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

Authors: Garrett M. Janzen[1], Li Wang[1], and Matthew B. Hufford[1,*]

[1]Department 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 [?]. However, archaeological and genetic evidence are beginning to reveal that, in many cases, domestication has been temporally protracted and geographically diffuse [? ? ? ? ?]. An 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 [? ?]. 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, [?]) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, [?]). The history of natural 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.

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 adaptive 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**). 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 [**? ? ? ? ? ?**]. 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) [**?**], diffusion approximations for demographic inference ($\delta a \delta i$) [**?**], isolation with migration models [**?**], and a series of methods utilizing the sequentially Markovian coalescent (PSMC, MSMC and SMC++) [**? ? ?**]). 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 derivates including G_{min} [**?**] and RND_{min}[**?**] 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:

$$RND_{min} = \frac{d_{min}}{d_{out}} \tag{1}$$

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) [**?**]. 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 (C). 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) [**?**]. Further modifications have allowed specification of more than one source population (see details in [**?**]).

Second, local ancestry deconvolution (also known as chromosome painting) assigns genomic regions to various source populations based on patterns of allele/haplotype sharing [**?**]. 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 [**?**]). Another clusters admixed populations with reference samples using a sliding-window approach (*e.g.*, PCAdmix [**?**] and LAMP [**?**]). And finally, introgression can be detected through chromosome painting by using a Bayesian model [**?**] in which deviations from Hardy-Weinberg equilibrium are minimized through creation of genetic groups (*e.g.*, fineSTRUCTURE [**?**]).

Li, are you familiar with the analytical tools MIGRATE-N and BAYESASS? Rieseberg comapres these two to STRUCTURE at some length in the discussion of this paper: https://biology.unm.edu/Whitney/WhitneyShould we include these two methods, even in passing, in this portion of the paper?

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 $\hat{f}_d$ [**?**] and the five-taxon D-statistic [**?**]. 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.

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 [**?**]); introgression has conferred Müllerian mimicry *I would explain a bit more here...wing coloration loci, protects against predation...* across butterfly species [**?**]; introgression has spread insecticide resistance across mosquito species [**?**], and introgression across *Mimulus* (*i.e.*, monkeyflower) species has resulted in adaptation to pollinator preference and contributed to speciation [**?**].

Crop adaptation through introgression

Over the last few years, several high-profile publications based on genome-wide data have documented introgression between crops and their wild relatives outside putative domestication centers. Recent empirical studies have revealed that introgression has occurred in many of the world's most important crops (**Table 2**).

1. Maize:

   The relationship between maize (*Zea mays* ssp. *mays*) and the teosinte *Zea mays* ssp. *mexicana* (hereafter referred to as *mexicana*) offers a prime case study of adaptive wild-to-crop introgression. Maize was domesticated from *Zea mays* ssp. *parviglumis* (hereafter referred to as *parviglumis*) approximately 9,000 years ago in the lowlands of the Balsas River Valley in Mexico [**?**]. 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 long been reported based on both morphological [**? ? ?**] and molecular [**? ? ? ? ?**] data. However, [**?**] 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 using both HAPMIX and the linkage model of STRUCTURE (Figure 1). These introgressed segments showed low diversity and overlapped QTL that had previously been found to control anthocyanin content and leaf macrohairs [**?**], 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 Mexico, yet maize has colonized and adapted to high elevation in a number of other regions. A recent study [**?**] 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 maize landraces with *mexicana* introgression were transferred to nearby high elevation regions in Guatemala and the southwestern United States, but more distant high elevation regions (*e.g.,* the Andes) showed no *mexicana* ancestry.

2. Barley:

   Barley (*Hordeum vulgare* subsp. *vulgare*) was domesticated at least twice roughly 10,000 BP: once from the wild subsp. *spontaneum* in the Fertile Crescent and once from subsp. *spontaneum* var. *agriocrithon* in Tibet [**? ? ? ? ?**]. *the source morrell2007genetic will be used*

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 [**?**]. Barley/*spontaneum* hybrids are known to be fertile and are found spontaneously when wild and domesticated barleys co-occur. In some cases, wild-to-crop introgression has been shown to occur over distances of more than a kilometer [**?**].

I think here a more focused discussion of potential for adaptive introgression based on the Poets paper is needed The authors of [**?**] used STRUCTURE to look for patterns of introgression from wild relatives in a dataset of 803 landraces, and found a high amount variability in the amount of contribution from wild relatives, as well as its location in the genome, within barley populations. This is indicative of contribution from numerous wild populations. Furthermore, the authors found that wild introgression contribution is generally greatest from geographically-proximate populations, and that introgressed regions might be combined from geographically-separate wild populations. Low linkage disequilibrium and small blocks of identity by state indicate that these introgressed regions are old, perhaps dating back to the beginning of barley domestication. As landraces and nearby wild relatives share similar genomic sequences, the introgressed regions that are exclusive to that landrace are more likely to contain adaptive alleles. Such alleles were not identified specifically, though wild-domesticate breeding experiments have shown that wild barleys have alleles for several important agronomic phenotypes, including powdery mildew resistance, brittleness, flowering time, plant height, lodging, and yield [**? ? ?**].

3. Sunflower:

   The common sunflower (*Helianthus annuus*) shows evidence of domestication in the eastern United States [**? ?**] with potential for a second domestication center in Mexico [**?**]. The pre-Columbian *H. annuus* distribution of cultivated sunflower spanned much of the Great Plains, from what is now north-central Texas to Montana and North Dakota (see figure 1 of [**?**]). Domesticated sunflower has long lived in sympatry with wild relatives such as *H. petiolaris* and *H. bolanderi* and forms stable hybrid populations with these taxa [**? ? ?**]. Wild sunflowers are known to be locally-adapted, and weedy hybrid populations often share these adaptations [**?**]. However, the most striking example of adaptive introgression within *Helianthus* is that of the cucumberleaf sunflower, *H. debilis* ssp. *cucumerifolius*. *I believe* H. annuus *ssp.* texanus *is a wild sunflower, right? If so, this is more an example of hybrid speciation rather then adaptive wild-to-crop gene flow* Cucumberleaf sunflower is endemic to south-central Texas, and exhibits several adaptations to the region. Introgressive hybridization imparted locally-adapted alleles from *H. debilis* to *H. annuus* via introgressive hybridization [**?**]. These introgressed hybrids formed a new lineage of sunflower (*H. annuus* ssp. *texanus*, *H. a. texanus* hereafter) which displays *H. debilis*-like traits adaptive to south-central Texas climate and ecology. These adaptive *debilis*-like traits include resistance to herbivorous pests and an increased branching plant architecture, as well as higher overall fitness than *H. annuus* (as measured by higher seed production [**?**]). Although H. annuus and *H. a. texanus* are interfertile, *H. a. texanus* displays persistent phenotypic differences from *H. annuus* [**?**].

   The genome of the common sunflower has been greatly influenced by introgression from wild relatives, due to both natural outcrossing events and concerted breeding efforts in crop improvement. *Helianthus* has several genes for downy mildew resistance, and each imparts

resistance to one or more races of *Plasmopara halstedii*, one of the most agronomically important diseases in sunflower cultivation [?]. Some of these downy mildew resistence genes were found in wild relatives (including *H. argophyllus*, *H. tuberosus*, and *H. praecox*) and have been successfully bred into modern *H. annuus* [?]. PlArg, an allele found in wild silverleaf sunflowers (*H. argophyllus*, inbred line Arg1575-2), confers resistance to all known (20 or more) races of downey mildew [?] while others (Pl1-Pl11) are effective for one or more types [?]. Silverleaf sunflower has also been the focus of drought resistance breeding efforts [?] and *Phomopsis* resistance breeding efforts [?]. *H. annuus* shows signs of persistent introgressive hybridization with *H. petiolaris* with evidence of positive selection driving some of the genetic differentiation between the two species [?].

Recent investigations into the history of *Helianthus* introgression have implemented genomic methods. [?] analyzed transcriptome sequence variation on cultivated and wild *H. annuus*, *H. petiolaris*, and *H. argophyllus*. Using STRUCTURE, these authors found that introgressions from wild relatives exist on every chromosome in at least one modern line, covering over 10% of the genome. Of particular note is the modern line RHA 274, a modern line which was bred with *H. a. texanus* in the 1970s to restore a branching plant body architecture, which allows the plant to produce pollen for a longer period of time, increasing seed production. RHA 274 has several large introgression from *H. a. texanus*, including one at the site of HaGNAT, the domestication gene associated with branching. These introgressed regions are not found in the non-branching lines Sunrise and VNIIMK8931, further suggesting that the *H. a. texanus* introgressed regions are causative.

Notes for review:

http://biology.unm.edu/Whitney/Whitney

http://www.jstor.org/stable/pdf/2409227.pdf?refreqid=excelsior:a686b7f06f9daa2b89bd5a654dade3b4

file:///home/gjanzen/Downloads/Whitney$_e t_a l - 2010 - New_P hytologist$

http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2010.03234.x/epdf

https://biology.unm.edu/Whitney/Whitney

Genome scan of hybridizing sunflowers from Texas (Helianthus annuus and H. debilis) reveals asymmetric patterns of introgression and small islands of genomic differentiation:

In contrast, coalescent analyses of long-term migration rates imply that interspecific migration has been more important than mutation in providing new genetic variation in ssp. texanus populations. Immigration from H. debilis cucumerifolius into ssp. texanus (M = 4.05; Nem = 1.74) is higher than median estimates of intraspecific migration rates in plants (Morjan and Rieseberg 2004) and only slightly lower than immigration from ssp. annuus into ssp. texanus (M = 5.00). Immigration rates from ssp. texanus into H. debilis cucumerifolius and ssp. annuus are still higher (M = 6.30 and 7.29, respectively), implying that the putative hybrid lineage serves as a bridge for migration between the parental species.

One puzzle is that the estimated immigration rate between the allopatric taxa (H. debilis cucumerifolius and ssp. annuus) is approximately comparable with that between the sympatric taxa (H. debilis cucumerifolius and ssp. texanus ssp. texanus). This observation may point to the effectiveness of ssp. texanus as a bridge for the transfer of alleles between H. debilis and ssp. annuus. An alternative explanation is that the BAYESASS estimates are unreliable because homoplasy of microsatellite alleles is high in taxa with large-effective population sizes (Estoup et al. 2002) such as the annual sunflowers studied here.

Despite the apparent discrepancy between the results from STRUCTURE and BAYEASS, several conclusions can be made. Analyses with STRUCTURE support the distinctness of ssp. texanus, as well as its clear inclusion within H. annuus, as initially postulated by Heiser (1951). Likewise, the STRUCTURE, MIGRATE-N and BAYESASS analyses all confirm previous report of introgression between H. annuus and H. debilis (Rieseberg et al. 1990, 2007) and suggest that the introgression is genomewide when it occurs. Contrary to the Heisers scenario of a recent Holocene origin of ssp. texanus, however, our results are more consistent with a much longer history of contact between H. annuus and H. debilis. Otherwise, it is difficult to account for the presence of significant longterm migration between currently allopatric populations of ssp. annuus and H. debilis cucumerifolius. This revised scenario makes sense in the light of numerous glacialinterglacial cycles over the past million years (Hewitt 2000), which probably resulted in intermittent contact between H. annuus and H. debilis, with the last contact likely occurring during the Wisconsin glaciation, 18 000 BP. The current contact between the species differs from the previous ones in that it appears to have been human-aided. Possibly, some of the molecular evidence of introgression, particularly into allopatric populations, stems from past periods of contact.

One of the four outlier loci, HT0414, is associated with a heat-shock protein and was previously shown to have been the target of selection in H. annuus ssp. annuus salt marsh populations from the state of Utah (Kane and Rieseberg 2007), as well as in weedy sunflower populations from several different locations across the USA (Kane and Rieseberg 2008). In this study, populations of ssp. annuus from Kansas, Nebraska and Oklahoma monomorphic for a single allele, whereas populations of H. debilis and other populations of ssp. annuus are considerably more polymorphic. These results imply that HT0414 may be involved in adaptation to a range of different habitats or to conditions shared by several different habitats (Kane and Rieseberg 2008).

4. Asian Rice:

The story of Asian rice (*Oryza sativa*) domestication is still debated. Interspecific hybridization between wild and domesticated rices both ancient and recent has produced a phylogeny that defies clear and simple models. On one hand, some 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 and the Indian Ganges plain, respectively [?]. Others support a single center of domestication, arguing that present-day patterns of rice variation could be explained by a single domestication event in a population with high standing variation, followed by dispersal into sympatry with locally-adapted wild relatives in diverse environments, genetic admixture, and selection for adaptive alleles [?]. A recent investigation utilizing SNP data from wild and domesticated rice accesssions pointed to a single domestication occurring 8,200-13,500 BP in the Yangtze Basin in China from rufipogon, with later divergence of indica and japonica [?]. Huang and colleagues [?] developed a genetic map of rice variation, which they used to measure distance between wild and domesticated rices at and around domesication 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.

The high genetic diversity within domesticated rice is likely due to introgression from wild relatives both within the domestication center(s) and in new environments where rice has dispersed [?], and gene flow between domesticated Asian rice and its wild relatives outside of the Yangtze Basin could have imparted locally adaptive traits [?]. The wild relatives *rufipogon* and *nivara* both maintain high genotypic and phenotypic diversity and exhibit locally-adaptive traits (whereas *rufipogon* is

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

methods	data type	reference
chromosome paiting		
Hapmix	phased haplotype; reference panel	Price et al. 2009
RASPberry	phased haplotype	Wegmann et al. 2011
MultiMix	phased/unphased genotype; reference panel	Churchhouse and Marchini 201
PCAdmix	phased haplotype	Brisbin et al. 2012
LAMP	phased haplotypes; reference panel	Sankararaman et al. 2008
phylogenetic relationship		
ABBA-BABA/D-statistics	biallelic SNP	Durand et al. 2011
fd statistic	biallelic SNP	Martin et al. 2015
five taxon D statistics	biallelic SNP	Pease and Hahn 2015
divergence		
Gmin	biallelic SNP	Geneva et al. 2015
RNDmin	phased haplotype	Rosenzweig et al. 2016
(see .tex file for comment)	biallelic SNP	Racimo et al. 2016
population structure related		
fineStructure	phased haplotype	Lawson et al. 2012
Globetrotter	phased haplotype	Hellenthal et al. 2014

adapted to forested wetland environments, *nivara* is adapted to dryer conditions and has life cycle adaptations to survive grazing pressure). 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. Gene flow between wild and cultivated rices, though asymmetrical, is frequent enough that nuisance weedy hybrids commonly arise amongst cultivated populations and must be managed with difficulty [**?**].

To date, research into adaptive introgression in the domestication of rice has been insufficient to detect clearly-supported examples. [**?**] used a SNP panel and STRUCTURE analysis to uncover patterns of population structure, admixture, and introgrogression within domesticated rices, and the authors emphasize the importance of similar research that includes wild rice accessions. 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 [**?**]). This asymmetry is due in part to the closed floret architecture of the domesticated rice, which hinders outcrossing. However, 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 [**?**]. 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.

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

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. 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 completely confounded. The level of divergence of a crop from newly encountered populations and species will decrease due to introgression, a signal that could be mistaken for origin rather than gene flow. For example, when determining a single origin of maize from *parviglumis*, Matsuoka and colleagues [?] 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.* [?] 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 [? ?]. 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.*, [? ?]). 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 [?], may explain discrepancies between domestication dates based on genetic and archaeological data. More accurate estimates of the timing of domestication may be obtained from genetic data by excluding loci that show signatures of introgression.

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

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.

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 [?]. 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 demographic model, 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 [?]. Takuno and co-authors [?] 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 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 [?], likely due to ongoing selection favoring the domesticated phenotype.

Future studies in crop-wild introgression

Research has so far shown that adaptive crop-wild introgression has played a significant role in the domestication histories of many agronomically-important crops. However, the dynamics of the process in these cases are not yet fully understood. To what extent does the level of introgression across taxa depend on divergence time and/or mutation load between donor and recipient taxa? Can colonizing species and/or hybrid swarms serve as bridges for gene flow between previously allopatric taxa? At what geographic scale does adaptive introgression occur? Is introgression frequently restricted to very local populations, or is it often seen over broad geographic ranges? To what extent does this depend on the slope of environmental gradients such as temperature, precipitation, and elevation? How did the conscious, subconscious, and unconscious decisions of early farmers facilitate or hinder adaptive introgression into their crops during early domestication? How do the practices of contemporary farmers affect the process of adaptive introgression today?

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 signal of introgression. These loci are potentially beneficial targets for crop improvement. Furthermore, adaptive introgression that is clearly tied to a specific environment may include beneficial alleles that can be utilized in crop breeding.

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.
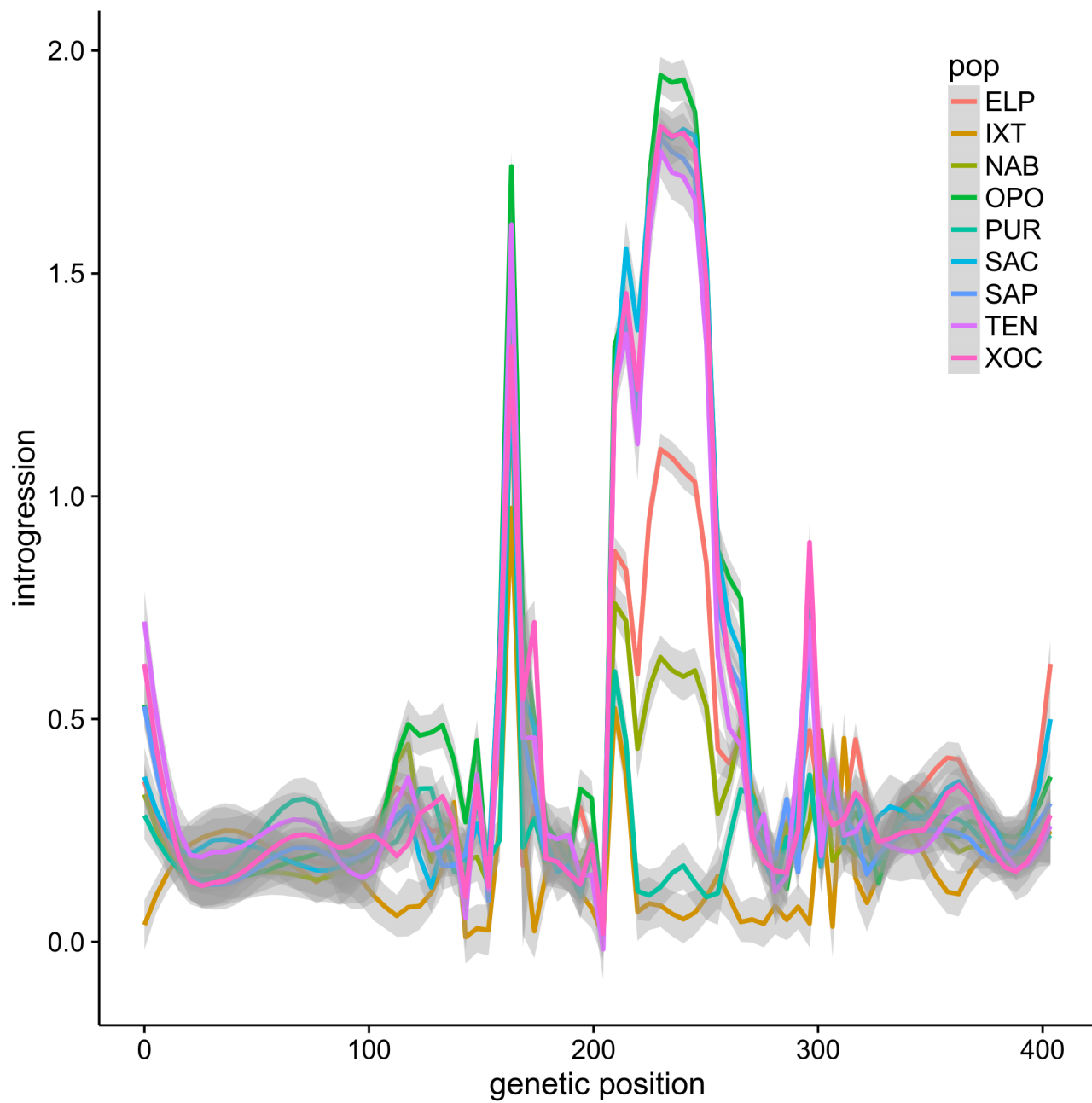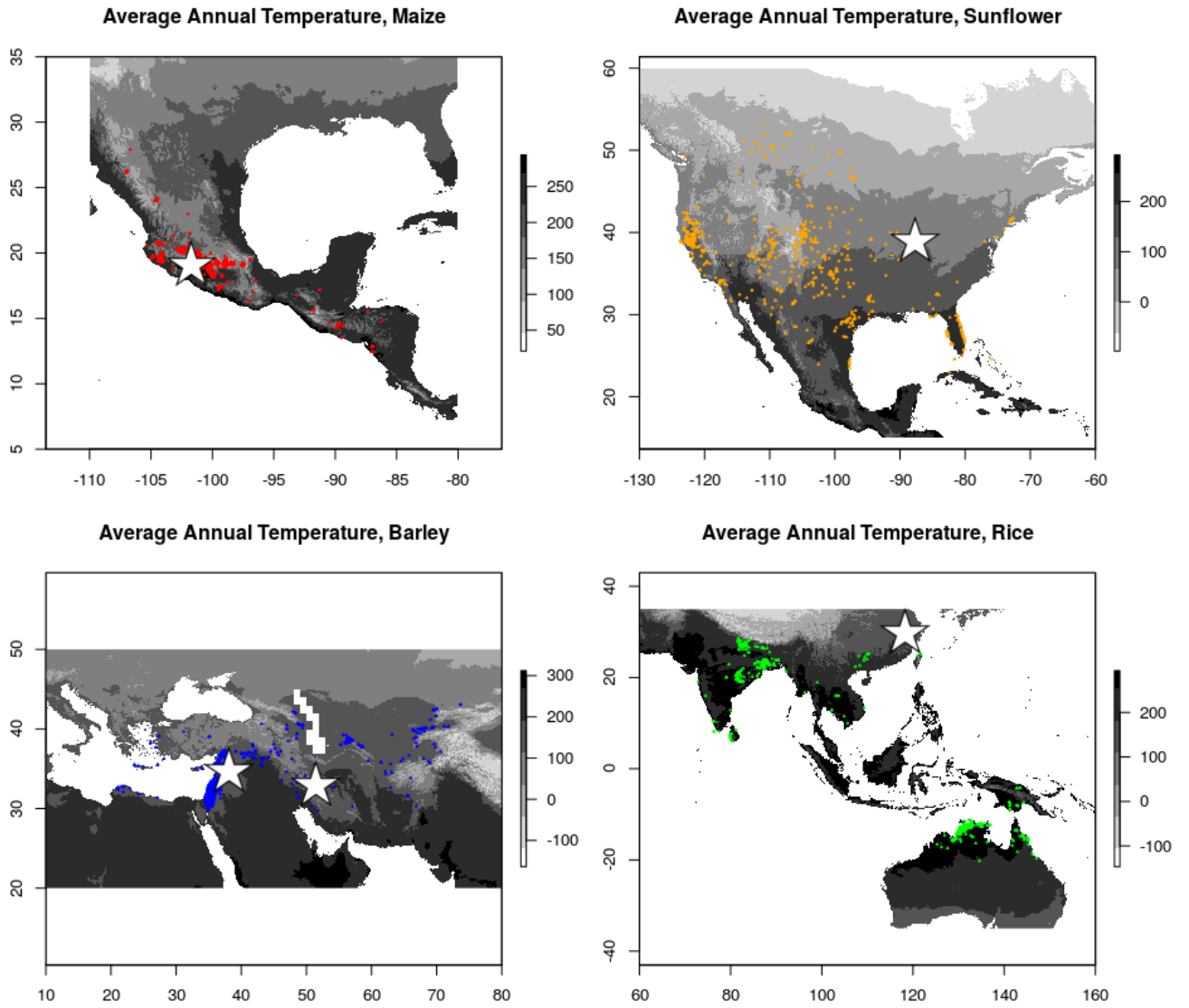
References

Figure 1: Li's caption here.

Figure 2: Map of the natural ranges of wild relatives of four domesticated crops, overlayed with average annual temperature.
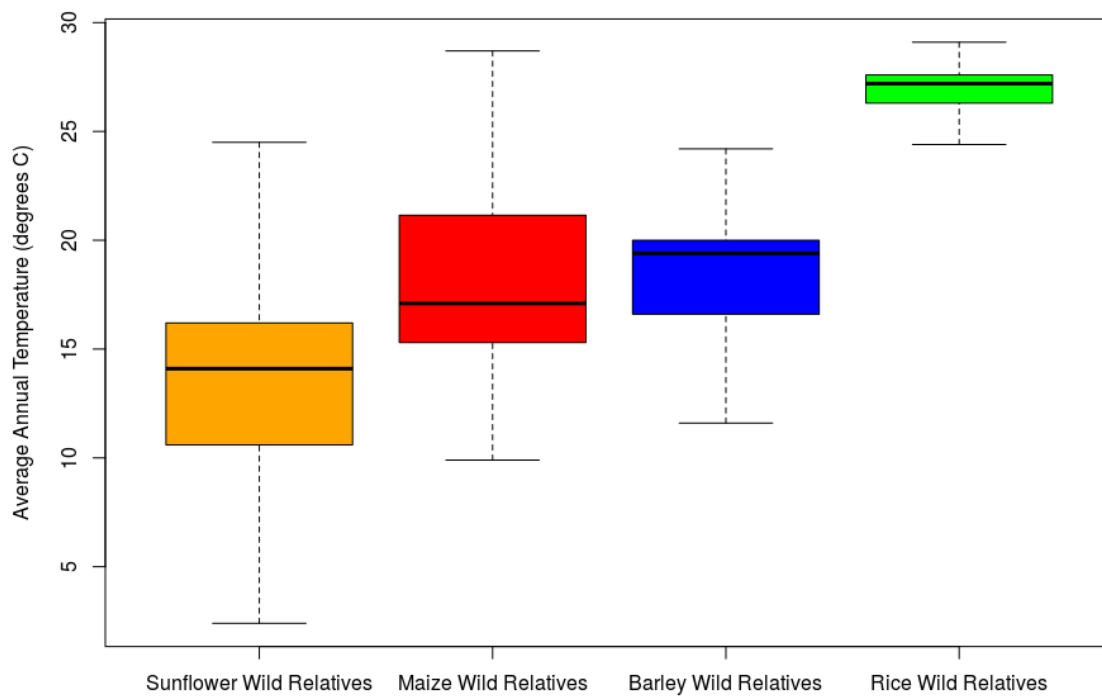
Figure 3: The distribution of average annual temperature experienced in the geographic home ranges of wild relatives interfertile with four crops