**Review: The Prevalence of Adaptive Introgression in Crop Histories**

Authors: Garrett M. Janzen

Li Wang

Matthew 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 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 differentiated taxa, backcrossing to one of the parents, and selection on recombinant genotypes with progressively diminished linkage drag . 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, ) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, ). 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. Emerging results from these studies suggest there is a need to expand our conception of domestication to include a broadening of the genetic base of crops that occurred through adaptive gene flow with newly encountered wild relatives during post-domestication expansion.

In this review, we will: 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 remaining questions regarding crop adaptation through gene flow with wild relatives.

**I. 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, has facilitated comprehensive study of introgression across a number of species (**Table 1**). 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 (*δaδi*) , 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, patterns of nucleotide/haplotype sharing, and phylogenetic relationships.

First, introgressed segments are expected to show low differentiation from their source population. The *Fst* and *dXY* statistics and their derivates including *Gmin* and *RNDmin* 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, *RNDmin* accounts for variable mutation rate, which is detected based on branch length to an outgroup . These statistics have been further developed by adding differentiation between both non-admixed (*A*) and admixed populations (*B*) and a source population (*C*) . For example, the *UA*,*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, *Q*95*A*,*B*,*C*(*w*,*y*), sets a hard cutoff at the 95*th* percentile of allele frequency 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 a hidden Markov model 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 ).

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 $\hat{f\_{d}}$ and the five-taxon D-statistic 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 detects introgression based on the localized phylogenetic pattern and is capable of determining introgression directionality.

Application of these approaches, in combination with population genetic statistics to detect selection, suggest introgression can play an adaptive role. For example, based on sequence divergence methods, introgression has been detected across *Mimulus* (*i.e.*, monkeyflower) species and appears to play a role in both adaptation to pollinator preference and speciation . Likewise, the HAPMIX chromosome painting method was applied by Jeong et al. to detect introgression in human populations from the Nepalese Sherpa to Tibetans at loci controlling high altitude adaptation (reviewed in ). Finally, the ABBA-BABA statistic has revealed introgression at wing coloration loci conferring Müllerian mimicry across butterfly species .

**II. Crop adaptation through introgression**

As range-wide genetic analyses of crops and their wild relatives have become feasible, evidence for substantial wild-to-crop 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 promising systems in which introgression from wild relatives appears 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 . 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 and molecular data. However, Hufford et al. 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 [fig:introgressionMaize]). These introgressed segments overlapped with 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, 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. A recent study 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 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.

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 . 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 . Cultivated barley is found throughout much of the *spontaneum* distribution and barley-*spontaneum* hybrids are fertile and commonly found when these taxa co-occur. In fact, signatures of contemporary gene flow between wild and domesticated barley have been detected across distances greater than a kilometer .

Poets and co-authors 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. Using chromosome painting methods, the authors identified genomic regions of shared ancestry linking 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 suggested even locally introgressed chromosomal regions are ancient, perhaps dating to the early expansion of barley post-domestication. While these results are consistent with adaptive introgression, wild barley haplotypes have yet to be definitively linked to specific adaptations in landraces.

3. Asian Rice:

The details of Asian rice (*Oryza sativa*) domestication are still debated. 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 China and the Indian Ganges plain, respectively , with a potential third domestication event giving rise to *aus* in Bangladesh or central India . Other studies support a single domestication in China, with later divergence of *japonica* and *indica* during crop expansion. For example, Huang and colleagues developed a genetic map of rice variation, which they used to measure genetic distance between a range-wide sample of wild and domesticated rice, finding *japonica* was domesticated near the Pearl River in Guangxi province, China, and that *indica* was likely the result of hybridization between *japonica* and local *rufipogon* populations in Southern and South-eastern Asia. In a re-examination of these same data, found evidence supporting independent domestications of *japonica*, *indica*, and *aus*, as well as a hybrid origin (*japonica* x *aus*) of *aromatic* rice. A recent third analysis by Choi and co-authors compared these two disparate results and the methods utilized therein, concluding that domestication alleles (including *LABA1*, *PROG1*, and *sh1*) arose during a single domestication event of *japonica*, and that these alleles introgressed into several wild *rufipogon* subpopulations which thereby became the progenitors of other Asian rice varieties.

The findings of bear similarities to a hypothesis posited by Vaughan and colleagues . According to this hypothesis, domestication alleles arose in a single cultivated rice population and subsequently introgressed into diverse cultivated populations (some *japonica*-like, some *indica*-like). As these now fully domesticated populations spread further into new environments, they potentially received introgression from locally adapted wild relatives, retaining alleles that improved fitness. While the precise history of domesticated rice remains in question, multiple lines of evidence indicate diverse wild populations have contributed to domesticated germplasm and suggest adaptive introgression may have played a role during the expansion of this important crop.

4. Potato

Modern potato (*Solanum tuberosum*) was likely domesticated in southern Peru in sympatry with several wild relatives approximately 6,000-10,000 BP. The exact progenitor has remained in question for some time , but a distance-based phylogeny constructed using genotypic data from a *Solanum* diversity panel recently identified *S. candolleanum* as the most probable progenitor . 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 . 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 . The prevalence of wild introgression was recently clarified in a broad survey of potato diversity by Hardigan and colleagues . 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 .

**III. 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 is extensive during crop geographic expansion. The addition of ongoing gene flow to our understanding of crop demography could therefore 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 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 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 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 taxon. This phenomenon, in combination with divergence of crop samples from true ancestral crop populations, ongoing evolution of crop progenitors, and problems with assuming evolution under a molecular clock , 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 was genome-wide diversity 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 (*Ne*). 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 *Ne* 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 *Ne* 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 . 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 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, 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 . 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 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 , likely due to ongoing selection favoring the domesticated phenotype.

**IV. Evaluating crop adaptation through introgression:**

Research has so far shown that adaptive crop-wild introgression has played a significant role during domestication and dispersal of multiple important crops. However, the dynamics of this process are not yet fully understood and many questions remain:

**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 general environmental or climatic variables. 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. Highly variable introgression across individuals would be more consistent with random gene flow than adaptation. Finally, introgression at genomic loci showing environmental association and/or a role in disease/pest resistance can be particularly suggestive of adaptive significance.

**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. Moreover, since recombination will progressively degrade tracts of introgression, the genome-wide distribution of tract lengths based on a genetic map can be used to date introgression events (as in ). Mutation rate, effective population size, and strength of selection may also influence the rate of breakup of introgressed regions.

**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 to hybridization. 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 .

**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. Furthermore, adaptive introgression that is demonstrably tied to a specific environment may include beneficial alleles that can be directly utilized in crop breeding. Identification of wild relative species or populations that harbor such adaptive genotypes may lead to their prioritization as part of conservation efforts.

**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. In support of these findings, population genetic theory reinforces the proclivity of native species (in this instance, wild relatives) to provide advantageous, locally adapted alleles to colonizing species (crops) as they disperse beyond their sites of origin into new habitats. The process of crop range expansion and adaptation through gene flow is closely linked to domestication and, as such, offers new questions and opportunities for practical and theoretical investigation. As the historic role of wild relatives in the adaptation of crops is clarified, their conservation may be more prioritized, particularly in the face of future climate volatility and change.

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.