Research 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: email: mhufford@iastate.edu, telephone: 1-515-294-8511

Number of Figures: 2, both in color

Number of Tables: 2

Word Count: 3,966

Summary

The study of crop evolution has focused primarily on the process of initial domestication. Post-domestication adaptation during the expansion of crops from their centers of origin has received considerably less attention. Recent research has revealed that, in at least some instances, crops have received introgression from their wild relatives that has facilitated adaptation to novel conditions encountered during expansion. Such adaptive introgression could bear importantly on the basic study of domestication, affecting estimates of several evolutionary processes of interest (*e.g.*, the strength of the domestication bottleneck, the timing of domestication, the targets of selection during domestication). Identification of haplotypes introgressed from the wild may also aid the identification of alleles that are beneficial under particular environmental conditions. Here we review mounting evidence for substantial adaptive wild introgression in several crops and consider the implications of such gene flow to our understanding of crop histories.

Key Words: adaptation, domestication, gene flow, introgression, wild relatives

Introduction

Plant domestication is often conceptualized as a geographically constrained process, with crops originating from a wild progenitor within defined centers followed by expansion to the modern-day range of cultivation (Harlan, 1992). However, archaeological and genetic evidence are beginning to reveal that, in many cases, domestication has been temporally protracted and geographically diffuse (Fuller *et al.*, 2014; Meyer *et al.*, 2016; Wang *et al.*, 2017). An additional 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, back-crossing to one of the parents, and selection on recombinant genotypes with progressively diminished linkage drag (Barton, 2001). 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, Stewart *et al.* 2003) and on modern plant breeding efforts to introgress desirable traits from wild relatives (for a review, Dempewolf *et al.* 2017). The potential for adaptive introgression of wild alleles into domesticated crops over evolutionary timescales has received considerably less attention. Recently developed 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 recent 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 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; Beaumont *et al.* 2002), diffusion approximations for demographic inference ($\delta a \delta i$; Gutenkunst *et al.* 2009), isolation with migration models (Hey & Nielsen, 2004), and methods utilizing the sequentially Markovian coalescent (*e.g.*, PSMC; Li & Durbin 2011). 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} (Geneva *et al.*, 2015) and RND_{min} (Rosenzweig *et al.*, 2016) 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. These statistics have been further developed by adding differentiation between both non-admixed (A) and admixed populations (B) and a source population (C) (Racimo *et al.*, 2016). 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 95^{th} percentile of allele frequency in the admixed population (B) (Racimo *et al.*, 2016). Further modifications have allowed specification of more than one source population (see details in Racimo *et al.* 2016). 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 source populations based on patterns of allele or haplotype sharing (Schraiber & Akey, 2015). 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 Price *et al.* 2009). Another clusters admixed populations with reference samples using a sliding-window approach (*e.g.*, PCAdmix, Brisbin *et al.* 2012 and LAMP, Sankararaman *et al.* 2008). A third version uses a Bayesian model (Pritchard *et al.*, 2000) in which deviations from Hardy-Weinberg equilibrium are minimized through creation of genetic groups (*e.g.*, fineSTRUCTURE, Lawson *et al.* 2012). Ancestry deconvolution methods are better suited to high-density marker data given the intent to assign ancestry genome-wide. Many such methods also require accurate phasing of haplotypes.

Phylogenetic relationships are applied to introgression detection using the ABBA-BABA statistic (also known as the D-statistic) and related metrics (Durand *et al.*, 2011). 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 $\hat{f}_d$ (Martin *et al.*, 2015) and D_{FOIL} tests (Pease & Hahn, 2015) are capable of localizing introgression to specific chromosomal regions. The former is quite similar to the D-statistic but is based on allele frequencies, 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 be available.

Collectively, introgression detection methods have been applied to several systems, frequently identifying instances of adaptive introgression (see applications in **Table 1**). For example, based on sequence divergence methods, introgression has been detected in *Mimulus* (*i.e.*, monkeyflower) species and appears to play a role in adaptation to pollinator preference and speciation (Stankowski & Streisfeld, 2015). Likewise, the HAPMIX ancestry deconvolution method was applied by Jeong *et al.* (2014) to detect introgression 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 (Consortium *et al.*, 2012). Below we describe the nascent application of these methods to crop systems as well as implications for the study of domestication and adaptation.

Crop adaptation through introgression

As range-wide genetic analyses of crops and their wild relatives have become feasible, evidence for substantial crop-wild introgression has been discovered in many 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 global distributions, a migration that brought them into contact with new populations of wild relatives that are distributed across broad environmental gradients (**Figure 1**). For each case study we describe both what is known about a crop's domestication history and the prevalence of adaptive introgression during 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 (Matsuoka *et al.*, 2002). 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 has been reported based on both morphological (Wilkes, 1977) and molecular (Doebley *et al.*, 1987; van Heerwaarden *et al.*, 2011) data. However, Hufford *et al.* (2013) 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 shown to control anthocyanin content and leaf macrohairs

(Lauter *et al.*, 2004), traits known to be adaptive at high elevation. In a growth chamber experiment, the authors demonstrated that maize populations with *mexicana* introgression had increased height (a proxy for fitness) under highland environmental conditions. 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 *et al.* (2017) 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. These findings are consistent with previous work suggesting high elevation adaptation in Andean maize likely occurred *de novo* (Takuno *et al.*, 2015).

2. Barley:

Barley (*Hordeum vulgare* ssp. *vulgare*) was likely domesticated multiple times from wild ssp. *spontaneum* roughly 8,000 to 10,000 BP. There is clear evidence of one domestication center in the Fertile Crescent (Badr *et al.*, 2000; Morrell & Clegg, 2007), and others have supported additional eastern domestication events, potentially from ssp. *spontaneum* east of the Zagros Mountains (Morrell & Clegg, 2007) or from ssp. *spontaneum* var. *agriocrithon* in modern-day Tibet (Dai *et al.*, 2012). 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 (Pourkheirandish *et al.*, 2018). Presently, the distribution of wild barley stretches from the eastern Mediterranean to west-central Asia, spanning clines in temperature, precipitation, soil type, and altitude (Morrell & Clegg, 2007). Cultivated barley is found throughout wild barley's distribution and crop-wild hybrids are fertile and common when these taxa co-occur.

Poets *et al.* (2015) 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 suggest 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.

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 (Fuller *et al.*, 2010), with a potential third domestication event giving rise to the varietal group *aus* in Bangladesh or central India (Civáň *et al.*, 2015). Other studies support a single domestication in China, with later divergence of *japonica* and *indica* (Huang *et al.*, 2012; Molina *et al.*, 2011) during crop expansion. For example, Huang *et al.* (2012) measured genetic distance between a range-wide sample of wild and domesticated rice, finding that *japonica* was likely domesticated near the Pearl River in Guangxi province, China, and that *indica* potentially resulted from hybridization between *japonica* and local *rufipogon* populations in southern and south-eastern Asia. In a re-examination of these same data, Civáň and colleagues (2015) 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 & Purugganan (2018) compared these two disparate results and concluded that domestication alleles (including *LABA1*, *PROG1*, and *sh1*) arose during a single domestication event of *japonica*, and were introgressed into several wild *rufipogon* subpopulations (crop-to-wild gene flow), which thereby became the progenitors of other Asian rice varietals.

The findings of Choi & Purugganan (2018) bear similarities to a hypothesis posited by

Vaughan and colleagues (2008) that stresses the potential adaptive significance of crop-wild gene flow in rice. 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 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 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 (Hawkes, 1988; Pickersgill & Heiser Jr, 1977; Spooner *et al.*, 2005), but a distance-based phylogeny constructed using genotypic data from a *Solanum* diversity panel recently identified *S. candolleanum* as the most probable progenitor (Hardigan *et al.*, 2015). 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 (Quiros *et al.*, 1992). 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 (Brush *et al.*, 1995). The prevalence of wild introgression was recently clarified in a broad survey of potato diversity by Hardigan and colleagues (2017). These authors discovered that tetraploid domesticates in particular had received extensive introgression from wild *Solanum*, documenting 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 (Hardigan *et al.*, 2017).

The four crop systems described in detail here represent particularly compelling examples of wild and potentially adaptive 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, some generalities can be observed. Data thus far indicate that wild introgression is often 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. The wild gene flow identified in these case studies raises a number of questions regarding both domestication and adaptation of crops.

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 (2002) 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.* (2011) 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.*, (Matsuoka *et al.*, 2002; Molina *et al.*, 2011)). 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 genetic structure among wild relative populations, 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 (Ross-Ibarra *et al.*, 2007). 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 (van Heerwaarden *et al.*, 2011). Takuno and co-authors (2015) 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 where *mexicana* introgression is most prevalent. 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 wild introgression (Hufford *et al.*, 2013), likely due to ongoing selection favoring the domesticated phenotype.

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

Investigating crop adaptation through introgression:

While research in a subset of crops suggests adaptive wild introgression likely occurred, the scope and dynamics of this process remain poorly described or 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 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 Poets *et al.* 2015). 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).

From which wild taxa will introgression occur?

As species become substantially diverged gene flow is limited due to 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 taxa. Hybridization may also be limited between a crop and a particular wild relative due to genetic load. 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 and relegated primarily to non-genic regions due to the high genetic load found within Neanderthal donor individuals (Harris & Nielsen, 2016).

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. 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

Recent innovations in both high-density marker data and methods for characterizing genome-wide patterns of introgression have helped reveal the extent and timing of gene flow in a number of species. Application of these data and techniques has led to mounting evidence of crop-wild gene flow following initial domestication in several species. Substantial post-domestication gene flow with wild relatives can affect inferences regarding domestication and may be an important mechanism through which crops adapted to novel conditions during global expansion. An accurate understanding of the extent of gene flow is therefore important to both the basic study of crop evolution and to the identification of adaptive alleles for continued crop improvement. While some studies in crop systems have identified wild introgression, even fewer have effectively linked introgressed alleles to adaptation. More comprehensive functional analyses and field evaluation will be critical for understanding the evolutionary and adaptive significance of introgression in crops.

Acknowledgements

The authors thank Daniel Gates, Peter Morrell, Jeffrey Ross-Ibarra and Jonathan Wendel for comments on a previous version of this manuscript. This work was supported by funding from the National Science Foundation Plant Genome Research Program (IOS-1546719).

References

Aldrich P, Doebley J, Schertz K , Stec A. 1992. Patterns of allozyme variation in cultivated and wild sorghum bicolor. *Theoretical and Applied Genetics*, **85**: 451–460.

Badr A, Sch R, El Rabey H, Effgen S, Ibrahim H, Pozzi C, Rohde W, Salamini F *et al.* 2000. On the origin and domestication history of barley (hordeum vulgare). *Molecular Biology and Evolution*, **17**: 499–510.

Barton N. 2001. The role of hybridization in evolution. *Molecular Ecology*, **10**: 551–568.

Beaumont MA, Zhang W , Balding DJ. 2002. Approximate bayesian computation in population genetics. *Genetics*, **162**: 2025–2035.

Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C *et al.* 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnology*, **34**: 562–570.

Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey JG , Bustamante CD. 2012. Pcadmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Human Biology*, **84**: 343.

Brush S, Kesseli R, Ortega R, Cisneros P, Zimmerer K , Quiros C. 1995. Potato diversity in the andean center of crop domestication. *Conservation Biology*, **9**: 1189–1198.

Choi JY , Purugganan MD. 2018. Multiple origin but single domestication led to oryza sativa. *G3: Genes, Genomes, Genetics*, pp. g3–300334.

Christe C, Stölting KN, Bresadola L, Fussi B, Heinze B, Wegmann D , Lexer C. 2016. Selection against recombinant hybrids maintains reproductive isolation in hybridizing populus species despite f1 fertility and recurrent gene flow. *Molecular Ecology*, **25**: 2482–2498.

Churchhouse C , Marchini J. 2013. Multiway admixture deconvolution using phased or unphased ancestral panels. *Genetic Epidemiology*, **37**: 1–12.

Civáň P, Craig H, Cox CJ , Brown TA. 2015. Three geographically separate domestications of asian rice. *Nature Plants*, **1**: 15164.

Consortium HG *et al.* 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**: 94–98.

Dai F, Nevo E, Wu D, Comadran J, Zhou M, Qiu L, Chen Z, Beiles A, Chen G , Zhang G. 2012. Tibet is one of the centers of domestication of cultivated barley. *Proceedings of the National Academy of Sciences*, **109**: 16969–16973.

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C , Guarino L. 2017. Past and future use of wild relatives in crop breeding. **57**: 1070–1082.

Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P , Gaut BS. 2015. Olive domestication and diversification in the mediterranean basin. *New Phytologist*, **206**: 436–447.

Doebley J, Goodman MM , Stuber CW. 1987. Patterns of isozyme variation between maize and mexican annual teosinte. *Economic Botany*, **41**: 234–246.

Durand EY, Patterson N, Reich D , Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**: 2239–2252.

Dvorak J, Akhunov ED, Akhunov AR, Deal KR , Luo MC. 2006. Molecular characterization of a diagnostic dna marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution*, **23**: 1386–1396.

Eyheramendy S, Martinez FI, Manevy F, Vial C , Repetto GM. 2015. Genetic structure characterization of chileans reflects historical immigration patterns. *Nature Communications*, **6**.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E *et al.* 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**: 1258524.

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG , Purugganan MD. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences*, **111**: 6147–6152.

Fuller DQ, Sato YI, Castillo C, Qin L, Weisskopf AR, Kingwell-Banham EJ, Song J, Ahn SM , Van Etten J. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences*, **2**: 115–131.

Geneva AJ, Muirhead CA, Kingan SB , Garrigan D. 2015. A new method to scan genomes for introgression in a secondary contact model. *PLoS ONE*, **10**: e0118621.

Gutenkunst RN, Hernandez RD, Williamson SH , Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS Genetics*, **5**: e1000695.

Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, Zhao L, Zhou G, Wang Z, Huang L *et al.* 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist*, **209**: 871–884.

Hardigan MA, Bamberg J, Buell CR , Douches DS. 2015. Taxonomy and genetic differentiation among wild and cultivated germplasm of solanum sect. petota. *The Plant Genome*, **8**.

Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farré EM *et al.* 2017. 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*, **114**: E9999–E10008.

Harlan JR. 1992. *Crops & Man.* American Society of Agronomy, Crop Science Society of America, Madison, WI.

Harris K , Nielsen R. 2016. The genetic cost of neanderthal introgression. *Genetics*, **203**: 881–891.

Hawkes JG. 1988. The evolution of cultivated potatoes and their tuber-bearing wild relatives. *Die Kulturpflanze*, **36**: 189–208.

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Gonzalez JdJS , Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences*, **108**: 1088–1092.

Hey J , Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of drosophila pseudoobscura and d. persimilis. *Genetics*, **167**: 747–760.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**: 497–501.

Hufford M, Lubinksy P, Pyhäjärvi T, Devengenzo M, Ellstrand N , Ross-Ibarra J. 2013. The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, **9**: e1003477.

Jeong C, Alkorta-Aranburu G, Basnyat B, Neupane M, Witonsky DB, Pritchard JK, Beall CM , Di Rienzo A. 2014. Admixture facilitates genetic adaptations to high altitude in tibet. *Nature Communications*, **5**: 3281.

Kingan SB, Geneva AJ, Vedanayagam JP , Garrigan D. 2015. Genome divergence and gene flow between drosophila simulans and d. mauritiana. *bioRxiv*, p. 024711.

Lauter N, Gustus C, Westerbergh A , Doebley J. 2004. The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics*, **167**: 1949–1959.

Lawson DJ, Hellenthal G, Myers S , Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics*, **8**: 1–16.

Li H , Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**: 493–496.

Ma B, Liao L, Peng Q, Fang T, Zhou H, Korban SS , Han Y. 2017. Reduced representation genome sequencing reveals patterns of genetic diversity and selection in apple. *Journal of Integrative Plant Biology*, **59**: 190–204.

Martin SH, Davey JW , Jiggins CD. 2015. Evaluating the use of abba–baba statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**: 244–257.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E , Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, **99**: 6080–6084.

Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, Dorph K, Barretto A, Gross B, Fuller DQ *et al.* 2016. Domestication history and geographical adaptation inferred from a snp map of african rice. *Nature Genetics*, **48**: 1083–1088.

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD *et al.* 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences*, **108**: 8351–8356.

Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, Acuña-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S *et al.* 2014. The genetics of mexico recapitulates native american substructure and affects biomedical traits. *Science*, **344**: 1280–1285.

Morrell PL , Clegg MT. 2007. Genetic evidence for a second domestication of barley (hordeum vulgare) east of the fertile crescent. *Proceedings of the National Academy of Sciences*, **104**: 3289–3294.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D *et al.* 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, **108**: 3530–3535.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T , Reich D. 2012. Ancient admixture in human history. *Genetics*, **192**: 1065–1093.

Pease JB , Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, **64**: 651–662.

Pickersgill B , Heiser Jr CB. 1977. Origins and distribution of plants domesticated in the new world tropics. *Origins of Agriculture*.

Poets AM, Fang Z, Clegg MT , Morrell PL. 2015. Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology*, **16**: 1–11.

Pourkheirandish M, Kanamori H, Wu J, Sakuma S, Blattner FR , Komatsuda T. 2018. 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*, **94**: 525–534.

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D , Myers S. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, **5**: 1–18.

Pritchard JK, Stephens M , Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945–959.

Quiros C, Ortega R, Van Raamsdonk L, Herrera-Montoya M, Cisneros P, Schmidt E , Brush S. 1992. 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*, **39**: 107–113.

Racimo F, Marnetto D , Huerta-Sánchez E. 2016. Signatures of archaic adaptive introgression in present-day human populations. *Molecular Biology and Evolution*, p. msw216.

Racimo F, Sankararaman S, Nielsen R , Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, **16**: 359–371.

Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutierrez S, Ordaz-Ortiz JJ, Aguilar OM, Vianello-Brondani RP, Santalla M, Delaye L *et al.* 2017. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biology*, **18**: 60.

Rick CM. 1958. The role of natural hybridization in the derivation of cultivated tomatoes of western south america. *Economic Botany*, **12**: 346–367.

Rieseberg LH, Kim SC, Randell RA, Whitney KD, Gross BL, Lexer C , Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129**: 149–165.

Roda F, Mendes FK, Hahn MW , Hopkins R. 2017. Genomic evidence of gene flow during reinforcement in texas phlox. *Molecular Ecology*, **26**: 2317–2330.

Rosenzweig BK, Pease JB, Besansky NJ , Hahn MW. 2016. Powerful methods for detecting introgressed regions from population genomic data. *Molecular Ecology*.

Ross-Ibarra J, Morrell PL , Gaut BS. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences*, **104**: 8641–8648.

Sams AJ, Dumaine A, Nédélec Y, Yotova V, Alfieri C, Tanner JE, Messer PW , Barreiro LB. 2016. Adaptively introgressed neandertal haplotype at the oas locus functionally impacts innate immune responses in humans. *Genome Biology*, **17**: 246.

Sankararaman S, Sridhar S, Kimmel G , Halperin E. 2008. Estimating local ancestry in admixed populations. *The American Journal of Human Genetics*, **82**: 290–303.

Schraiber JG , Akey JM. 2015. Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*.

Spooner DM, McLean K, Ramsay G, Waugh R , Bryan GJ. 2005. 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*, **102**: 14694–14699.

Stankowski S , Streisfeld MA. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**.

Stewart CN, Halfhill MD , Warwick SI. 2003. Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics*, **4**: 806–817.

Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB , Ross-Ibarra J. 2015. Independent molecular basis of convergent highland adaptation in maize. *Genetics*.

Vaughan DA, Lu BR , Tomooka N. 2008. The evolving story of rice evolution. *Plant Science*, **174**: 394–408.

Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J , Hufford MB. 2017. The interplay of demography and selection during maize domestication and expansion. *Genome Biology*, **18**: 215.

Wegmann D, Kessner DE, Veeramah KR, Mathias RA, Nicolae DL, Yanek LR, Sun YV, Torgerson DG, Rafaels N, Mosley T *et al.* 2011. Recombination rates in admixed individuals identified by ancestry-based inference. *Nature Genetics*, **43**: 847–853.

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

Zhang W, Dasmahapatra KK, Mallet J, Moreira GR , Kronforst MR. 2016. Genome-wide introgression among distantly related heliconius butterfly species. *Genome Biology*, **17**: 1.

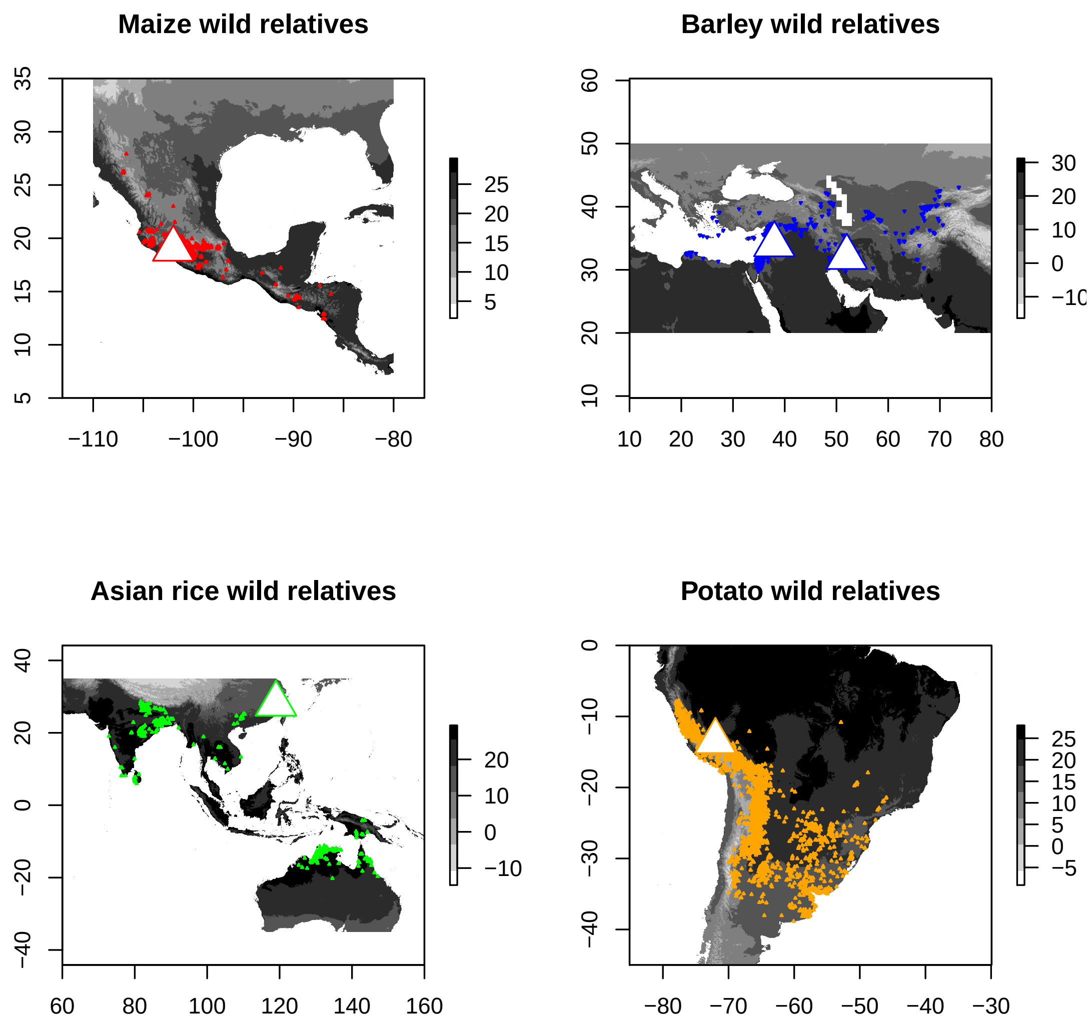

Figure 1: Map of the natural ranges of wild relatives of four domesticated crops overlayed on average annual temperature. Approximate domestication center for each crop is denoted by a triangle

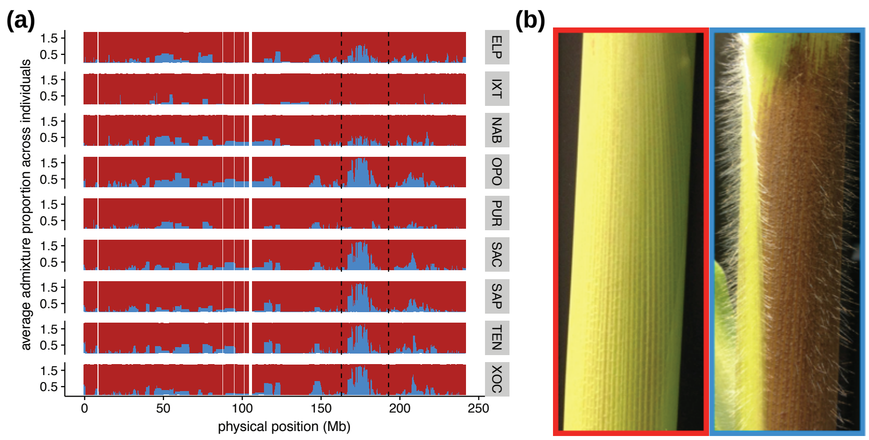

Figure 2: Evidence of adaptive introgression from *mexicana* to Mexican highland maize on chromosome 4. (a) Stacked bar plots of a HAPMIX introgression scan. For each population and chromosomal position, red indicates maize ancestry and blue indicates *mexicana* ancestry. Data obtained from Hufford *et al.* (2013). 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 a previously identified QTL for macrohairs and pigment density in Lauter *et al.* (2004). (b) Phenotypic differences between maize stems with (blue) and without (red) *mexicana* introgression in QTL affecting presence of pigment and macrohairs.

.

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	Geneva *et al.* (2015)	Kingan *et al.* (2015)
RNDmin	phased haplotype	Rosenzweig *et al.* (2016)	Roda *et al.* (2017)
$U_{A,B,C(w,x,y)}$ and $Q95_{A,B,C(w,y)}$	biallelic SNP	Racimo *et al.* (2015)	Sams *et al.* (2016)
Ancestry Deconvolution			
Hapmix	phased haplotype; reference panel	Price *et al.* (2009)	Hufford *et al.* (2013)
RASPberry	phased haplotype	Wegmann *et al.* (2011)	Christe *et al.* (2016)
MultiMix	phased/unphased genotype; reference panel	Churchhouse & Marchini (2013)	Eyheramendy *et al.* (2015)
PCAdmix	phased haplotype	Brisbin *et al.* (2012)	Moreno-Estrada *et al.* (2014)
LAMP	phased haplotypes; reference panel	Sankararaman *et al.* (2008)	Patterson *et al.* (2012)
Phylogenetic Relationship			
ABBA-BABA/D-statistics	biallelic SNP	Durand *et al.* (2011)	Consortium *et al.* (2012)
fd statistic	biallelic SNP	Martin *et al.* (2015)	Zhang *et al.* (2016)
five taxon D statistics	biallelic SNP	Pease & Hahn (2015)	Fontaine *et al.* (2015)

Table 2: Extent of evidence for adaptive introgression for major crops including whether hybrids are observed, introgression is detected, and introgression has been shown to be adaptive.

Domesticated Crop	**Compatible Wild Relatives**	**Hybrids**	**Evidence of Introgression**	**Evidence of Adaptation**	**Sources**
Apple (*Malus domesticus*)	*M. sylvestris*, *M. orientalis*, *M. baccata*, *M. sieversii*	X	X		Ma *et al.* (2017)
Asian Rice (*Oryza sativa*)	*O. rufipogon*	X	X	X	Huang *et al.* (2012)
Barley (*Hordeum vulgare*)	*H. v.* subsp. *spontaneum*	X	X	X	Poets *et al.* (2015)
Cassava (*Manihot esculenta*)	*M. glaziovii*	X	X	X	Bredeson *et al.* (2016)
Common Bean (*Phaseolus vulgaris*)	*P. v.* var. *aborigineus*, *P. v.* var. *mexicanus*	X	X		Rendón-Anaya *et al.* (2017)
Grapes (*Vitis vinifera* subsp. *vinifera*)	*V. v.* subsp. *sylvestris*	X	X		Myles *et al.* (2011)
Maize (*Zea mays* subsp. *mays*)	*Z. m.* subsp. *mexicana*, *Z. m.* subsp. *parviglumis*	X	X	X	Hufford *et al.* (2013)
Olive (*Olea europaea* ssp. *europaea* var. *sativa*)	*O. e.* ssp. *europaea* var. *sylvestris*	X	X		Diez *et al.* (2015)
Potato (*Solanum tuberosum*)	many	X	X	X	Hardigan *et al.* (2017)
Soybeans (*Glycine max*)	*G. soja*	X	X		Han *et al.* (2016)
Sorghum (*Sorghum bicolor* subsp. *bicolor*)	*S. b.* subsp. *arundinaceum*, *S. b.* subsp. drummondii	X	X		Aldrich *et al.* (1992)
Sunflower (*Helianthus annuus*)	*H. argophyllus*, *H. bolanderi*, *H. debilis*, *H. petiolaris*	X			Rieseberg *et al.* (2007)
Tomato (*Solanum lycopersicum*)	*S. pimpinellifolium*	X	X	X	Rick (1958)
Wheat (*Tritium monococcum, T. dicoccum, T. aestivum*)	*T. m. boeoticum, T. dioccoides, T. urartu, Aegilops speltoides, A. tauschii*	X	X		Dvorak *et al.* (2006)