*Solanum tuberosum*) is believed to have been domesticated in southern Peru in sympatry with a multitude of wild relatives about 6000 BP, although the exact location and formal classification and phylogenetic relationships between these taxa have long been disputed [? ? ? ?]. The northern members of the polyphyletic *S. brevicaulle* wild potato complex have been identified as likely progenitors of modern potatoes [?], but determination of a single progenitor species is unlikely, either because widespread gene flow in the complex will mask signs of this phylogeny or because *s. tuberosum* has a polyphyletic origin. Also, these northern members of the *S. breviaule* complex are not clearly defined, and may in fact be one singular species (in which case the species name would be *S. bukasovii*, [?]).

Although potatoes are usually propagated clonally, farmers also promote sexual hybridization at times to improve disease resistance and develop new cultivars [?]. Farmers continue to grow potatoes in close proximity to wild relatives, resulting in domesticate-weedy-wild hybrid

Table 1: List and brief description of recently developed methods and examples of empirical studies employing these methods.

methods	data type	reference	empirical stu
chromosome painting			
Hapmix	phased haplotype; reference panel	[12]	[? ?]
RASPBerry	phased haplotype	[?]	[?]
MultiMix	phased/unphased genotype; reference panel	[?]	[?]
PCAdmix	phased haplotype	[25]	[? ?]
LAMP	phased haplotypes; reference panel	[26]	[?]
phylogenetic relationship			
ABBA-BABA/D-statistics	biallelic SNP	[?]	[?]
fd statistic	biallelic SNP	[?]	[? ?]
five taxon D statistics	biallelic SNP	[14]	[? ?]
divergence			
Gmin	biallelic SNP	[16]	[?]
RNDmin	phased haplotype	[15]	NA
$U_{A,B,C(w,x,y)}$ and $Q95_{A,B,C(w,y)}$	biallelic SNP	[28]	[?]

complexes which promote introgressive hybridization [? ? ?]. These complexes, combined with a diverse range of biotic and environmental selective pressures and local farming practices (human-mediated migration, isolated farmsteads in fertile valleys, clonal propagation, and intentional maintenance of a variety of landraces), have fostered expansion of genetic diversity within potatoes subsequent to domestication [?]. However, as farmers tend to abandon fields after being used for potato cultivation, it is less likely that hybrids have an opportunity to form stable populations for maintained introgressive gene flow into the domesticated potato gene pool [?]. Also, natural selection may be subverted by farmer preference and artificial selection [?].

Introgressive hybridization is widespread in potatoes [?]. The various cultivars of Andean potatoes are interfertile, forming one large plastic gene pool [?]. Andean potatoes exhibit high ecological versatility, due in part either to allelic diversity in polyploids or introgression of desirable alleles from wild relatives in diploids [?]. Cultivated diploid potatoes have been shown introgressed with genes from *S. sparsipilum* [?] and *S. megistacrolobum* [? ?]. Drought- and frost-resistance genes introgressed from wild relatives expedited the spread of cultivated potatoes into the central Andean Altiplano [56? ?]. Resistance genes have been experimentally introgressed from wild potato relatives (*S. tuberosum subsp. andigena*, [?], *S. bulbocastanum*, [?]) into cultivated potato, although this process is hindered by compatibility issues including ploidy levels and Endosperm Balance Number [?].

Re-evaluating domestication

A framework in which crops are domesticated from a single wild population or even a single species is an oversimplification when introgression during the geographic expansion of crops is extensive.

Crop	Compatible Wild Relatives	Hybrids and/or Hybridization	Evidence of Crop Introgression	Evidence of Adaptiveness	Source
Maize (<i>Zea mays</i> subsp. <i>mays</i>)	<i>Z. m.</i> subsp. <i>mexicana</i> , <i>Z. m.</i> subsp. <i>parviglumis</i>	X	X	X	[53]
Asian Rice (<i>Oryza sativa</i>)	<i>O. rufipogon</i>	X	X	X	[48]
Barley (<i>Hordeum vulgare</i>)	<i>H. v.</i> subsp. <i>spontaneum</i>	X	X	X	[45]
Sunflower (<i>Helianthus annuus</i>)	<i>H. argophyllus</i> , <i>H. bolanderi</i> , <i>H. debilis</i> , <i>H. petiolaris</i>	X			[54]
Cassava (<i>Manihot esculenta</i>)	<i>M. glaziovii</i>	X	X	X	[55]
Potato (<i>Solanum tuberosum</i>)	many	X	X	X	[56? ?]
Tomato (<i>Solanum lycopersicum</i>)	<i>S. pimpinellifolium</i>	X	X	X	[57]
Olive (<i>Olea europaea</i> ssp. <i>europaea</i> var. <i>sativa</i>)	<i>O. e.</i> ssp. <i>europaea</i> var. <i>sylvestris</i>	X	X		[58]
Soybeans (<i>Glycine max</i>)	<i>G. soja</i>	X	X		[59]
Common Bean (<i>Phaseolus vulgaris</i>)	<i>P. v.</i> var. <i>aborigineus</i> , <i>P. v.</i> var. <i>mexicanus</i> [[not in this source]]	X	X		[60]
Grapes (<i>Vitis vinifera</i> subsp. <i>vinifera</i>)	<i>V. v.</i> subsp. <i>sylvestris</i>	X	X		[61]
Sorghum (<i>Sorghum bicolor</i> subsp. <i>bicolor</i>)	<i>S. b.</i> subsp. <i>arundinaceum</i> , <i>S. b.</i> subsp. <i>drummondii</i>	X	X		[62]
Wheat (<i>Triticum monococcum</i> , <i>T. dicoccum</i> , <i>T. aestivum</i>)	<i>T. m. boeoticum</i> , <i>T. diocoides</i> , <i>T. urartu</i> , <i>Aegilops speltoides</i> , <i>A. tauschii</i>	X	X		[63]
Apple (<i>Malus domestica</i>)	<i>M. sylvestris</i> , <i>M. orientalis</i> , <i>M. baccata</i> , <i>M. sieversii</i>	X	X		[64]

The addition of ongoing gene flow to our understanding of crop demography could bear importantly on fundamental questions of crop domestication:

What is the progenitor of a crop?

Depending on the extent of post-domestication gene flow with new wild relatives, identification of a crop’s progenitor can be complicated or confounded entirely. Introgression between a crop and newly-encountered populations and species decreases the level of divergence. This signal could be mistaken for origin rather than gene flow. For example, when determining a single origin of maize from *parviglumis*, Matsuoka and colleagues [32] identified a paradox: while *parviglumis* is found exclusively in the lowlands of southwest Mexico, maize with allele frequencies most similar to *parviglumis* was found in the highlands of the Mexican Central Plateau. Several years later, van Heerwaarden *et al.* [34] resolved the paradox by determining that widespread introgression in the highlands from *mexicana*, which is closely related to *parviglumis*, has caused maize from this region to appear ancestral. Similarly, extensive post-domestication adaptive introgression from potato wild relatives long obscured this crop’s origin. Recent work has shown that, following the original domestication event of *Solanum tuberosum* in the central Andes, potato received introgression from as many as four additional species during colonization of the highest elevations of the Andes and the lowlands of the Chilean coast [65, 66]. *This sentence needs to be updated with the new potato publication.*

Beyond confounding detection of progenitor taxa, extensive introgression may necessitate a reevaluation of crop origins. In cases like maize and potato it is important to recognize the substantial contributions of introgressing taxa to the genetic base of modern crops. Broad recognition of the role these wild relatives have played in crop adaptation could further their use in breeding and elevate their conservation status.

When was a crop domesticated?

Estimates of the timing of initial domestication are often based on levels of sequence divergence between a crop and populations of its presumed progenitor (*e.g.*, [32, 47]). In highly introgressed domesticates, these estimates will be based on comparison of either crop or introgressant haplotypes

to those of the progenitor. In such cases, divergence is a mixture of time since domestication and time since split of the progenitor and the introgressing species. This phenomenon, in combination with divergence of samples from true ancestral populations, ongoing evolution of crop progenitors, and problems with assuming evolution under a molecular clock [67], may explain discrepancies between domestication dates based on genetic and archaeological data. *maybe some citations for a protracted domestication time: African rice NG paper, our paper, grape PNAS paper* More accurate estimates of the timing of domestication may be obtained from genetic data by excluding loci that show signatures of introgression. *excluding the introgression regions might cause some inference method not working properly, such as MSMC, as it would reduce heterozygosity. Suggestions: "More accurate estimates of the timing of domestication may be obtained from genetic data by excluding loci that show signatures of introgression and / or developing new statistic tools which taking potential introgression into account."*

How was genome-wide diversity impacted by a domestication bottleneck?

I suggest to delete "impacted by a domestication bottleneck" because that might confuse readers. Adaptive introgression could affect genome-wide diversity, owing to that introgressing haplotype have a higher historical N_e and domesticates have a lower historical N_e due to bottleneck. Adding "impacted by a domestication bottleneck" seems making the message not direct and not complete. Measurement of the strength of the initial domestication bottleneck may also be impacted by adaptive introgression during the spread of crops. Crop wild relatives have distinct demographies when compared to domesticates and may therefore have contrasting effective population sizes (N_e). The influence of wild relative introgression on estimates of the domestication bottleneck will depend on a number of factors including the magnitude of gene flow, the N_e of the introgressing taxon, and the strength of selection on haplotypes following introgression. For example, substantial introgression at neutral loci from a wild taxon with a historically higher N_e will lead to underestimates of the overall strength of the domestication bottleneck. *Does Li's research fit in well here? Should we add her citation?*

What candidate genes were targeted by selection during domestication?

Loci targeted by selection during domestication can be identified through so-called "bottom-up" approaches based on population genetic signatures [68]. Ideally, candidate loci will be identified by first constructing a demographic model representing the history of the domesticate. In this approach, diversity data from neutral loci are fit to potential models of a crop's demography and then statistical tests of selection are used to identify candidate domestication genes under the most likely model. Due to the difficulty of this approach and the uncertainty associated with any given demography, many studies identify domestication loci using a strict outlier approach in which loci showing the greatest reduction in nucleotide diversity or the highest allele frequency differentiation in the domesticate relative to the wild progenitor are identified as candidates. Introgression during crop expansion may influence candidate gene detection using both demographic-modeling and strict-outlier approaches. For example, *mexicana* introgression into maize described above accounts for approximately 20% of the genome of maize in the highlands of Mexico [34]. Takuno and co-authors [69] have shown that a demographic model incorporating this introgression is a significantly better fit to empirical data than a model lacking introgression. Failure to account for introgression in maize would therefore compromise domestication candidate detection, particularly if a study contained maize samples from the Mexican highlands. Likewise, introgression that increased nucleotide diversity in the domesticate or decreased differentiation at domestication loci would confound a strict outlier approach. However, previous work, also in maize, has shown that known domestication loci are particularly resistant to introgression [53], likely due to ongoing

selection favoring the domesticated phenotype.

Future studies in crop-wild introgression

Basic:

Research has so far shown that adaptive crop-wild introgression has played a significant role in the domestication and dispersal of many agronomically-important crops. However, the dynamics of this process are not yet fully understood, especially in the context of individual case examples, and many questions remain.

What is the genomic architecture of this introgression, and does the architecture suggest that it has been adaptive? At what geographic scale is introgression adaptive? *Barley and maize most strongly show geographic patterns that we can talk about here, but I'm not sure how to discuss them again.* To what extent does this depend on the slope of environmental gradients such as temperature, precipitation, and elevation across these ranges? *Refer to the map and bar graph here?* We can look at conservation of genomic architecture across landscapes and between populations, and make predictions about introgressions and their relations to local adaptation. If the genomic architecture of an introgressed region is conserved across a broad ecogeographical region, this suggests that the introgression imparts adaptation to general environmental or climatic variables. On the other hand, if the genomic architecture is conserved within populations but not between nearby populations in the region, this suggests that the introgressed regions offer adaptations to more local selective pressures. If the genomic architecture of an introgressed region is not conserved within a population, there is little evidence that the introgression is adaptive. *Determination of adaptiveness could be accomplished genomically (looking for high LD regions that indicate selective sweeps) or by functional identification?*

After hybridization events that lead to introgression, how long might the detectable genomic signals of introgression persist? Introgressed regions are easier to detect when there has been limited recombination to break them apart. Therefore, introgressions are easiest to detect when they are either recent (few generations means few recombination events) or involve structural variation (which diminishes recombination rate). *involve $-r$ involving* Because recombination progressively breaks apart LD in introgressed regions, measurements of LD can be used to date introgression events (as in [45]).

At what taxonomic scale does introgression occur? When do species become so diverged that introgression is either maladaptive or impossible (due to Dobzhansky-Muller incompatibilities or other mechanisms)? Theory indicates that the most significant limiting factor to gene flow between progenitor and domesticate is divergence time. Over time, diverged populations drift and become increasingly incompatible. Small effective population size, and correspondingly high genetic load, of the introgressive population also limits gene flow. Although perhaps less applicable in crop systems, this effect is seen in other well-documented cases of introgression (for example, Neanderthal introgression into humans, [?]). *I tracked down the sources that indicate that introgression is suppressed around genes. The case example was Neanderthal alleles into humans, the paper titled "The genetic history of Ice Age Europe" [?], cited by the Graham Coop paper. However, I'm unclear if this is a general rule, or simply a consequence of Neanderthal being high in deleterious genetic load. Wouldn't crop systems, with wild relatives harboring putatively adaptive alleles with low load, show a different pattern? I could also talk about cross-incompatibility factors in maize/teosinte hybridization vs. highly compatible asian rice and its relatives. I probably wouldn't need a new source for the rice, but I might need to cite a paper about tcb1, ga1, and ga2.*

Applied:

Our identification and understanding of introgression in agroecosystems would be augmented by the development of wild relative genomic resources, such as annotated genomic sequence assemblies and functional genomic data sets [?]. Additional study of introgression in agroecosystems could lead to advances in both basic and applied genetics, and specifically the continued improvement of modern crops. Loci underlying the domesticated phenotype can be more clearly identified by removing the confounding population genetic signals of introgression. These loci are potentially beneficial targets for crop improvement. Furthermore, adaptive introgression that is demonstrably tied to a specific environment may include beneficial alleles that can be utilized in crop breeding.

I could here include a handful of examples of crops that farmers have bred adaptive alleles into. Also, we could mention that identification of wild relatives responsible for valuable agronomic introgressed alleles could inform conservation efforts.

Conclusions

The study of crop domestication has been revolutionized by the advent and application of genomic tools. The genomes of crops and their wild relatives tell a story of give-and-take that extends well beyond the initial stages of domestication. Likewise, population genetic theory reinforces the proclivity of wild relatives to provide advantageous, locally-adapted alleles to crops as they disperse beyond their domestication centers into new geographies with new ecological pressures and niches.

References

- [1] J. R. Harlan, *Crops & Man*. Madison, WI: American Society of Agronomy, Crop Science Society of America, 1992.
- [2] T. A. Brown, M. K. Jones, W. Powell, and R. G. Allaby, “The complex origins of domesticated crops in the fertile crescent,” *Trends in Ecology & Evolution*, vol. 24, no. 2, pp. 103–109, 2009.
- [3] R. S. Meyer, J. Y. Choi, M. Sanches, A. Plessis, J. M. Flowers, J. Amas, K. Dorph, A. Barretto, B. Gross, D. Q. Fuller, I. K. Bimpong, M.-N. Ndjiondjop, K. M. Hazzouri, G. B. Gregorio, and M. D. Purugganan, “Domestication history and geographical adaptation inferred from a snp map of african rice,” *Nat Genet*, vol. 48, pp. 1083–1088, 09 2016.
- [4] L. Wang, T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, and M. Hufford, “The interplay of demography and selection during maize domestication and expansion,” *bioRxiv*, p. 114579, 2017.
- [5] Y. Zhou, M. Massonnet, J. Sanjak, D. Cantu, and B. S. Gaut, “The evolutionary genomics of grape (*vitis vinifera* ssp. *vinifera*) domestication,” *bioRxiv*, p. 146373, 2017.
- [6] D. Q. Fuller, T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, and M. D. Purugganan, “Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 17, pp. 6147–6152, 2014.
- [7] N. Barton, “The role of hybridization in evolution,” *Molecular Ecology*, vol. 10, no. 3, pp. 551–568, 2001.

- [8] C. Feuillet, P. Langridge, and R. Waugh, “Cereal breeding takes a walk on the wild side,” *Trends in Genetics*, vol. 24, no. 1, pp. 24 – 32, 2008.
- [9] C. N. Stewart, M. D. Halfhill, and S. I. Warwick, “Transgene introgression from genetically modified crops to their wild relatives,” *Nature Reviews Genetics*, vol. 4, no. 10, pp. 806–817, 2003.
- [10] H. Dempewolf, G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino, “Past and future use of wild relatives in crop breeding,” vol. 57, pp. 1070–1082, 2017.
- [11] S. H. Martin, J. W. Davey, and C. D. Jiggins, “Evaluating the use of abba–baba statistics to locate introgressed loci,” *Molecular biology and evolution*, vol. 32, no. 1, pp. 244–257, 2015.
- [12] A. L. Price, A. Tandon, N. Patterson, K. C. Barnes, N. Rafaels, I. Ruczinski, T. H. Beaty, R. Mathias, D. Reich, and S. Myers, “Sensitive detection of chromosomal segments of distinct ancestry in admixed populations,” *PLoS Genet*, vol. 5, no. 6, pp. 1–18, 2009.
- [13] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush, “Inference of population structure using dense haplotype data,” *PLoS Genet*, vol. 8, no. 1, pp. 1–16, 2012.
- [14] J. B. Pease and M. W. Hahn, “Detection and polarization of introgression in a five-taxon phylogeny,” *Systematic biology*, vol. 64, no. 4, pp. 651–662, 2015.
- [15] B. K. Rosenzweig, J. B. Pease, N. J. Besansky, and M. W. Hahn, “Powerful methods for detecting introgressed regions from population genomic data,” *Molecular ecology*, 2016.
- [16] A. J. Geneva, C. A. Muirhead, S. B. Kingan, and D. Garrigan, “A new method to scan genomes for introgression in a secondary contact model,” *PloS one*, vol. 10, no. 4, p. e0118621, 2015.
- [17] M. A. Beaumont, W. Zhang, and D. J. Balding, “Approximate bayesian computation in population genetics,” *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.
- [18] R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, “Inferring the joint demographic history of multiple populations from multidimensional snp frequency data,” *PLoS Genet*, vol. 5, no. 10, p. e1000695, 2009.
- [19] J. Hey and R. Nielsen, “Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of drosophila pseudoobscura and d. persimilis,” *Genetics*, vol. 167, no. 2, pp. 747–760, 2004.
- [20] H. Li and R. Durbin, “Inference of human population history from individual whole-genome sequences,” *Nature*, vol. 475, no. 7357, pp. 493–496, 2011.
- [21] S. Schiffels and R. Durbin, “Inferring human population size and separation history from multiple genome sequences,” *Nature genetics*, vol. 46, no. 8, pp. 919–925, 2014.
- [22] J. Terhorst, J. A. Kamm, and Y. S. Song, “Robust and scalable inference of population history from hundreds of unphased whole genomes,” tech. rep., 2017.
- [23] F. Racimo, D. Marnetto, and E. Huerta-Sánchez, “Signatures of archaic adaptive introgression in present-day human populations,” *Molecular Biology and Evolution*, p. msw216, 2016.
- [24] J. G. Schraiber and J. M. Akey, “Methods and models for unravelling human evolutionary history,” *Nature Reviews Genetics*, 2015.

- [25] A. Brisbin, K. Bryc, J. Byrnes, F. Zakharia, L. Omberg, J. Degenhardt, A. Reynolds, H. Ostrer, J. G. Mezey, and C. D. Bustamante, “Pcadmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations,” *Human biology*, vol. 84, no. 4, p. 343, 2012.
- [26] S. Sankararaman, S. Sridhar, G. Kimmel, and E. Halperin, “Estimating local ancestry in admixed populations,” *The American Journal of Human Genetics*, vol. 82, no. 2, pp. 290–303, 2008.
- [27] J. K. Pritchard, M. Stephens, and P. Donnelly, “Inference of population structure using multilocus genotype data,” *Genetics*, vol. 155, no. 2, pp. 945–959, 2000.
- [28] F. Racimo, S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez, “Evidence for archaic adaptive introgression in humans,” *Nature Reviews Genetics*, vol. 16, no. 6, pp. 359–371, 2015.
- [29] T. H. G. Consortium, “Butterfly genome reveals promiscuous exchange of mimicry adaptations among species,” *Nature*, vol. 487, no. 7405, pp. 94–98, 2012.
- [30] L. C. Norris, B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro, “Adaptive introgression in an african malaria mosquito coincident with the increased usage of insecticide-treated bed nets,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 3, pp. 815–820, 2015.
- [31] S. Stankowski and M. A. Streisfeld, “Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 282, no. 1814, 2015.
- [32] Y. Matsuoka, Y. Vigouroux, M. M. Goodman, J. Sanchez, E. Buckler, and J. Doebley, “A single domestication for maize shown by multilocus microsatellite genotyping,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 6080–6084, 2002.
- [33] H. Wilkes, “Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize,” *Economic Botany*, pp. 254–293, 1977.
- [34] J. van Heerwaarden, J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. d. J. S. Gonzalez, and J. Ross-Ibarra, “Genetic signals of origin, spread, and introgression in a large sample of maize landraces,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 3, pp. 1088–1092, 2011.
- [35] J. Doebley, M. M. Goodman, and C. W. Stuber, “Patterns of isozyme variation between maize and mexican annual teosinte,” *Economic Botany*, vol. 41, no. 2, pp. 234–246, 1987.
- [36] M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra, “The genomic signature of crop-wild introgression in maize,” *Plos genetics*, 2013.
- [37] N. Lauter, C. Gustus, A. Westerbergh, and J. Doebley, “The inheritance and evolution of leaf pigmentation and pubescence in teosinte,” *Genetics*, vol. 167, no. 4, pp. 1949–1959, 2004.
- [38] R. Takahashi, “The origin and evolution of cultivated barley,” *Advances in genetics*, vol. 7, pp. 227–266, 1955.
- [39] A. Badr, R. Sch, H. El Rabey, S. Effgen, H. Ibrahim, C. Pozzi, W. Rohde, F. Salamini, *et al.*, “On the origin and domestication history of barley (*hordeum vulgare*),” *Molecular Biology and Evolution*, vol. 17, no. 4, pp. 499–510, 2000.

- [40] P. Azhaguvel and T. Komatsuda, "A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley," *Annals of botany*, vol. 100, no. 5, pp. 1009–1015, 2007.
- [41] G. Haberer and K. F. Mayer, "Barley: From brittle to stable harvest," *Cell*, vol. 162, no. 3, pp. 469–471, 2015.
- [42] F. Dai, E. Nevo, D. Wu, J. Comadran, M. Zhou, L. Qiu, Z. Chen, A. Beiles, G. Chen, and G. Zhang, "Tibet is one of the centers of domestication of cultivated barley," *Proceedings of the National Academy of Sciences*, vol. 109, no. 42, pp. 16969–16973, 2012.
- [43] E. Nevo and G. Chen, "Drought and salt tolerances in wild relatives for wheat and barley improvement," *Plant, cell & environment*, vol. 33, no. 4, pp. 670–685, 2010.
- [44] G. Hillman, R. Hedges, A. Moore, S. Colledge, and P. Pettitt, "New evidence of lateglacial cereal cultivation at abu hureyra on the euphrates," *The Holocene*, vol. 11, no. 4, pp. 383–393, 2001.
- [45] A. M. Poets, Z. Fang, M. T. Clegg, and P. L. Morrell, "Barley landraces are characterized by geographically heterogeneous genomic origins," *Genome Biology*, vol. 16, no. 1, pp. 1–11, 2015.
- [46] D. Q. Fuller, Y.-I. Sato, C. Castillo, L. Qin, A. R. Weisskopf, E. J. Kingwell-Banham, J. Song, S.-M. Ahn, and J. Van Etten, "Consilience of genetics and archaeobotany in the entangled history of rice," *Archaeological and Anthropological Sciences*, vol. 2, no. 2, pp. 115–131, 2010.
- [47] J. Molina, M. Sikora, N. Garud, J. M. Flowers, S. Rubinstein, A. Reynolds, P. Huang, S. Jackson, B. A. Schaal, C. D. Bustamante, *et al.*, "Molecular evidence for a single evolutionary origin of domesticated rice," *Proceedings of the National Academy of Sciences*, vol. 108, no. 20, pp. 8351–8356, 2011.
- [48] X. Huang, N. Kurata, X. Wei, Z.-X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu, W. Li, Y. Guo, Y. Lu, C. Zhou, D. Fan, Q. Weng, C. Zhu, T. Huang, L. Zhang, Y. Wang, L. Feng, H. Furuumi, T. Kubo, T. Miyabayashi, X. Yuan, Q. Xu, G. Dong, Q. Zhan, C. Li, A. Fujiyama, A. Toyoda, T. Lu, Q. Feng, Q. Qian, J. Li, and B. Han, "A map of rice genome variation reveals the origin of cultivated rice," *Nature*, vol. 490, pp. 497–501, 2012.
- [49] D. A. Vaughan, B.-R. Lu, and N. Tomooka, "The evolving story of rice evolution," *Plant science*, vol. 174, no. 4, pp. 394–408, 2008.
- [50] G. Second, "Origin of the genic diversity of cultivated rice (*oryza* spp.): study of the polymorphism scored at 40 isozyme loci.," , vol. 57, no. 1, pp. 25–57, 1982.
- [51] K. Zhao, M. Wright, J. Kimball, G. Eizenga, A. McClung, M. Kovach, W. Tyagi, M. L. Ali, C.-W. Tung, A. Reynolds, *et al.*, "Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome," *PloS one*, vol. 5, no. 5, p. e10780, 2010.
- [52] H. Wang, F. G. Vieira, J. E. Crawford, C. Chu, and R. Nielsen, "Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice," *Genome Research*, vol. 27, no. 6, pp. 1029–1038, 2017.

- [53] M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra, “The genomic signature of crop-wild introgression in maize,” 2013.
- [54] L. H. Rieseberg, S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay, “Hybridization and the colonization of novel habitats by annual sunflowers,” *Genetica*, vol. 129, no. 2, pp. 149–165, 2007.
- [55] J. V. Bredeson, J. B. Lyons, S. E. Prochnik, G. A. Wu, C. M. Ha, E. Edsinger-Gonzales, J. Grimwood, J. Schmutz, I. Y. Rabbi, C. Egesi, *et al.*, “Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity,” *Nature biotechnology*, vol. 34, no. 5, pp. 562–570, 2016.
- [56] T. Johns and S. L. Keen, “Ongoing evolution of the potato on the altiplano of western bolivia,” *Economic Botany*, vol. 40, no. 4, pp. 409–424, 1986.
- [57] C. M. Rick, “The role of natural hybridization in the derivation of cultivated tomatoes of western south america,” *Economic Botany*, vol. 12, no. 4, pp. 346–367, 1958.
- [58] C. M. Diez, I. Trujillo, N. Martinez-Urdiroz, D. Barranco, L. Rallo, P. Marfil, and B. S. Gaut, “Olive domestication and diversification in the mediterranean basin,” *New Phytologist*, vol. 206, no. 1, pp. 436–447, 2015.
- [59] H.-M. Lam, X. Xu, X. Liu, W. Chen, G. Yang, F.-L. Wong, M.-W. Li, W. He, N. Qin, B. Wang, *et al.*, “Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection,” *Nature genetics*, vol. 42, no. 12, pp. 1053–1059, 2010.
- [60] R. Papa and P. Gepts, “Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*phaseolus vulgaris* l.) from mesoamerica,” *Theoretical and Applied Genetics*, vol. 106, no. 2, pp. 239–250, 2003.
- [61] S. Myles, A. R. Boyko, C. L. Owens, P. J. Brown, F. Grassi, M. K. Aradhya, B. Prins, A. Reynolds, J.-M. Chia, D. Ware, *et al.*, “Genetic structure and domestication history of the grape,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 9, pp. 3530–3535, 2011.
- [62] P. Aldrich, J. Doebley, K. Schertz, and A. Stec, “Patterns of allozyme variation in cultivated and wild sorghum bicolor,” *Theoretical and Applied Genetics*, vol. 85, no. 4, pp. 451–460, 1992.
- [63] D. Zohary, J. R. Harlan, and A. Vardi, “The wild diploid progenitors of wheat and their breeding value,” *Euphytica*, vol. 18, no. 1, pp. 58–65, 1969.
- [64] A. Cornille, P. Gladieux, M. J. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, J. Clavel, M. Olonova, L. Feugey, *et al.*, “New insight into the history of domesticated apple: secondary contribution of the european wild apple to the genome of cultivated varieties,” *PLoS Genet*, vol. 8, no. 5, p. e1002703, 2012.
- [65] D. M. Spooner, M. Ghislain, R. Simon, S. H. Jansky, and T. Gavrilenko, “Systematics, diversity, genetics, and evolution of wild and cultivated potatoes,” *The Botanical Review*, vol. 80, no. 4, pp. 283–383.
- [66] T. Gavrilenko, O. Antonova, A. Shuvalova, E. Krylova, N. Alpatyeva, D. M. Spooner, and L. Novikova, “Genetic diversity and origin of cultivated potatoes based on plastid microsatellite polymorphism,” *Genetic Resources and Crop Evolution*, vol. 60, no. 7, pp. 1997–2015, 2013.

- [67] M. A. Zeder, E. Emshwiller, B. D. Smith, and D. G. Bradley, “Documenting domestication: the intersection of genetics and archaeology,” *Trends in Genetics*, vol. 22, no. 3, pp. 139–155, 2006.
- [68] J. Ross-Ibarra, P. L. Morrell, and B. S. Gaut, “Plant domestication, a unique opportunity to identify the genetic basis of adaptation,” *Proceedings of the National Academy of Sciences*, vol. 104, no. suppl 1, pp. 8641–8648, 2007.
- [69] S. Takuno, P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra, “Independent molecular basis of convergent highland adaptation in maize,” *Genetics*, 2015.

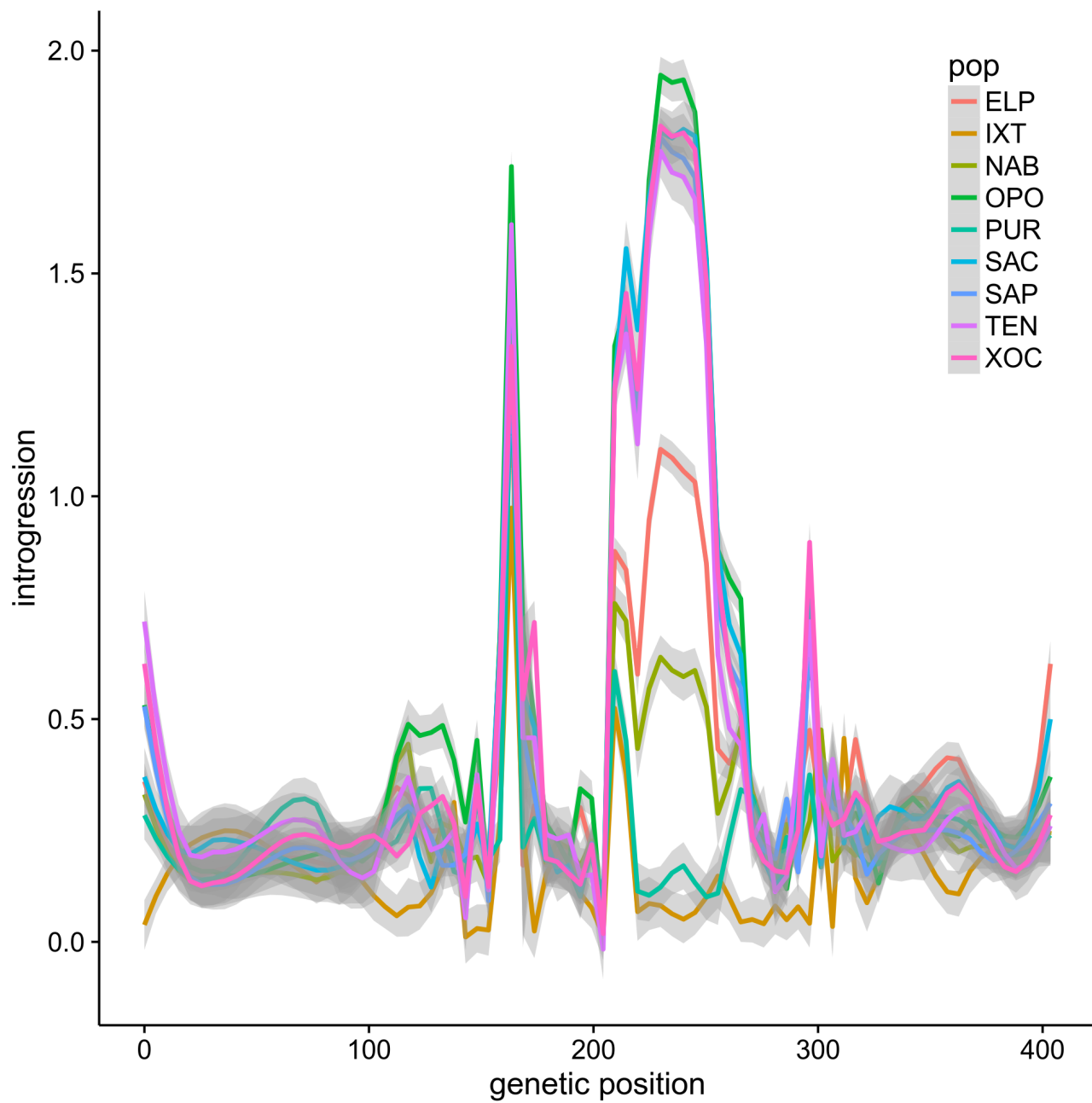


Figure 1: Li's caption here.

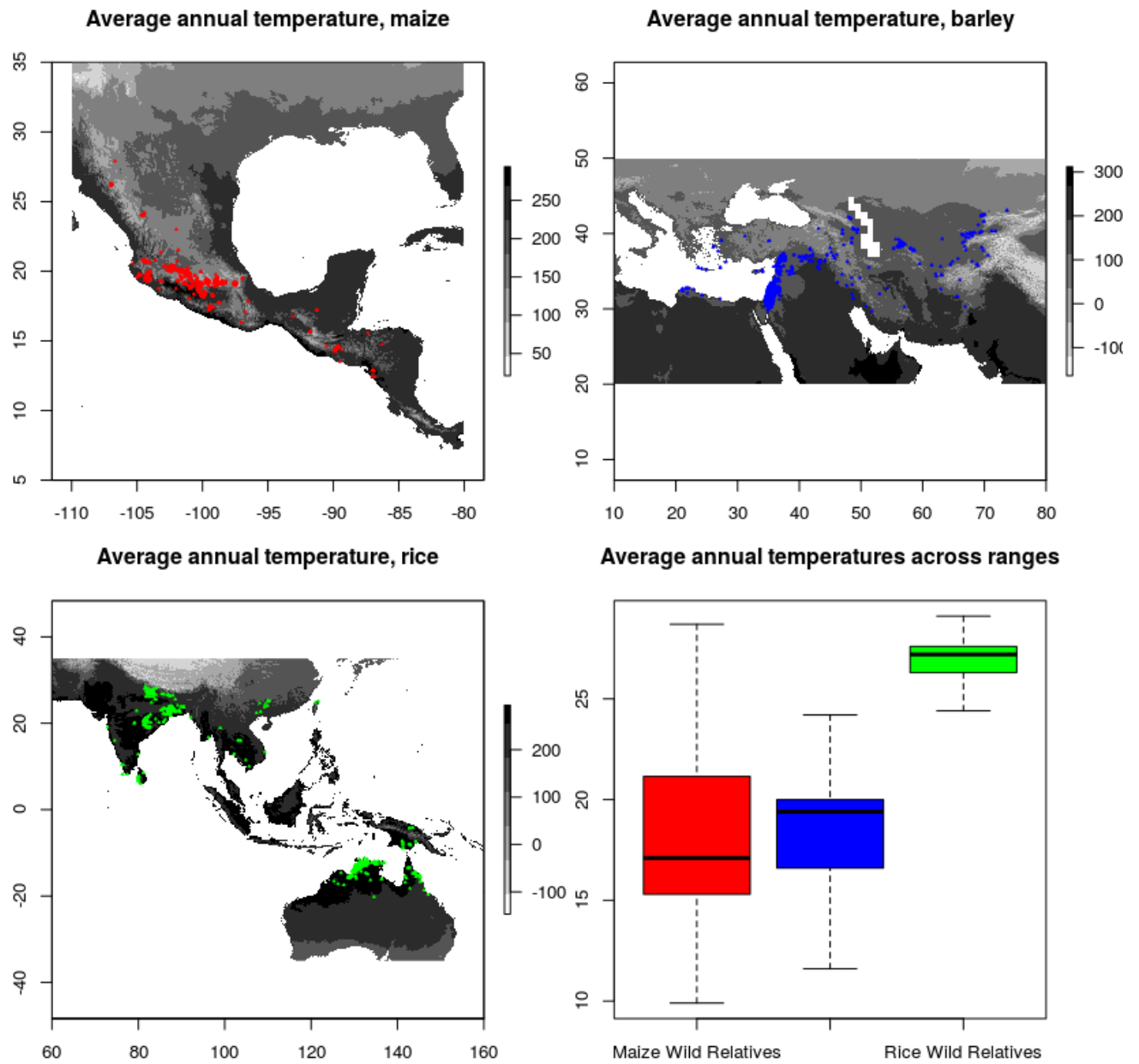


Figure 2: Map of the natural ranges of wild relatives of three domesticated crops, overlaid with average annual temperature. The distribution of average annual temperature experienced in the geographic home ranges of wild relatives interfertile with four crops