Review: The Extent of Adaptive Wild Introgression in Crops

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)

Plant domestication is often conceptualized as a geographically constrained process, with crops originating from a wild progenitor within one or more defined centers followed by expansion to the modern-day range 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]. An additional important aspect of the emerging complexity of domestication is the occurrence of beneficial gene flow from locally adapted wild relatives to crops during their expansion following domestication. It is this adaptive introgression that is the subject of this review.

Adaptive introgression has three components: hybridization between differentiated taxa, backcrossing to one of the parents, and selection on recombinant genotypes with progressively diminished linkage drag [5]. In domesticated species, adaptive introgression would consist of crop-wild hybrids backcrossing to a crop followed by increase in frequency of adaptive wild alleles in the crop and selection against undesirable wild background. To date, the literature on crop-wild gene flow has largely focused on the risk of transgene introgression from domesticated crops into wild relatives (for a review, [6]) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, [7]). The history of natural and potentially adaptive introgression of wild alleles into domesticated crops over evolutionary timescales has received considerably less attention. Recently developed tools and methods have been applied to high-density marker data to detect genome-wide patterns of introgression, granting novel insight into the prevalence of adaptive introgression in crop histories. Results from these studies suggest there is a need to expand our conception of domestication to encompass the broadening of the genetic base of crops that occurred through adaptive gene flow from newly encountered wild relatives during post-domestication expansion.

In this review, we: 1) briefly describe recently developed methods for detecting 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 key questions regarding crop adaptation through gene flow from wild relatives.

Introgression methods and their application

The decreasing cost of genome-wide resequencing and availability of reduced-representation genotyping (*e.g.*, GBS and RAD-Seq), combined with new analytical methods (**Table 1**), has facilitated the comprehensive study of introgression across a broad spectrum of species. The methods reviewed here do not include those estimating introgression/migration rate as a component of broader demographic history (*e.g.*, Approximate Bayesian Computation (ABC) [8], diffusion approximations for demographic inference ($\delta a \delta i$) [9], isolation with migration models [10], and methods utilizing the

sequentially Markovian coalescent (*e.g.*, PSMC) [11]). Rather, we focus on three categories of methods that explicitly identify introgressed genomic segments based on the extent of differentiation, 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 derivates of d_{XY} including G_{min} [12] and RND_{min}[13] gauge differentiation. The former two statistics are insensitive to rare migrants in a population and therefore lack power to detect very 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 [13]. These statistics have been further developed by adding differentiation between both non-admixed (A) and admixed populations (B) and a source population (C) [14]. For example, the $U_{A,B,C(w,x,y)}$ statistic summarizes the 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 frequency in the admixed population (B) [14]. Further modifications have allowed specification of more than one source population (see details in [14]). Since differentiation-based methods can be calculated site-by-site, high-density, genome-wide data are not necessarily required. However, accuracy of introgression estimates is improved with more comprehensive data. Phased data are also not a prerequisite for differentiation-based methods.

Second, ancestry deconvolution (also known as local-ancestry inference and chromosome painting) assigns genomic regions to various source populations based on patterns of allele or haplotype sharing [15]. One form of ancestry deconvolution utilizes a hidden Markov model to evaluate ancestry across admixed genomes through comparison to reference, non-admixed individuals (*e.g.*, HAPMIX [16]). Another clusters admixed populations with reference samples using a sliding-window approach (*e.g.*, PCAdmix [17] and LAMP [18]). And, finally, a third version uses a Bayesian model [19] in which deviations from Hardy-Weinberg equilibrium are minimized through creation of genetic groups (*e.g.*, fineSTRUCTURE [20]). Ancestry deconvolution methods are better suited to high-density marker or whole genome sequencing data given the intent to assign ancestry genome-wide. Many such methods also require accurate phasing of haplotypes.

Phylogenetic relationships are evaluated and applied to introgression detection using the ABBA-BABA statistic (also known as the D-statistic) and related metrics. These statistics make inferences regarding introgression based on genomic patterns of derived variants that are shared between populations or species. While the D-statistic is best suited to detection of introgression at the genome level, elaborations of the D-statistic including $\hat{f}_d$ [21] and D_{FOIL} tests [22] are capable of localizing introgression to specific chromosomal regions. The former is quite similar to the D-statistic but uses allele frequencies from each population/species, and the latter can identify donor and recipient lineages of introgression in a more complex, five-taxon phylogeny. Like differentiation methods, phylogeny-based detection of introgression can be employed using low-density data. However, because these methods require knowledge of whether an allele is ancestral or derived, data from a sufficiently diverged outgroup must also be included in analyses.

Collectively, introgression detection methods have been applied across a wide range of systems, identifying several instances of adaptive introgression. For example, based on sequence divergence methods, introgression has been detected in *Mimulus* (*i.e.*, monkeyflower) species and appears to play a role in both adaptation to pollinator preference and speciation [23]. Likewise, the HAPMIX ancestry deconvolution method was applied by Jeong et al. [24] to detect introgression in human populations from the Nepalese Sherpa to Tibetans at loci controlling high altitude adaptation. Finally, the ABBA-BABA statistic has revealed introgression at wing coloration loci conferring Müllerian mimicry across butterfly species [25]. Below we describe the nascent application of these

methods to crop systems as well as implications for the study of domestication and adaptation.

Table 1: List of recently developed methods for detecting introgression and examples of their use in empirical studies.

Methods	Data Type	References	Applications
Divergence			
Gmin	biallelic SNP	[12]	[26]
RNDmin	phased haplotype	[13]	[27]
$U_{A,B,C(w,x,y)}$ and $Q95_{A,B,C(w,y)}$	biallelic SNP	[28]	[29]
Ancestry Deconvolution			
Hapmix	phased haplotype; reference panel	[16]	[30]
RASPberry	phased haplotype	[31]	[32]
MultiMix	phased/unphased genotype; reference panel	[33]	[34]
PCAdmix	phased haplotype	[17]	[35]
LAMP	phased haplotypes; reference panel	[18]	[36]
Phylogenetic Relationship			
ABBA-BABA/D-statistics	biallelic SNP	[37]	[25]
fd statistic	biallelic SNP	[21]	[38]
five taxon D statistics	biallelic SNP	[22]	[39]

Crop adaptation through introgression

As comprehensive genetic analyses of crops and their wild relatives have become feasible across their geographic ranges, evidence for substantial crop-wild introgression has been discovered in some of the world's most important crops (**Table 2**). Below we present a summary of findings from maize, barley, rice, and potato, four systems in which crop-wild gene flow appears to have played an adaptive role. All four of these crops were domesticated in defined centers and have subsequently expanded to much broader distributions, a migration that brought them into contact with new populations of wild relatives (Figure 1). Below we describe for each case study what is known about a crop's domestication history and the prevalence of adaptive introgression during post-domestication expansion.

1. Maize:

   The relationship between maize (*Zea mays* ssp. *mays*) and the teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*) offers a prime example 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 [40]. 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 [41] and molecular [42, 43] data. However, Hufford et al. [30] 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 ancestry deconvolution methods including HAPMIX (Figure 2). These introgressed segments overlapped with QTL that had previously been found to control anthocyanin content and leaf macrohairs [44], 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, providing further evidence of local adaptation.

Populations of *mexicana* cannot be found outside the highlands of Mexico, yet maize has colonized and adapted to high elevation in a number of other regions. Wang and co-authors [3] employed the ABBA-BABA and $\hat{f}_d$ statistics to evaluate if 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* introgression was pervasive in maize from Mesoamerican high elevation regions (the highlands of Mexico, Guatemala, and the southwestern United States), but that more distant high elevation regions (*e.g.,* the Andes) showed no *mexicana* ancestry. Maize from the lowlands of northern South America appears to have been derived from Mesoamerican lowland maize and high elevation adaptation in Andean maize likely occurred *de novo* from this genetic base [45].

2. Barley:

   Barley (*Hordeum vulgare* subsp. *vulgare*) was likely domesticated multiple times from wild subsp. *spontaneum* roughly 8,000 to 10,000 BP. There is clear evidence of one domestication center in the Fertile Crescent [46, 47], and others have supported additional eastern domestication events, potentially from subsp. *spontaneum* east of the Zagros Mountains [47] or from subsp. *spontaneum* var. *agriocrithon* in modern-day Tibet [48]. However, recent research casts doubt on Tibetan domestication and suggests that var. *agriocrithon* is not a wild relative, but rather a hybrid of domesticated landraces [49]. Presently, the distribution of wild barley stretches from the eastern Mediterranean through the Middle-East to west-central Asia, spanning clines in temperature, precipitation, soil type, and altitude [47]. Cultivated barley is found throughout much of wild barley's distribution and crop-wild hybrids are fertile and common when these taxa co-occur.

   Poets and co-authors [50] recently investigated the range-wide contribution of wild barley to landraces, assessing both genome-wide and geographical patterns of introgression. This study identified several lines of evidence consistent with wild introgression aiding the expansion and adaptation of domesticated barley. The authors utilized ancestry deconvolution methods to identify genomic regions of shared ancestry, which linked particular landraces to numerous wild relative populations. These results suggested landraces may have received wild introgression on a continual basis during post-domestication expansion. However, barley landraces also showed an excess of ancestry from nearby wild relatives, indicating a prevalence of local and potentially adaptive gene flow. Limited admixture linkage disequilibrium and small tracts of identity by state suggest substantial recombination has occurred since initial crop-wild hybridization and that even locally introgressed chromosomal regions are ancient, perhaps dating to the early expansion of barley post-domestication. While these patterns suggest the possibility of adaptive introgression, wild barley haplotypes have yet to be definitively linked to specific adaptations in landraces post-domestication.

3. Potato

   Modern potato (*Solanum tuberosum*) was likely domesticated approximately 6,000-10,000 BP

in southern Peru in sympatry with several wild relatives. The exact progenitor has remained in question for some time [51, 52, 53], but a distance-based phylogeny constructed using genotypic data from a *Solanum* diversity panel recently identified *S. candolleanum* as the most probable progenitor [54]. The lack of clarity regarding a progenitor has been due, in part, to extensive post-domestication hybridization between potato and a number of related species.

While potatoes are primarily propagated clonally, farmers do at times promote sexual reproduction for improvement of the crop and development of new cultivars [55]. Close proximity of domesticated potatoes and wild relatives, active hybridization, and local selection pressure favoring wild haplotypes across a diverse range of biotic and environmental conditions have likely fostered an expansion of genetic diversity within potatoes subsequent to domestication [56]. The prevalence of wild introgression was recently clarified in a broad survey of potato diversity by Hardigan and colleagues [57]. These authors discovered that both diploid and tetraploid domesticates had received extensive introgression from more than ten species of wild *Solanum*, with a continued broadening of the genetic base of potato as it spread away from its Peruvian origin. In certain cultivars, wild ancestry was estimated at upwards of 30%. Genes located within these introgressed regions were more likely to be highly-expressed and stress-inducible, and contained loci related to disease resistance, drought tolerance, and heat tolerance, suggesting introgression conferred adaptations critical to survival, possibly facilitating tolerance for new environmental pressures during range expansion [57].

The four crop systems described in detail here represent particularly compelling examples of putatively adaptive wild introgression. However, given their similar histories, many additional crops have likely benefited from wild-to-crop gene flow during post-domestication expansion (**Table 1**). Across these four cases studies, potential generalities can be observed. Data thus far indicate that wild introgression is often local or regional in its extent, but that, in certain cases (*e.g.*, *mexicana* haplotypes detected in maize landraces from the Guatemalan or southwestern U.S. highlands), newly introgressed wild haplotypes can be disseminated more broadly. Additionally, when functional information is evaluated, as in the maize and potato studies, introgression has been found to occur at loci conferring adaptation to novel conditions not found in a crop's center of origin. Beyond these observations associated with post-domestication adaptation, pervasive wild-to-crop gene flow is also relevant to the study of domestication itself.

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 has been extensive throughout a crop's history. The addition of ongoing gene flow to our understanding of crop demography could therefore bear on fundamental questions of crop domestication:

Where and from what taxa did a crop originate?

Depending on the extent of post-domestication gene flow with new wild relatives, identification of a crop's origin may be complicated or confounded entirely. Introgression between a crop and newly-encountered taxa decreases divergence of the crop from these donors. 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 [40] 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.* [42] 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, as described above, extensive post-domestication adaptive introgression from potato wild relatives long obscured this crop's origin.

Beyond confounding detection of progenitor taxa, extensive introgression may necessitate a more nuanced view of crop origins. In cases like maize and potato it is important to recognize the substantial contributions of introgressing taxa to modern crops. While these crops may have originated from a single species or subspecies, the crops as we know them today have a broader genetic base.

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.*, [40, 58]). In highly introgressed domesticates, these estimates will be based on comparison of both crop and introgressant haplotypes to those of the presumed progenitor. In such cases, divergence time is a mixture of time since domestication and time since split of the progenitor and the introgressing taxa. This phenomenon, in combination with divergence of modern crop samples from true ancestral crop populations, ongoing evolution of crop progenitors, and problems with assuming evolution under a molecular clock [59], may help 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 or by explicitly including estimates of introgression when modeling a crop's demographic history.

How has genome-wide diversity been shaped by domestication?

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 from a wild taxon with a historically higher N_e will lead to underestimates of the overall strength of the initial 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 [60]. Ideally, candidate loci will be identified by first constructing a demographic model representing the history of the domesticate. In this approach, polymorphism 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 uncertainty associated with any given demography, many studies identify domestication loci using a strict outlier approach in which loci showing, for example, 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 [42]. Takuno and co-authors [45] 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 [30], likely due to ongoing selection favoring the domesticated phenotype.

In summary, should post-domestication gene flow with wild relatives be pervasive during crop histories, investigations seeking to unravel fundamental questions of crops' initial domestication must take this into account in order to accurately estimate parameters of interest.

Investigating crop adaptation through introgression:

Research has so far shown that adaptive crop-wild introgression has played a significant role during domestication and dispersal of many important crops. However, the scope and dynamics of this process are not yet fully described and remain unexplored in many systems. In determining the extent and nature of adaptation due to introgression, several questions should be considered:

Do geographic patterns of introgression inform our understanding of adaptation?

Conservation of the genomic architecture of introgression across individuals, between populations, and across landscapes can help illuminate whether introgression is, in fact, adaptive. For example, if an introgressed chromosomal region is conserved across a broad ecogeographic region, this suggests it may impart adaptation to more widespread environmental or climatic variables (*e.g.*, cool temperatures at high elevation). On the other hand, if genetic architectures of introgression are conserved across individuals within a population but not across populations in the region, this suggests much more local selective pressures (*e.g.*, locally prevalent biotic pressures). Highly variable introgression across individuals would be more consistent with random gene flow than adaptation.

Over what timescales and in what genomic regions can we reliably detect adaptive introgression?

Introgressed haplotypes are most easily detected with limited recombination post-hybridization. Therefore, recent introgressions (limited meioses) or those occurring in low recombination regions such as centromeres or inversions are preferentially detected. While this can be problematic for the detection of ancient introgression, the fact that recombination degrades tracts of introgression at a relatively constant and predictable rate allows use of the genome-wide distribution of introgression tract lengths to date initial hybridization (as in [50]). Detection of introgression will also be affected by mutation rate, effective population size, the strength of selection on introgressed alleles, and the extent of divergence between donor taxa and a crop's wild progenitor (*e.g.*, highly divergent introgressed haplotypes will be easier to identify).

At what taxonomic scale does introgression occur?

As species become substantially diverged, introgression can become maladaptive or impossible due to, for example, Dobzhansky-Muller incompatibilities and other pre- and post-zygotic barriers. Divergence time may therefore be a useful predictor of the possibility of gene flow between a particular wild relative and a domesticate. Hybridization may also be limited between a crop and a particular wild relative due to demography. For instance, gene flow from a wild relative with a small long-term effective population size, and correspondingly high genetic load may not be favored by selection. This effect has been observed in the case of Neanderthal introgression into humans, which was likely limited in extent and relegated largely to non-genic regions due to the high genetic load found within Neanderthal donor individuals [61].

Can adaptive introgression inform crop improvement?

Additional study of introgression in agroecosystems could lead to advances in crop improvement. As described above, 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 and their accurate identification is crucial. Furthermore, adaptive introgression that is demonstrably tied to a specific environment represents a promising source of beneficial alleles that can be directly utilized in breeding to adapt crops to similar conditions. Finally, as the historic role of wild relatives in the adaptation of crops is clarified, their conservation may be more prioritized, particularly as a resource for breeding in the face of future climate volatility and change.

Conclusions

As recognized early on by Darwin, crop species represent promising model systems for evolutionary study. The application of genomic tools has generated mounting evidence of crop-wild gene flow during and well beyond the initial stages of domestication. The process of crop range expansion and adaptation through gene flow is closely linked to domestication and offers new questions and opportunities for practical and theoretical investigation. Domesticated crops are among humanity's first and greatest inventions. Interwoven in the genetic history of these crops is the story of our ancestors as they transitioned through the Agronomic Revolution. In a more nuanced and dynamic understanding of crop domestication that includes sustained gene flow with wild relatives, we gain insight into the evolutionary journey that crops and humanity took, and continue to take, together.

Domesticated Crop	Compatible Wild Relatives	Hybrids and/or Hybridization	Evidence of Crop Introgression	Evidence of Adaptativeness	Sources
Apple (*Malus domesticus*)	*M. sylvestris, M. orientalis, M. baccata, M. sieversii*	X	X		[62]
Asian Rice (*Oryza sativa*)	*O. rufipogon*	X	X	X	[63]
Barley (*Hordeum vulgare*)	*H. v.* subsp. *spontaneum*	X	X	X	[50]
Cassava (*Manihot esculenta*)	*M. glaziovii*	X	X	X	[64]
Common Bean (*Phaseolus vulgaris*)	*P. v.* var. *aborigineus, P. v.* var. *mexicanus* [[not in this source]]	X	X		[65]
Grapes (*Vitis vinifera* subsp. *vinifera*)	*V. v.* subsp. *sylvestris*	X	X		[66]
Maize (*Zea mays* subsp. *mays*)	*Z. m.* subsp. *mexicana, Z. m.* subsp. *parviglumis*	X	X	X	[30]
Olive (*Olea europaea* ssp. *europaea* var. *sativa*)	*O. e.* ssp. *europaea* var. *sylvestris*	X	X		[67]
Potato (*Solanum tuberosum*)	many	X	X	X	[57]
Soybeans (*Glycine max*)	*G. soja*	X	X		[68]
Sorghum (*Sorghum bicolor* subsp. *bicolor*)	*S. b.* subsp. *arundinaceum, S. b.* subsp. *drummondii*	X	X		[69]
Sunflower (*Helianthus annuus*)	*H. argophyllus, H. bolanderi, H. debilis, H. petiolaris*	X			[70]
Tomato (*Solanum lycopersicum*)	*S. pimpinellifolium*	X	X	X	[71]
Wheat (*Tritium monococcum, T. dicoccum, T. aestivum*)	*T. m. boeoticum, T. dioccoides, T. urartu, Aegilops speltoides, A. tauschii*	X	X		[72]

References

[1] J. R. Harlan, *Crops & Man.* Madison, WI: American Society of Agronomy, Crop Science Society of America, 1992.

[2] 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.

[3] L. Wang, T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, and M. B. Hufford, "The interplay of demography and selection during maize domestication and expansion," *Genome Biology*, vol. 18, p. 215, Nov 2017.

[4] 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.

[5] N. Barton, "The role of hybridization in evolution," *Molecular Ecology*, vol. 10, no. 3, pp. 551–568, 2001.

[6] 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.

[7] 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.

[8] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.

[9] 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.

[10] 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.

[11] H. Li and R. Durbin, "Inference of human population history from individual whole-genome sequences," *Nature*, vol. 475, no. 7357, pp. 493–496, 2011.

[12] 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.

[13] 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.

[14] 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.

[15] J. G. Schraiber and J. M. Akey, "Methods and models for unravelling human evolutionary history," *Nature Reviews Genetics*, 2015.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] C. Jeong, G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D. B. Witonsky, J. K. Pritchard, C. M. Beall, and A. Di Rienzo, "Admixture facilitates genetic adaptations to high altitude in tibet," *Nature communications*, vol. 5, p. 3281, 2014.

[25] H. G. Consortium *et al.*, "Butterfly genome reveals promiscuous exchange of mimicry adaptations among species," *Nature*, vol. 487, no. 7405, pp. 94–98, 2012.

[26] S. B. Kingan, A. J. Geneva, J. P. Vedanayagam, and D. Garrigan, "Genome divergence and gene flow between drosophila simulans and d. mauritiana," *bioRxiv*, p. 024711, 2015.

[27] F. Roda, F. K. Mendes, M. W. Hahn, and R. Hopkins, "Genomic evidence of gene flow during reinforcement in texas phlox," *Molecular ecology*, vol. 26, no. 8, pp. 2317–2330, 2017.

[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] A. J. Sams, A. Dumaine, Y. Nédélec, V. Yotova, C. Alfieri, J. E. Tanner, P. W. Messer, and L. B. Barreiro, "Adaptively introgressed neandertal haplotype at the oas locus functionally impacts innate immune responses in humans," *Genome Biology*, vol. 17, no. 1, p. 246, 2016.

[30] M. Hufford, P. Lubinksy, T. Pyhäjärvi, M. Devengenzo, N. Ellstrand, and J. Ross-Ibarra, "The genomic signature of crop-wild introgression in maize," *PLoS Genetics*, vol. 9, no. 5, p. e1003477, 2013.

[31] D. Wegmann, D. E. Kessner, K. R. Veeramah, R. A. Mathias, D. L. Nicolae, L. R. Yanek, Y. V. Sun, D. G. Torgerson, N. Rafaels, T. Mosley, *et al.*, "Recombination rates in admixed individuals identified by ancestry-based inference," *Nature genetics*, vol. 43, no. 9, pp. 847–853, 2011.

[32] C. Christe, K. N. Stölting, L. Bresadola, B. Fussi, B. Heinze, D. Wegmann, and C. Lexer, "Selection against recombinant hybrids maintains reproductive isolation in hybridizing populus species despite f1 fertility and recurrent gene flow," *Molecular ecology*, vol. 25, no. 11, pp. 2482–2498, 2016.

[33] C. Churchhouse and J. Marchini, "Multiway admixture deconvolution using phased or unphased ancestral panels," *Genetic epidemiology*, vol. 37, no. 1, pp. 1–12, 2013.

[34] S. Eyheramendy, F. I. Martinez, F. Manevy, C. Vial, and G. M. Repetto, "Genetic structure characterization of chileans reflects historical immigration patterns," *Nature communications*, vol. 6, 2015.

[35] A. Moreno-Estrada, C. R. Gignoux, J. C. Fernández-López, F. Zakharia, M. Sikora, A. V. Contreras, V. Acuña-Alonzo, K. Sandoval, C. Eng, S. Romero-Hidalgo, *et al.*, "The genetics of mexico recapitulates native american substructure and affects biomedical traits," *Science*, vol. 344, no. 6189, pp. 1280–1285, 2014.

[36] N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich, "Ancient admixture in human history," *Genetics*, vol. 192, no. 3, pp. 1065–1093, 2012.

[37] E. Y. Durand, N. Patterson, D. Reich, and M. Slatkin, "Testing for ancient admixture between closely related populations," *Molecular biology and evolution*, vol. 28, no. 8, pp. 2239–2252, 2011.

[38] W. Zhang, K. K. Dasmahapatra, J. Mallet, G. R. Moreira, and M. R. Kronforst, "Genome-wide introgression among distantly related heliconius butterfly species," *Genome biology*, vol. 17, no. 1, p. 1, 2016.

[39] M. C. Fontaine, J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, A. B. Hall, F. Catteruccia, E. Kakani, *et al.*, "Extensive introgression in a malaria vector species complex revealed by phylogenomics," *Science*, vol. 347, no. 6217, p. 1258524, 2015.

[40] 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.

[41] H. Wilkes, "Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize," *Economic Botany*, pp. 254–293, 1977.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] P. L. Morrell and M. T. Clegg, "Genetic evidence for a second domestication of barley (hordeum vulgare) east of the fertile crescent," *Proceedings of the National Academy of Sciences*, vol. 104, no. 9, pp. 3289–3294, 2007.

[48] 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.

[49] M. Pourkheirandish, H. Kanamori, J. Wu, S. Sakuma, F. R. Blattner, and T. Komatsuda, "Elucidation of the origin of agriocrithonbased on domestication genes questions the hypothesis that tibet is one of the centers of barley domestication," *The Plant Journal*, vol. 94, no. 3, pp. 525–534, 2018.

[50] 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.

[51] D. M. Spooner, K. McLean, G. Ramsay, R. Waugh, and G. J. Bryan, "A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 41, pp. 14694–14699, 2005.

[52] B. Pickersgill and C. B. Heiser Jr, "Origins and distribution of plants domesticated in the new world tropics," *Origins of agriculture*, 1977.

[53] J. G. Hawkes, "The evolution of cultivated potatoes and their tuber-bearing wild relatives," *Die Kulturpflanze*, vol. 36, no. 1, pp. 189–208, 1988.

[54] M. A. Hardigan, J. Bamberg, C. R. Buell, and D. S. Douches, "Taxonomy and genetic differentiation among wild and cultivated germplasm of solanum sect. petota," *The Plant Genome*, vol. 8, no. 1, 2015.

[55] C. Quiros, R. Ortega, L. Van Raamsdonk, M. Herrera-Montoya, P. Cisneros, E. Schmidt, and S. Brush, "Increase of potato genetic resources in their center of diversity: the role of natural outcrossing and selection by the andean farmer," *Genetic Resources and Crop Evolution*, vol. 39, no. 2, pp. 107–113, 1992.

[56] S. Brush, R. Kesseli, R. Ortega, P. Cisneros, K. Zimmerer, and C. Quiros, "Potato diversity in the andean center of crop domestication," *Conservation Biology*, vol. 9, no. 5, pp. 1189–1198, 1995.

[57] M. A. Hardigan, F. P. E. Laimbeer, L. Newton, E. Crisovan, J. P. Hamilton, B. Vaillancourt, K. Wiegert-Rininger, J. C. Wood, D. S. Douches, E. M. Farré, *et al.*, "Genome diversity of tuber-bearing solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato," *Proceedings of the National Academy of Sciences*, vol. 114, no. 46, pp. E9999–E10008, 2017.

[58] 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.

[59] 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.

[60] 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.

[61] K. Harris and R. Nielsen, "The genetic cost of neanderthal introgression," *Genetics*, vol. 203, no. 2, pp. 881–891, 2016.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.
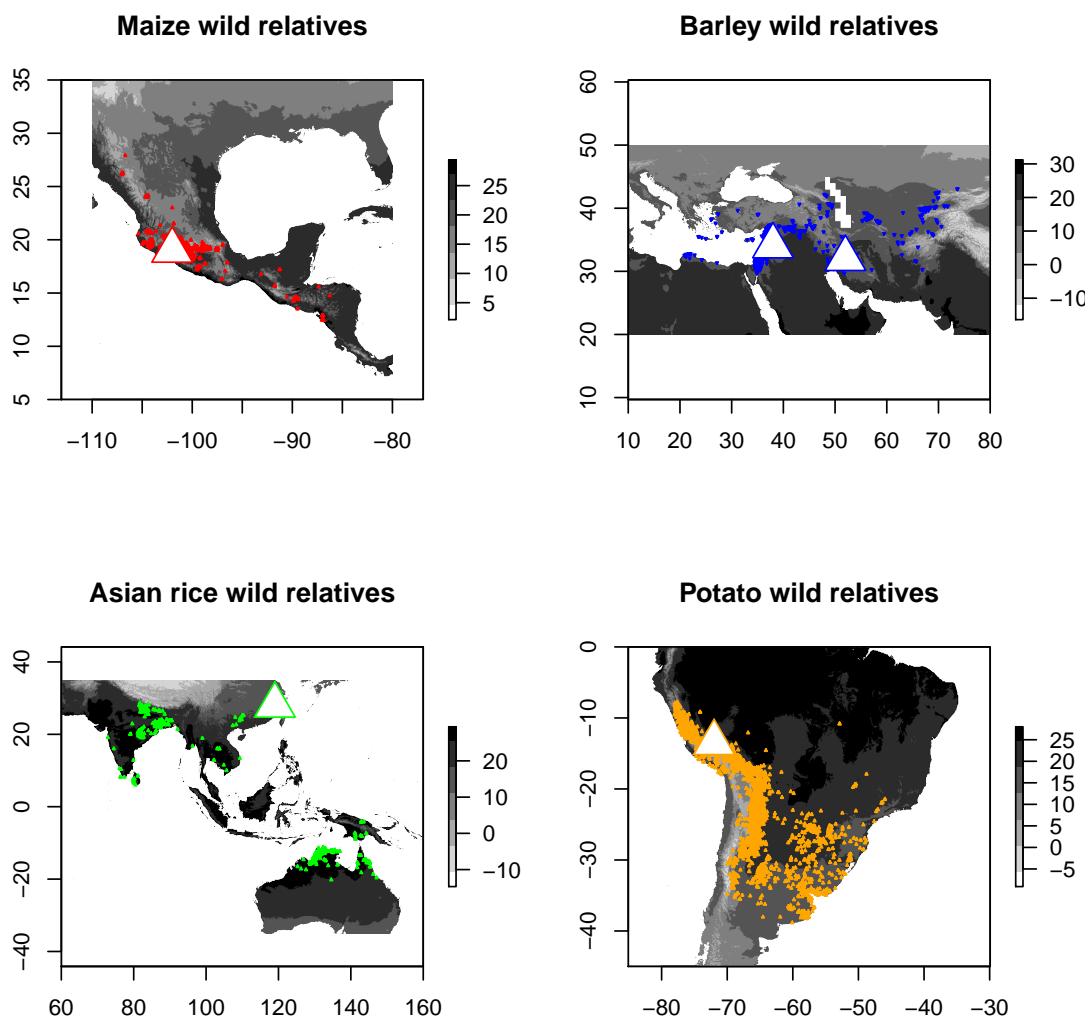
Figure 1: Map of the natural ranges of wild relatives of four domesticated crops, overlayed with average annual temperature. Approximate domestication center for each crop is denoted by a triangle
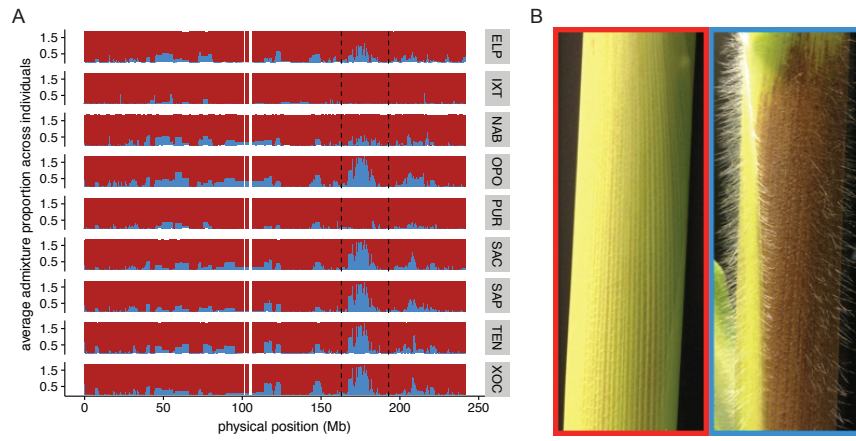
Figure 2: Evidence of adaptive introgression from *mexicana* to Mexican highland maize on chromosome 4. A. Stacked bar plots of the HAPMIX introgression scan. Data adopted from Hufford et al. [30]. ELP: EL Porvenir; IXT: Ixtlan ; NAB: Nabogame; OPO: Opopeo; PUR: Puruandiro; SAC: Santa Clara; SAP: San Pedro; TEN: Tenango del Aire; XOC: Xochimilco. The dashed vertical lines indicate one of the QTLs for macrohairs and pigment density in Lauter et al. [44]. B. Phenotypic differences between lowland (red) and highland (blue) maize stems.

.