

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

Authors: Garrett M. Janzen<sup>1</sup>, Li Wang<sup>1</sup>, and Matthew B. Hufford<sup>1,\*</sup>

<sup>1</sup>Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

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

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

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

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

## Introgression methods and their application

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

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

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

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

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

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

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

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

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

Application of these approaches across a number of plant and animal species suggests introgression can play an adaptive role. For example, introgression from ancient hominins (*e.g.*, Neanderthals and Denisovans) to humans has been detected at loci controlling skin pigmentation, defense against pathogens, and toleration of high altitude (reviewed in [28]); introgression has conferred Müllerian mimicry *I would explain a bit more here...wing coloration loci, protects against predation...* across butterfly species [29]; introgression has spread insecticide resistance across mosquito species [30], and introgression across *Mimulus* (*i.e.*, monkeyflower) species has resulted in adaptation to pollinator preference and contributed to speciation [31].

## Crop adaptation through introgression

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

### 1. Maize:

The relationship between maize (*Zea mays* ssp. *mays*) and the teosinte *Zea mays* ssp. *mexicana* (hereafter referred to as *mexicana*) offers a prime case study of adaptive wild-to-crop introgression. Maize was domesticated from *Zea mays* ssp. *parviglumis* (hereafter referred to as *parviglumis*) approximately 9,000 years ago in the lowlands of the Balsas River Valley in Mexico [32]. From this domestication center, maize spread into the highlands of the Mexican Central Plateau, where it came into sympatry with *mexicana*. Introgression from *mexicana* to maize in the Central Plateau has long been reported based on both morphological [33, 34, 35] and molecular [32, 36, 37, 38, 39] data. However, [40] first localized *mexicana* introgression to chromosomal regions and provided evidence that it was likely adaptive. The authors identified nine genomic regions in several maize populations which consistently showed evidence of *mexicana* introgression based on chromosome painting using both HAPMIX and the linkage model of STRUCTURE (Figure 1). These introgressed segments showed low diversity and overlapped QTL that had previously been found to control anthocyanin content and leaf macrohairs [34], traits known to be adaptive at high elevation. In a growth chamber experiment, the authors demonstrated that maize populations with *mexicana* introgression showed greater plant height (a proxy for fitness) under highland environmental conditions than populations that lacked introgression. Height differences were not detected under lowland conditions.

Populations of *mexicana* cannot be found outside of Mexico, yet maize has colonized and adapted to high elevation in a number of other regions. A recent study [4] employed the ABBA-BABA and  $\hat{f}_d$  statistics to evaluate whether maize with *mexicana* introgression was transferred to other highland regions or whether highland adaptation was obtained *de novo* outside of Mexico. Overall, analyses revealed that maize landraces with *mexicana* introgression were transferred to nearby high elevation regions in Guatemala and the southwestern United States, but more distant high elevation regions (*e.g.*, the Andes) showed no *mexicana* ancestry.

### 2. Barley:

Barley (*Hordeum vulgare* subsp. *vulgare*) was domesticated at least twice roughly 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 [41, 42, 43, 44, 45, 46? ]. *While morrell2007 cites the Zargos mountains as a center of domestication, this paper is cited by ren2013tibet, which states that the findings of morrell2007genetic are just that a second origin exists east of the Fertile Crescent. It is cited by dai2012tibet and the poets paper, but little is said of its mention of the zargos mountains. I think that the mention of the Zargos mountains was just as a dividing line between the two spheres of influence between the two domestication centers.* 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 [47]. Barley/*spontaneum* hybrids are known to be fertile and are found spontaneously when wild and domesticated barleys co-occur. Introgression between wild and domesticated barleys appears to be frequent [46]at times occurring over distances of more than a kilometer [48].

The authors of [49] investigated the contribution of wild barleys to domesticated barley landraces from either center of domestication, looking specifically for patterns of contribution across regions of the genome and across geographical ranges. Broadly, these authors found several lines of evidence that indicate that introgression from many wild populations of barley may have been important in the dispersal and adaptation of domesticated barleys. Barley landraces show greater genetic contributions from proximate wild relatives than from distant ones, suggesting that these contributions may be locally-adaptive. Furthermore, genomic segments with greater contribution from a wild population than the average are more likely to contain regionally-unique, locally-adaptive alleles, and unequal genetic contributions at specific genomic regions to individuals within landraces may suggest that adaptive alleles from diverse geographical areas may have been combined. *I'll be honest, I don't actually follow the reasoning in the second half of the above sentence. It is a point made in the Poets paper. I think my confusion arises from their specific "ancestry" term they define. I'll have to return to this at another point.* Low linkage disequilibrium and small blocks of identity by state indicate that these introgressed regions are old, perhaps dating back to the beginning of barley domestication.

Genomic evidence points to a history of wild-crop introgression in the dispersal of barleys beyond their centers of domestication, but for the most part, the identity of the alleles critical to these adaptations have yet to be uncovered. One exception comes from the authors of [46], who hypothesize that alleles for cold stress tolerance and spring habit *"spring habit" are the words from the paper, but they're not defined. Does this mean that it grows more quickly in the spring?* from wild Tibetan 2-rowed barley allowed the barley to spread to the cold northern regions of Europe. Wild-domesticate breeding experiments have shown that wild barleys have alleles for several important fitness and agronomic phenotypes, including powdery mildew resistance, brittleness, flowering time, plant height, lodging, and yield [50, 51, 52], and these individually or in combination may have been facilitators of the dispersal of domesticated barley.

### 3. Sunflower:

The common sunflower (*Helianthus annuus*) shows evidence of domestication in the eastern United States [53, 54] with potential for a second domestication center in Mexico [55]. The pre-Columbian *H. annuus* distribution of cultivated sunflower spanned much of the Great Plains, from what is now north-central Texas to Montana and North Dakota (see figure 1 of [56]). Domesticated sunflower has long lived in sympatry with wild relatives such as *H. petiolaris* and *H. bolanderi* and forms stable hybrid populations with these taxa [57, 58, 59]. Wild sunflowers are known to be locally-adapted, and weedy hybrid populations often share these adaptations [60]. However, the most striking example of adaptive introgression within *Helianthus* is that of the cucumberleaf sunflower, *H. debilis* ssp. *cucumerifolius*. *I believe H.*

*annuus* ssp. *texanus* is a wild sunflower, right? If so, this is more an example of hybrid speciation rather than adaptive wild-to-crop gene flow

Cucumberleaf sunflower is endemic to south-central Texas, and exhibits several adaptations to the region. Introgressive hybridization imparted locally-adapted alleles from *H. debilis* to *H. annuus* via introgressive hybridization [61]. These introgressed hybrids formed a new lineage of sunflower (*H. annuus* ssp. *texanus*, *H. a. texanus* hereafter) which displays *H. debilis*-like traits adaptive to south-central Texas climate and ecology. These adaptive *debilis*-like traits include resistance to herbivorous pests and an increased branching plant architecture, as well as higher overall fitness than *H. annuus* (as measured by higher seed production [62]). Although *H. annuus* and *H. a. texanus* are interfertile, *H. a. texanus* displays persistent phenotypic differences from *H. annuus* [63].

The genome of the common sunflower has been greatly influenced by introgression from wild relatives, due to both natural outcrossing events and concerted breeding efforts in crop improvement. *Helianthus* has several genes for downy mildew resistance, and each imparts resistance to one or more races of *Plasmopara halstedii*, one of the most agronomically important diseases in sunflower cultivation [64]. Some of these downy mildew resistance genes were found in wild relatives (including *H. argophyllus*, *H. tuberosus*, and *H. praecox*) and have been successfully bred into modern *H. annuus* [65]. PlArg, an allele found in wild silverleaf sunflowers (*H. argophyllus*, inbred line Arg1575-2), confers resistance to all known (20 or more) races of downy mildew [66] while others (Pl1-Pl11) are effective for one or more types [67]. Silverleaf sunflower has also been the focus of drought resistance breeding efforts [68] and *Phomopsis* resistance breeding efforts [69]. *H. annuus* shows signs of persistent introgressive hybridization with *H. petiolaris* with evidence of positive selection driving some of the genetic differentiation between the two species [70].

Recent investigations into the history of *Helianthus* introgression have implemented genomic methods. [71] analyzed transcriptome sequence variation on cultivated and wild *H. annuus*, *H. petiolaris*, and *H. argophyllus*. Using STRUCTURE, these authors found that introgressions from wild relatives exist on every chromosome in at least one modern line, covering over 10% of the genome. Of particular note is the modern line RHA 274, a modern line which was bred with *H. a. texanus* in the 1970s to restore a branching plant body architecture, which allows the plant to produce pollen for a longer period of time, increasing seed production. RHA 274 has several large introgression from *H. a. texanus*, including one at the site of HaGNAT, the domestication gene associated with branching. These introgressed regions are not found in the non-branching lines Sunrise and VNIIMK8931, further suggesting that the *H. a. texanus* introgressed regions are causative.

Notes for review:

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

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

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

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

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

In contrast, coalescent analyses of long-term migration rates imply that interspecific migration has been more important than mutation in providing new genetic variation in ssp. *texanus* populations. Immigration from *H. debilis cucumerifolius* into ssp. *texanus* ( $M = 4.05$ ;  $Nem = 1.74$ ) is higher than median estimates of intraspecific migration rates in plants (Morjan

and Rieseberg 2004) and only slightly lower than immigration from ssp. *annuus* into ssp. *texasus* ( $M = 5.00$ ). Immigration rates from ssp. *texasus* into *H. debilis cucumerifolius* and ssp. *annuus* are still higher ( $M = 6.30$  and  $7.29$ , respectively), implying that the putative hybrid lineage serves as a bridge for migration between the parental species.

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

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

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

#### 4. Asian Rice:

The story of Asian rice (*Oryza sativa*) domestication is still debated. Interspecific hybridization between wild and domesticated rices both ancient and recent has produced a phylogeny that defies clear and simple models. On one hand, some genetic and archaeobotanical evidence point toward independent domestications of the two prominent varietal groups *japonica* and *indica* from the wild species *Oryza rufipogon* (*rufipogon* hereafter) in the Yangtze Basin and the Indian Ganges plain, respectively [72]. Other studies support a single center of domestication, arguing that present-day patterns of rice variation could be explained by a single domestication event. A recent investigation utilizing SNP data from wild and domesticated rice accessions pointed to a single domestication occurring 8,200-13,500 BP in the Yangtze

Basin in China from *rufipogon*, with later divergence of *indica* and *japonica* [73]. Huang and colleagues [74] developed a genetic map of rice variation, which they used to measure genetic distance between wild and domesticated rices at and around domestication loci for various geographical locations, finding that *japonica* was likely domesticated in the Pearl River area of Guangxi province, China (just south of the Yangtze Basin), and that *indica* was likely the result of hybridization between *japonica* and local *rufipogon* populations in Southern and South-eastern Asia.

Vaughan and colleagues [75] posit a hypothesis combining elements of both theories. According to this hypothesis, the non-shattering allele arose in a single cultivated rice population and then introgressed into many diverse cultivated populations (some japonica-like, some indica-like) in the region. This would have facilitated the subsequent spread of other domestication alleles among those populations. As these domesticated populations then spread into new environments, they would have received introgressions from locally-adapted wild relatives, retaining alleles that improved fitness (in agreement with [76]). Contemporary populations of wild and domesticated rice show evidence of this history of local adaptation. The wild relatives *rufipogon* and *nivara* both maintain high genotypic and phenotypic diversity and exhibit locally-adaptive traits (*rufipogon* is adapted to forested wetland environments, whereas *nivara* is adapted to dryer conditions and has life cycle adaptations to survive grazing pressure), suggesting that either offered (and offers) adaptive genetic potential. Likewise, domesticated rice varieties display patterns of local adaptation (for examples, two of the domesticated rice deepwater varieties (*rayada* and *ashwina*) are said to be selected for the environment along the Ganges river, the *japonicas* are split into temperate and tropical subgroups, and the *indicas* are best suited for lowland environments).

To date, research into adaptive introgression in the domestication of rice has been insufficient to detect clearly-supported examples (though [77] emphasizes the necessity of this research). There are perhaps some practical reasons why research has not yet been devoted to this inquiry. As with many other domesticated crops, gene flow between wild and domesticated rices is highly asymmetric (estimates of wild rice admixture in domesticated rice are less than 5 percent [78]). This asymmetry is due in part to the closed floret architecture of the domesticated rice, which hinders outcrossing. Crop-wild hybridization is frequent enough that nuisance weedy hybrids commonly arise amongst cultivated populations and must be managed with difficulty [75]. During early domestication, introgression may have been more prevalent than at present because barriers to crop-to-wild introgression may have been less severe and because the inbreeding reproductive system of rice would not have been as firmly established [75]. Furthermore, the contemporary distribution of wild rice does not capture the range and diversity of wild rice during early domestication and range expansion of rice.

## Re-evaluating domestication

A framework in which crops are domesticated from a single wild population or even a single species is an oversimplification when introgression during the geographic expansion of crops is extensive. The addition of ongoing gene flow to our understanding of crop demography could bear importantly on fundamental questions of crop domestication:



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

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

| Crop   | Compatible Wild Relatives   | Hybrids and/or Hybridization | Evidence of Crop Introgression | Evidence of Adaptiveness | Source |
|--|---|------------------------------|--------------------------------|--------------------------|--------|
| Maize ( <i>Zea mays</i> subsp. <i>mays</i> )                                   | <i>Z. m.</i> subsp. <i>mexicana</i> , <i>Z. m.</i> subsp. <i>parviglumis</i>                                      | X                            | X                              | X                        | [79]   |
| Asian Rice ( <i>Oryza sativa</i> )   | <i>O. rufipogon</i>   | X                            | X                              | X                        | [74]   |
| Barley ( <i>Hordeum vulgare</i> )  | <i>H. v.</i> subsp. <i>spontaneum</i>   | X                            | X                              | X                        | [49]   |
| Sunflower ( <i>Helianthus annuus</i> )   | <i>H. argophyllus</i> , <i>H. bolanderi</i> , <i>H. debilis</i> , <i>H. petiolaris</i>                            | X                            | X                              | X                        | [63]   |
| Cassava ( <i>Manihot esculenta</i> )   | <i>M. glaziovii</i>   | X                            | X                              | X                        | [80]   |
| Potato ( <i>Solanum tuberosum</i> )  | many  | X                            | X                              | X                        | [81? ] |
| Tomato ( <i>Solanum lycopersicum</i> )   | <i>S. pimpinellifolium</i>  | X                            | X                              | X                        | [82]   |
| Olive ( <i>Olea europaea</i> ssp. <i>europaea</i> var. <i>sativa</i> )         | <i>O. e.</i> ssp. <i>europaea</i> var. <i>sylvestris</i>  | X                            | X                              |                          | [83]   |
| Soybeans ( <i>Glycine max</i> )  | <i>G. soja</i>  | X                            | X                              |                          | [84]   |
| Common Bean ( <i>Phaseolus vulgaris</i> )                                      | <i>P. v.</i> var. <i>aborigineus</i> , <i>P. v.</i> var. <i>mexicanus</i> [[not in this source]]                  | X                            | X                              |                          | [85]   |
| Grapes ( <i>Vitis vinifera</i> subsp. <i>vinifera</i> )                        | <i>V. v.</i> subsp. <i>sylvestris</i>   | X                            | X                              |                          | [86]   |
| Sorghum ( <i>Sorghum bicolor</i> subsp. <i>bicolor</i> )                       | <i>S. b.</i> subsp. <i>arundinaceum</i> , <i>S. b.</i> subsp. <i>drummondii</i>                                   | X                            | X                              |                          | [87]   |
| Wheat ( <i>Triticum monococcum</i> , <i>T. dicoccum</i> , <i>T. aestivum</i> ) | <i>T. m. boeoticum</i> , <i>T. diocoides</i> , <i>T. urartu</i> , <i>Aegilops speltoides</i> , <i>A. tauschii</i> | X                            | X                              |                          | [88]   |
| Apple ( <i>Malus domestica</i> )   | <i>M. sylvestris</i> , <i>M. orientalis</i> , <i>M. baccata</i> , <i>M. sieversii</i>                             | X                            | X                              |                          | [89]   |



## What is the progenitor of a crop?

Depending on the extent of post-domestication gene flow with new wild relatives, identification of a crop’s progenitor can be complicated or completely confounded. The level of divergence of a crop from newly encountered populations and species will decrease due to introgression, a signal that could be mistaken for origin rather than gene flow. For example, when determining a single origin of maize from *parviglumis*, Matsuoka and colleagues [32] identified a paradox: while *parviglumis* is found exclusively in the lowlands of southwest Mexico, maize with allele frequencies most similar to *parviglumis* was found in the highlands of the Mexican Central Plateau. Several years later, van Heerwaarden *et al.* [36] 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 [90, 91]. Beyond confounding detection of progenitor taxa, extensive introgression may necessitate a reevaluation of crop origins. In cases like maize and potato it is important to recognize the substantial contributions of introgressing taxa to the genetic base of modern crops. Broad recognition of the role these wild relatives have played in crop adaptation could further their use in breeding and elevate their conservation status.

## When was a crop domesticated?

Estimates of the timing of initial domestication are often based on levels of sequence divergence between a crop and populations of its presumed progenitor (*e.g.*, [32, 73]). In highly introgressed domesticates, these estimates will be based on comparison of either crop or introgressant haplotypes to those of the progenitor. In such cases, divergence is a mixture of time since domestication and time since split of the progenitor and the introgressing species. This phenomenon, in combination with divergence of samples from true ancestral populations, ongoing evolution of crop progenitors, and problems with assuming evolution under a molecular clock [92], may explain discrepancies between domestication dates based on genetic and archaeological data. More accurate estimates of the timing of domestication may be obtained from genetic data by excluding loci that show signatures of introgression.

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

Measurement of the strength of the initial domestication bottleneck may also be impacted by adaptive introgression during the spread of crops. Crop wild relatives have distinct demographies when compared to domesticates and may therefore have contrasting effective population sizes ( $N_e$ ). The influence of wild relative introgression on estimates of the domestication bottleneck will depend on a number of factors including the magnitude of gene flow, the  $N_e$  of the introgressing taxon, and the strength of selection on haplotypes following introgression. For example, substantial introgression at neutral loci from a wild taxon with a historically higher  $N_e$  will lead to underestimates of the overall strength of the domestication bottleneck.

## What candidate genes were targeted by selection during domestication?

Loci targeted by selection during domestication can be identified through so-called “bottom-up” approaches based on population genetic signatures [93]. Ideally, candidate loci will be identified

by first constructing a demographic model representing the history of the domesticate. In this approach, diversity data from neutral loci are fit to potential models of a crop’s demography and then statistical tests of selection are used to identify candidate domestication genes under the most likely model. Due to the difficulty of this approach and the uncertainty associated with any given demographic model, many studies identify domestication loci using a strict outlier approach in which loci showing the greatest reduction in nucleotide diversity or the highest allele frequency differentiation in the domesticate relative to the wild progenitor are identified as candidates. Introgression during crop expansion may influence candidate gene detection using both demographic-modeling and strict-outlier approaches. For example, *mexicana* introgression into maize described above accounts for approximately 20% of the genome of maize in the highlands of Mexico [36]. Takuno and co-authors [94] have shown that a demographic model incorporating this introgression is a significantly better fit to empirical data than a model lacking introgression. Failure to account for introgression in maize would therefore compromise candidate detection, particularly if a study contained maize samples from the Mexican highlands. Likewise, introgression that increased nucleotide diversity in the domesticate or decreased differentiation at domestication loci would confound a strict outlier approach. However, previous work, also in maize, has shown that known domestication loci are particularly resistant to introgression [79], likely due to ongoing selection favoring the domesticated phenotype.

## Future studies in crop-wild introgression

Basic: We’re beginning to see the introgression has occurred across a number of crops during their expansion. What is the genomic architecture of this introgression and does the architecture suggest it’s been adaptive? At what geographic scale is introgression adaptive? Is it very local? Do you see different architectures of introgression across different regions? At what taxonomic scale does introgression occur? When do species become so diverged that introgression is either not possible or maladaptive?

Research has so far shown that adaptive crop-wild introgression has played a significant role in the domestication histories of many agronomically-important crops. However, the dynamics of the process in these cases are not yet fully understood. To what extent does the level of introgression across taxa depend on divergence time and/or mutation load between donor and recipient taxa? At what geographic scale does adaptive introgression occur? Is introgression frequently restricted to very local populations, or is it often seen over broad geographic ranges? To what extent does this depend on the slope of environmental gradients such as temperature, precipitation, and elevation?

Applied: Additional study of introgression in agroecosystems could lead to advances in both basic and applied genetics, and specifically the continued improvement of modern crops. Loci underlying the domesticated phenotype can be more clearly identified by removing the confounding population genetic signal of introgression. These loci are potentially beneficial targets for crop improvement. Furthermore, adaptive introgression that is clearly tied to a specific environment may include beneficial alleles that can be utilized in crop breeding.

## Conclusions

The study of crop domestication has been revolutionized by the advent and application of genomic tools. The genomes of crops and their wild relatives tell a story of give-and-take that extends well beyond the initial stages of domestication. Likewise, population genetic theory reinforces the

proclivity of wild relatives to provide advantageous, locally-adapted alleles to crops as they disperse beyond their domestication centers into new geographies with new ecological pressures and niches.

## References

- [1] J. R. Harlan, *Crops & Man*. Madison, WI: American Society of Agronomy, Crop Science Society of America, 1992.
- [2] T. A. Brown, M. K. Jones, W. Powell, and R. G. Allaby, “The complex origins of domesticated crops in the fertile crescent,” *Trends in Ecology & Evolution*, vol. 24, no. 2, pp. 103–109, 2009.
- [3] R. S. Meyer, J. Y. Choi, M. Sanches, A. Plessis, J. M. Flowers, J. Amas, K. Dorph, A. Barretto, B. Gross, D. Q. Fuller, I. K. Bimpong, M.-N. Ndjiondjop, K. M. Hazzouri, G. B. Gregorio, and M. D. Purugganan, “Domestication history and geographical adaptation inferred from a snp map of african rice,” *Nat Genet*, vol. 48, pp. 1083–1088, 09 2016.
- [4] L. Wang, T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, and M. Hufford, “The interplay of demography and selection during maize domestication and expansion,” *bioRxiv*, p. 114579, 2017.
- [5] Y. Zhou, M. Massonnet, J. Sanjak, D. Cantu, and B. S. Gaut, “The evolutionary genomics of grape (*vitis vinifera* ssp. *vinifera*) domestication,” *bioRxiv*, p. 146373, 2017.
- [6] D. Q. Fuller, T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, and M. D. Purugganan, “Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 17, pp. 6147–6152, 2014.
- [7] N. Barton, “The role of hybridization in evolution,” *Molecular Ecology*, vol. 10, no. 3, pp. 551–568, 2001.
- [8] C. Feuillet, P. Langridge, and R. Waugh, “Cereal breeding takes a walk on the wild side,” *Trends in Genetics*, vol. 24, no. 1, pp. 24 – 32, 2008.
- [9] C. N. Stewart, M. D. Halfhill, and S. I. Warwick, “Transgene introgression from genetically modified crops to their wild relatives,” *Nature Reviews Genetics*, vol. 4, no. 10, pp. 806–817, 2003.
- [10] H. Dempewolf, G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino, “Past and future use of wild relatives in crop breeding,” vol. 57, pp. 1070–1082, 2017.
- [11] S. H. Martin, J. W. Davey, and C. D. Jiggins, “Evaluating the use of abba–baba statistics to locate introgressed loci,” *Molecular biology and evolution*, vol. 32, no. 1, pp. 244–257, 2015.
- [12] A. L. Price, A. Tandon, N. Patterson, K. C. Barnes, N. Rafaels, I. Ruczinski, T. H. Beaty, R. Mathias, D. Reich, and S. Myers, “Sensitive detection of chromosomal segments of distinct ancestry in admixed populations,” *PLoS Genet*, vol. 5, no. 6, pp. 1–18, 2009.
- [13] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush, “Inference of population structure using dense haplotype data,” *PLoS Genet*, vol. 8, no. 1, pp. 1–16, 2012.

- [14] J. B. Pease and M. W. Hahn, "Detection and polarization of introgression in a five-taxon phylogeny," *Systematic biology*, vol. 64, no. 4, pp. 651–662, 2015.
- [15] B. K. Rosenzweig, J. B. Pease, N. J. Besansky, and M. W. Hahn, "Powerful methods for detecting introgressed regions from population genomic data," *Molecular ecology*, 2016.
- [16] A. J. Geneva, C. A. Muirhead, S. B. Kingan, and D. Garrigan, "A new method to scan genomes for introgression in a secondary contact model," *PloS one*, vol. 10, no. 4, p. e0118621, 2015.
- [17] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.
- [18] R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, "Inferring the joint demographic history of multiple populations from multidimensional snp frequency data," *PLoS Genet*, vol. 5, no. 10, p. e1000695, 2009.
- [19] J. Hey and R. Nielsen, "Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*," *Genetics*, vol. 167, no. 2, pp. 747–760, 2004.
- [20] H. Li and R. Durbin, "Inference of human population history from individual whole-genome sequences," *Nature*, vol. 475, no. 7357, pp. 493–496, 2011.
- [21] S. Schiffels and R. Durbin, "Inferring human population size and separation history from multiple genome sequences," *Nature genetics*, vol. 46, no. 8, pp. 919–925, 2014.
- [22] J. Terhorst, J. A. Kamm, and Y. S. Song, "Robust and scalable inference of population history from hundreds of unphased whole genomes," tech. rep., 2017.
- [23] F. Racimo, D. Marnetto, and E. Huerta-Sánchez, "Signatures of archaic adaptive introgression in present-day human populations," *Molecular Biology and Evolution*, p. msw216, 2016.
- [24] J. G. Schraiber and J. M. Akey, "Methods and models for unravelling human evolutionary history," *Nature Reviews Genetics*, 2015.
- [25] A. Brisbin, K. Bryc, J. Byrnes, F. Zakharia, L. Omberg, J. Degenhardt, A. Reynolds, H. Ostrer, J. G. Mezey, and C. D. Bustamante, "Pcadmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations," *Human biology*, vol. 84, no. 4, p. 343, 2012.
- [26] S. Sankararaman, S. Sridhar, G. Kimmel, and E. Halperin, "Estimating local ancestry in admixed populations," *The American Journal of Human Genetics*, vol. 82, no. 2, pp. 290–303, 2008.
- [27] J. K. Pritchard, M. Stephens, and P. Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, vol. 155, no. 2, pp. 945–959, 2000.
- [28] F. Racimo, S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez, "Evidence for archaic adaptive introgression in humans," *Nature Reviews Genetics*, vol. 16, no. 6, pp. 359–371, 2015.
- [29] T. H. G. Consortium, "Butterfly genome reveals promiscuous exchange of mimicry adaptations among species," *Nature*, vol. 487, no. 7405, pp. 94–98, 2012.

- [30] L. C. Norris, B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro, "Adaptive introgression in an african malaria mosquito coincident with the increased usage of insecticide-treated bed nets," *Proceedings of the National Academy of Sciences*, vol. 112, no. 3, pp. 815–820, 2015.
- [31] S. Stankowski and M. A. Streisfeld, "Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 282, no. 1814, 2015.
- [32] Y. Matsuoka, Y. Vigouroux, M. M. Goodman, J. Sanchez, E. Buckler, and J. Doebley, "A single domestication for maize shown by multilocus microsatellite genotyping," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 6080–6084, 2002.
- [33] H. Wilkes, "Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize," *Economic Botany*, pp. 254–293, 1977.
- [34] 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.
- [35] J. F. Doebley, "Maize introgression into teosinte-a reappraisal," *Annals of the Missouri Botanical Garden*, pp. 1100–1113, 1984.
- [36] 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.
- [37] 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.
- [38] M. L. Warburton, G. Wilkes, S. Taba, A. Charcosset, C. Mir, F. Dumas, D. Madur, S. Dreisigacker, C. Bedoya, B. Prasanna, *et al.*, "Gene flow among different teosinte taxa and into the domesticated maize gene pool," *Genetic resources and crop evolution*, vol. 58, no. 8, pp. 1243–1261, 2011.
- [39] K. Fukunaga, J. Hill, Y. Vigouroux, Y. Matsuoka, J. Sanchez, K. Liu, E. S. Buckler, and J. Doebley, "Genetic diversity and population structure of teosinte," *Genetics*, vol. 169, no. 4, pp. 2241–2254, 2005.
- [40] M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra, "The genomic signature of crop-wild introgression in maize," *Plos genetics*, 2013.
- [41] R. Takahashi, "The origin and evolution of cultivated barley," *Advances in genetics*, vol. 7, pp. 227–266, 1955.
- [42] 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.
- [43] H.-I. Oka, *Origin of cultivated rice*. Elsevier, 2012.
- [44] P. Azhaguvel and T. Komatsuda, "A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley," *Annals of botany*, vol. 100, no. 5, pp. 1009–1015, 2007.

- [45] G. Haberer and K. F. Mayer, “Barley: From brittle to stable harvest,” *Cell*, vol. 162, no. 3, pp. 469–471, 2015.
- [46] 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.
- [47] E. Nevo and G. Chen, “Drought and salt tolerances in wild relatives for wheat and barley improvement,” *Plant, cell & environment*, vol. 33, no. 4, pp. 670–685, 2010.
- [48] G. Hillman, R. Hedges, A. Moore, S. Colledge, and P. Pettitt, “New evidence of lateglacial cereal cultivation at abu hureyra on the euphrates,” *The Holocene*, vol. 11, no. 4, pp. 383–393, 2001.
- [49] 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.
- [50] A. Dreiseitl, “Heterogeneity of powdery mildew resistance revealed in accessions of the icarda wild barley collection,” *Frontiers in Plant Science*, vol. 8, 2017.
- [51] M. Von Korff, H. Wang, J. Léon, and K. Pillen, “Ab-qt1 analysis in spring barley: Ii. detection of favourable exotic alleles for agronomic traits introgressed from wild barley (h. vulgare ssp. spontaneum),” *Theoretical and Applied Genetics*, vol. 112, no. 7, pp. 1221–1231, 2006.
- [52] L. L. Handley, E. Nevo, J. A. Raven, R. MartInez-Carrasco, C. M. Scrimgeour, H. Pakniyat, and B. P. Forster, “Chromosome 4 controls potential water use efficiency ( $\delta^{13}C$ ) in barley,” *Journal of Experimental Botany*, vol. 45, no. 11, pp. 1661–1663, 1994.
- [53] A. V. Harter, K. A. Gardner, D. Falush, D. L. Lentz, R. A. Bye, and L. H. Rieseberg, “Origin of extant domesticated sunflowers in eastern north america,” *Nature*, vol. 430, no. 6996, pp. 201–205, 2004.
- [54] D. M. Wills and J. M. Burke, “Chloroplast dna variation confirms a single origin of domesticated sunflower (helianthus annuus l.),” *Journal of Heredity*, vol. 97, no. 4, pp. 403–408, 2006.
- [55] D. L. Lentz, M. D. Pohl, J. L. Alvarado, S. Tarighat, and R. Bye, “Sunflower (helianthus annuus l.) as a pre-columbian domesticate in mexico,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 17, pp. 6232–6237, 2008.
- [56] K. D. Whitney, R. A. Randell, and L. H. Rieseberg, “Adaptive introgression of abiotic tolerance traits in the sunflower helianthus annuus,” *New Phytologist*, vol. 187, no. 1, pp. 230–239, 2010.
- [57] A. Schwarzbach and L. Rieseberg, “Likely multiple origins of a diploid hybrid sunflower species,” *Molecular Ecology*, vol. 11, no. 9, pp. 1703–1715, 2002.
- [58] L. H. Rieseberg, D. E. Soltis, and J. D. Palmer, “A molecular reexamination of introgression between helianthus annuus and h bolanderi (compositae),” *Evolution*, pp. 227–238, 1988.
- [59] M. E. Welch and L. H. Rieseberg, “Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species helianthus paradoxus,” *Evolution*, vol. 56, no. 11, pp. 2126–2137, 2002.

- [60] N. C. Kane and L. H. Rieseberg, "Genetics and evolution of weedy helianthus annuus populations: adaptation of an agricultural weed," *Molecular Ecology*, vol. 17, no. 1, 2008.
- [61] C. B. Heiser Jr, "Hybridization in the annual sunflowers: *Helianthus annuus* × *h. debilis* var. *cucumerifolius*," *Evolution*, pp. 42–51, 1951.
- [62] K. D. Whitney, R. A. Randell, and L. H. Rieseberg, "Adaptive introgression of herbivore resistance traits in the weedy sunflower *helianthus annuus*," *The American Naturalist*, vol. 167, no. 6, pp. 794–807, 2006.
- [63] 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.
- [64] Y. Cohen and W. Sackston, "Factors affecting infection of sunflowers by *plasmopara halstedii*," *Canadian Journal of Botany*, vol. 51, no. 1, pp. 15–22, 1973.
- [65] J. Miller and T. Gulya, "Inheritance of resistance to race 4 of downy mildew derived from interspecific crosses in sunflower," *Crop science*, vol. 31, no. 1, pp. 40–43, 1991.
- [66] C. Dussle, V. Hahn, S. Knapp, and E. Bauer, "Pl arg from *helianthus argophyllus* is unlinked to other known downy mildew resistance genes in sunflower," *Theoretical and applied genetics*, vol. 109, no. 5, pp. 1083–1086, 2004.
- [67] M. Rahim, C. Jan, and T. Gulya, "Inheritance of resistance to sunflower downy mildew races 1, 2 and 3 in cultivated sunflower," *Plant Breeding*, vol. 121, no. 1, pp. 57–60, 2002.
- [68] F. Sauca *et al.*, "Introgression of drought-resistance gene (s) from *helianthus argophyllus* to *helianthus annuus* specie, using embryo rescue techniques," *Romanian Agricultural Research*, no. 27, pp. 47–51, 2010.
- [69] G. Besnard, Y. Griveau, M. Quillet, H. Serieys, P. Lambert, D. Vares, and A. Bervillé, "Specifying the introgressed regions from *h. argophyllus* in cultivated sunflower (*helianthus annuus* l.) to mark phomopsis resistance genes," *Theoretical and applied genetics*, vol. 94, no. 1, pp. 131–138, 1997.
- [70] Y. Yatabe, N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg, "Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *helianthus annuus* and *h. petiolaris*," *Genetics*, vol. 175, no. 4, pp. 1883–1893, 2007.
- [71] G. J. Baute, N. C. Kane, C. J. Grassa, Z. Lai, and L. H. Rieseberg, "Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives," *New Phytologist*, vol. 206, no. 2, pp. 830–838, 2015.
- [72] D. Q. Fuller, Y.-I. Sato, C. Castillo, L. Qin, A. R. Weisskopf, E. J. Kingwell-Banham, J. Song, S.-M. Ahn, and J. Van Etten, "Consilience of genetics and archaeobotany in the entangled history of rice," *Archaeological and Anthropological Sciences*, vol. 2, no. 2, pp. 115–131, 2010.
- [73] 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.



- [74] 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.
- [75] D. A. Vaughan, B.-R. Lu, and N. Tomooka, “The evolving story of rice evolution,” *Plant science*, vol. 174, no. 4, pp. 394–408, 2008.
- [76] G. Second, “Origin of the genic diversity of cultivated rice (*oryza* spp.): study of the polymorphism scored at 40 isozyme loci,” , vol. 57, no. 1, pp. 25–57, 1982.
- [77] K. Zhao, M. Wright, J. Kimball, G. Eizenga, A. McClung, M. Kovach, W. Tyagi, M. L. Ali, C.-W. Tung, A. Reynolds, *et al.*, “Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome,” *PloS one*, vol. 5, no. 5, p. e10780, 2010.
- [78] H. Wang, F. G. Vieira, J. E. Crawford, C. Chu, and R. Nielsen, “Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice,” *Genome Research*, vol. 27, no. 6, pp. 1029–1038, 2017.
- [79] M. B. Hufford, P. Lubinsky, T. Pyhäjärvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra, “The genomic signature of crop-wild introgression in maize,” 2013.
- [80] 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.
- [81] T. Johns and S. L. Keen, “Ongoing evolution of the potato on the altiplano of western bolivia,” *Economic Botany*, vol. 40, no. 4, pp. 409–424, 1986.
- [82] 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.
- [83] 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.
- [84] 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.
- [85] 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.
- [86] 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.
- [87] 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.

- [88] 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.
- [89] 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.
- [90] D. M. Spooner, M. Ghislain, R. Simon, S. H. Jansky, and T. Gavrilenko, “Systematics, diversity, genetics, and evolution of wild and cultivated potatoes,” *The Botanical Review*, vol. 80, no. 4, pp. 283–383.
- [91] T. Gavrilenko, O. Antonova, A. Shuvalova, E. Krylova, N. Alpatyeva, D. M. Spooner, and L. Novikova, “Genetic diversity and origin of cultivated potatoes based on plastid microsatellite polymorphism,” *Genetic Resources and Crop Evolution*, vol. 60, no. 7, pp. 1997–2015, 2013.
- [92] 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.
- [93] 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.
- [94] S. Takuno, P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra, “Independent molecular basis of convergent highland adaptation in maize,” *Genetics*, 2015.

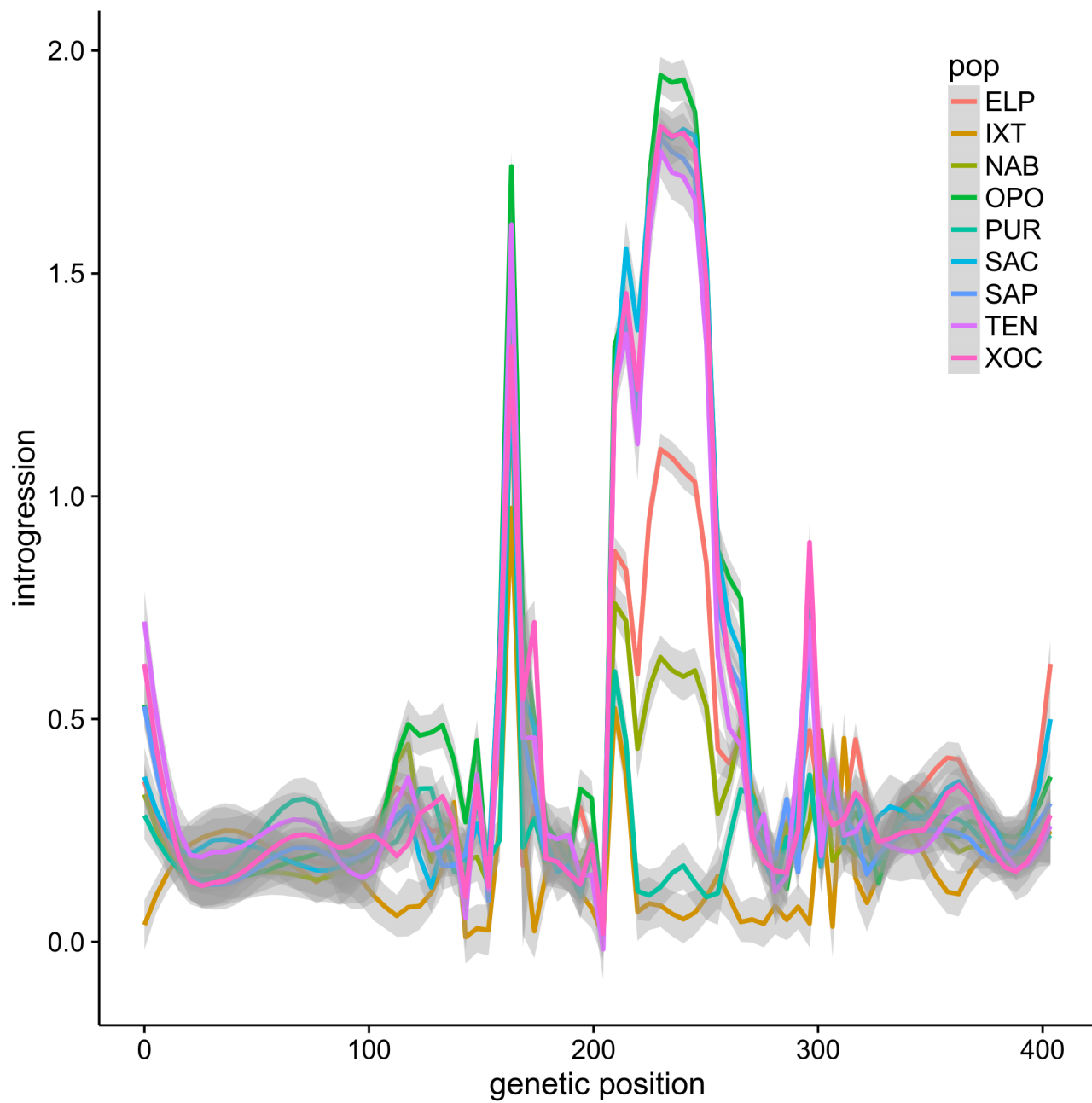


Figure 1: Li's caption here.

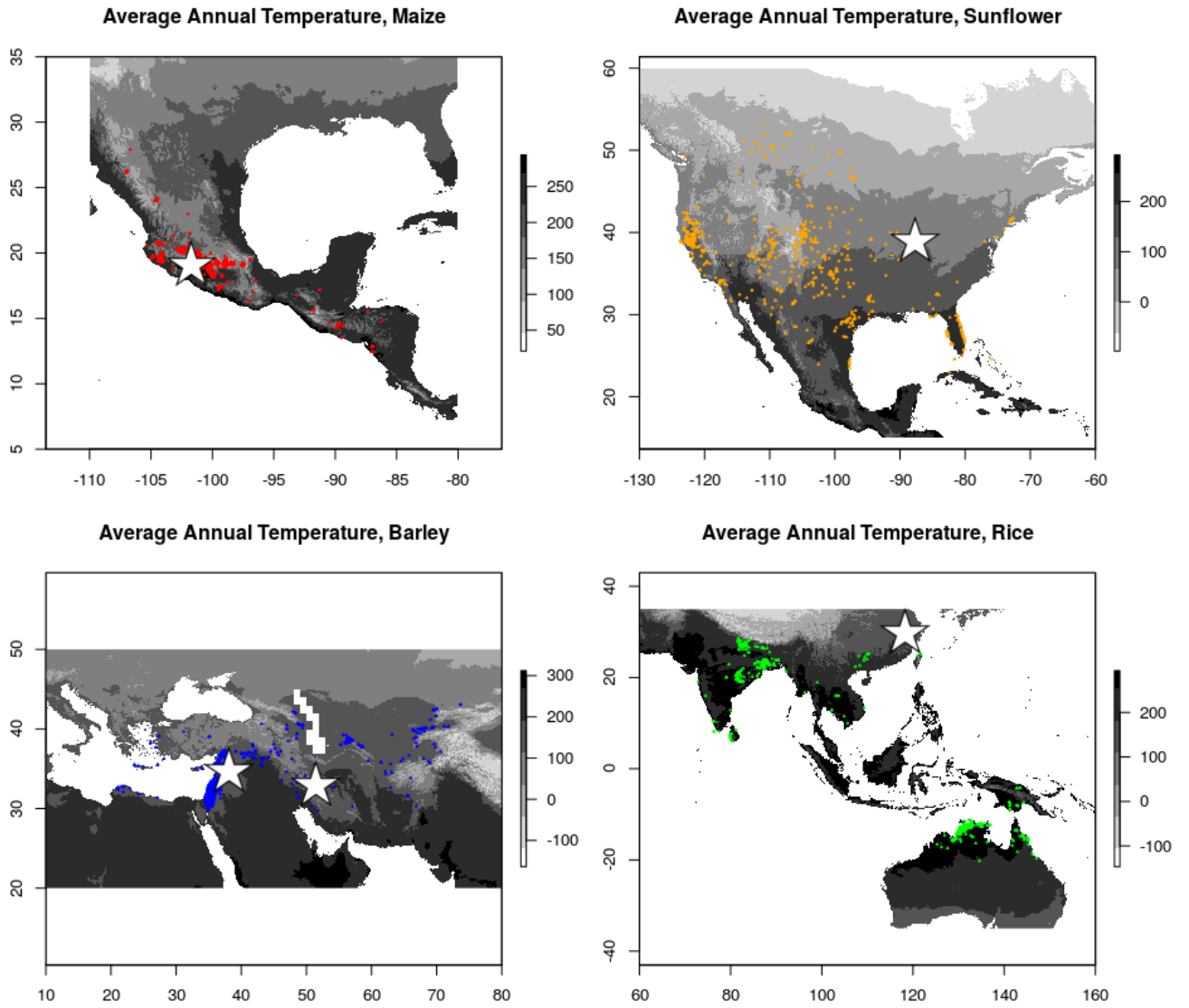


Figure 2: Map of the natural ranges of wild relatives of four domesticated crops, overlaid with average annual temperature.

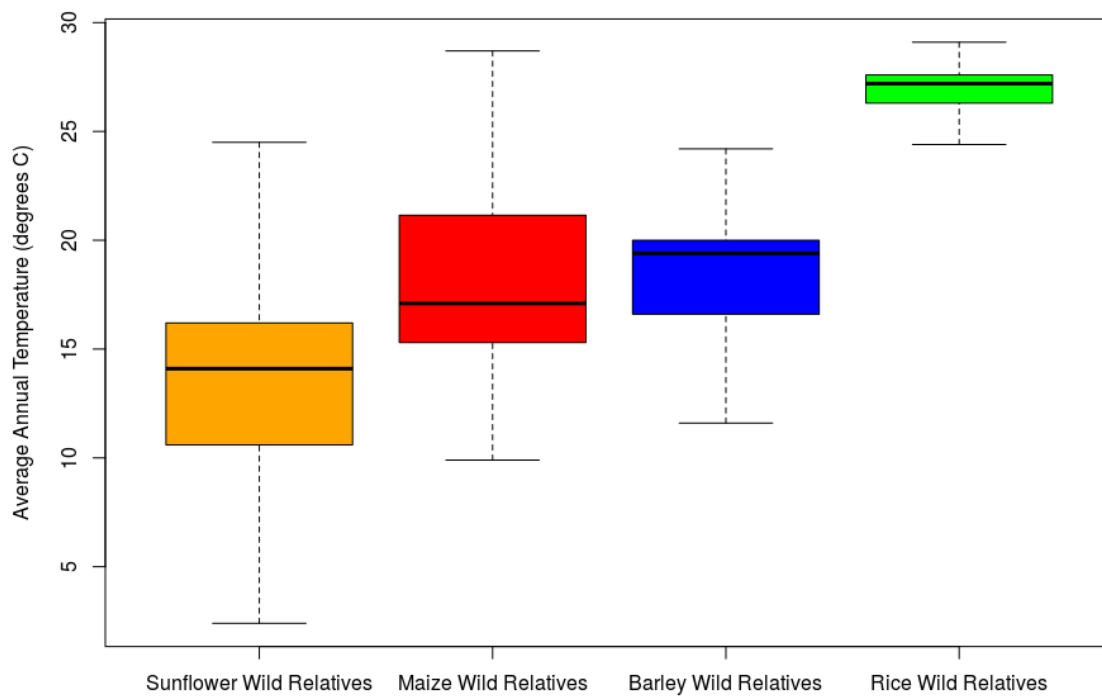


Figure 3: The distribution of average annual temperature experienced in the geographic home ranges of wild relatives interfertile with four crops