

# Review: Redefining Domestication: Adaptive Introgression during Crop Expansion

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 was once thought to be rapid and geographically constrained, with crops originating from a wild progenitor within one or more geographically defined centers followed by expansion to the modern-day extent of cultivation. *will need citation for this* However, archaeological and genetic evidence are beginning to reveal that, in many cases, domestication has been temporally protracted and more diffuse [1]. *citations here to Li's paper, Purugganan rice paper, Brandon Gaut's recent grape paper* This new conception of domestication emphasizes the role of beneficial gene flow (*i.e.*, adaptive introgression) from locally adapted wild relatives during crop expansion after 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 [2, 3]. 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 removal of undesirable wild background. Literature on crop-wild gene flow has focused on the risk of introgression of transgenes from domesticated crops into wild relatives (for a review, [4]) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, [5, 6, 7]). The history of natural introgression of wild alleles into domesticated crops over evolutionary timescales has received considerably less attention. Recent tools and methods have been employed to detect genome-wide patterns of introgression, granting new insights into the prevalence of adaptive introgression in crop histories. Emerging results suggest a need for reevaluation of the existing domestication paradigm.

In this review, we will: 1) briefly describe recently developed methods for detecting adaptive introgression and provide a summary of their application for detecting crop-wild introgression, 2) review evidence supporting the hypothesis that wild-to-crop introgression has conferred local adaptation, 3) consider how the prevalence of this introgression alters traditional concepts 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

The decreasing cost of genome-wide resequencing and availability of reduced-representation genotyping (*e.g.*, GBS and RAD-Seq), combined with new analytical methods, has facilitated comprehensive study of introgression across a number of species (**Table 1**). High-density marker data can be used with haplotype-based and other methods to assign specific genomic regions to a taxon of

origin and identify introgression across taxa [? ? ? ? ? ? ]. The methods reviewed here do not include those marginally estimating introgression/migration rate as a component of demographic history (*e.g.*, Approximate Bayesian Computation (ABC) [? ], diffusion approximations for demographic inference ( $\delta a \delta i$ ) [? ], isolation with migration models [? ], and the multiple sequentially Markovian coalescent (MSMC) [? ]). Rather, we focus on methods that explicitly identify introgressed genomic segments based on patterns of nucleotide/haplotype diversity and differentiation and on 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 such as  $G_{min}$  [? ] and  $RND_{min}$  [? ] gauge differentiation. The former two 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.

*abrupt transition here to Gmin equation...can you smooth this into the text?*

$$G_{min} = \frac{d_{min}}{d_{XY}} \quad (1)$$

where  $d_{min}$  is the minimum sequence distance between haplotypes in species X and Y.

$$RND_{min} = \frac{d_{min}}{d_{out}} \quad (2)$$

where  $d_{out}$  equals  $(d_{XO} + d_{YO})/2$ , the average sequence distance between each species and the outgroup ( $O$ ).

These statistics have recently been further developed by adding differentiation between both non-admixed ( $A$ ) and admixed populations ( $B$ ) and a source population ( $C$ ) [? ]. For example, the  $U_{A,B,C(w,x,y)}$  statistic summarizes number of sites, where an allele at frequency  $y$  in the source population ( $C$ ) has a frequency higher than  $x$  in the admixed population ( $B$ ) and lower than  $w$  in the non-admixed population ( $A$ ). 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 ( $B$ ) [? ]. *not quite sure I follow here, needs to be made more clear* These statistics were further developed to allow specification two source populations (see details in [? ]).

Second, local ancestry reconvolution (also known as chromosome painting) can identify the genomic regions that come from different source populations [? ]. Such methods usually take phased haplotypes as input, but MULTIMIX [? ] can take both phased and unphased haplotypes as input, unveiling the genomic ancestry from multiple source populations.

Third, ABBA-BABA statistics (also known as D-statistics) and their derivatives are widely applied to the search for signs of introgression in genomic patterns of shared derived variants between populations or species. Based on D-statistics,  $\hat{f}_d$  [? ] and five-taxon D statistics [? ] were developed to localize genomic regions under introgression. The former took allele frequencies from each population/species, and the latter detected introgression from the localized phylogenetic pattern. Their strength in identifying specific regions enables one to explore the relationships between introgression and recombination rate/gene density/distribution of deleterious alleles and to better understand the genomic regions penetrable to foreign gene flow.

Fourth, structure-related approaches are based on the popular software STRUCTURE [? ], which provides an estimate of the fraction of each individual's genome coming from each population. Many approaches in this respect have been developed. Two of these, fineSTRUCTURE [? ] and GLOBETROTTER [? ], were selected as representations. Different from STRUCTURE, fineSTRUCTURE not only identifies the number of populations, but also estimates the probability that genomic regions belong to one or the other populations. GLOBETROTTER is a development

of fineSTRUCTURE that allows for tracing back to an unsampled source population and dating admixture events.

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

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

**Figure 1:** Wing coloration patterns in *Heliconius* and evidence for introgression across species based on Patterson’s *D*-statistic; adapted from [? ].

## 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 crop/wild introgression. Maize was domesticated from (*Zea mays* ssp. *parviglumis*) approximately 9,000 years ago in the Balsas River Valley in Mexico [8]. From this domestication center, maize spread into the highlands of the Mexican central plateau, where it came into sympatry with wild *mexicana*. Introgression between *mexicana* and Mexican highland maize has been reported based on evidence from both morphological data [? ? ? ] and molecular analyses [? ? ? ? ? ]. However, [? ] was the first to reveal evidence that introgression from *mexicana* into Mexican highland maize is adaptive. The authors identified nine genomic regions which showed evidence of introgression from *mexicana* to maize in both the HAMP-MIX and the linkage model of STRUCTURE analyses with over seven sympatric population pairs among the nine pairs sampled. Among the nine regions, three span the centromeres of chromosomes 5, 6, and 10, and one is located in the inversion polymorphism on chromosome 4, suggesting a significant role of genome structures restricting recombination in adaptive introgression. By further characterizing the nine introgression regions, it is found that most regions contain long tracts of zero diversity, enriched with QTL linked with anthocyanin content and leaf macrohairs [? ] and over-represented with the SNPs demonstrating high association with temperature seasonality. It was also revealed that in the growth chamber experiment, maize populations with introgression from *mexicana* on chromosomes 4 (associated with QTLs controlling pigment density and macrohairs) and 9 (overlapped with QTLs for macrohairs) exhibited more macrohairs and greater pigmentation under the highland environmental settings than the populations with absence of introgression from *mexicana*. In summary, the introgressed alleles/haplotypes from *mexicana* to maize conferred adaptation to highland habitats when maize migrated from Mexican lowlands.

Even though introgression from *mexicana* to Mexican highland maize has been highly supported, it is yet unknown whether and to what extent such introgression could be found in other highland maize populations which are allopatric to *mexicana*. A recent study [?] found little empirical or theoretical support for parallel highland adaptation in the Mexican and South American highland regions, which was explained partly by the difference in potential for adaptive introgression from wild relatives. Adaptive SNPs in Mexican highland population were more likely located in the introgressed regions than those in South American highland populations. Furthermore, the adaptive SNPs in the Mexican highland population were more likely also showing signatures of local adaptation in *mexicana* and *pulvigliumis* populations than those South American highland population. For these reasons, adaptive introgression from wild relatives may play a significant role in patterning the genetic differences of maize highland populations.

Several questions regarding *mexicana* introgression yet remain. Has the introgression from *mexicana* spread to multiple highland populations? How do highland-adaptation traits differ among populations with and without introgression from *mexicana*? A recent study [?] first utilized the ABBA-BABA statistics to evaluate the existence or absence of introgression from *mexicana* to multiple highland populations, and then  $\hat{f}_d$  statistic proposed by Martin et al. (2015) [?] were calculated to locate the introgressed loci. Figure ?? shows the loess regression of  $\hat{f}_d$  in 10kb nonoverlapping windows across chromosomes 4 and 5 in multiple comparisons. On chromosome 4 (Figure ??), Mexican highland (MexHigh) and Guatamalan highland (GuaHigh) exhibited strong evidence of introgression from *mexicana*, and the peaks of distribution corresponded to the region identified in Hufford et al. (2013) [?]. The signal of introgression is absent for the other three populations. On chromosome 5, the signals of introgression (the peak region of the distribution) are present in MexHigh, GuaHigh and Southwestern US Highland (SW\_US), but not the other two. More details on the other chromosomes can be found in [?]. Overall, both analyses revealed that gene flow from *mexicana* exists, not only to MexHigh maize, but also to GuaHigh maize, as well as to some individuals in SW\_US, but not to the Andes and South American lowland (SA\_Low) populations, suggesting different extent contributions *mexicana* made to multiple maize highland landrace populations. The Andean maize, the population totally isolated from the occurrence of any teosinte species, underwent the severest historical bottleneck, as a population in the front wave of the serial founder effects. The high frequency of deleterious alleles caused by stronger genetic drift in the Andes population, together with the reduced efficiency of selection against deleterious sites, contributes to the observed higher mutation load in the well-isolated maize population. Although it is clear that the absence of introgression from wild relatives provided fewer genetic resources for highland adaptation (making the highland adaptation in the Andes unique), it is yet unknown whether or not being out of reach of wild relatives is a reason for reduced fitness in the Andes population. Furthermore, the question of whether convergent evolution occurs between populations with and without introgression from wild relatives is a key topic for future studies in maize.

## 2. Asian Rice:

Introgression appears to have played an important role in the domestication history of rice. The center of Asian rice (*Oryza sativa*) domestication is not known with complete confidence, but genetic and archaeobotanical evidence points towards both the Yangzee Basin in China and the Ganges plains in India for *O. sativa*, 8,200-13,500 years ago, from wild *O. rufipogon* [9, 10, 11, 12] or "the Asian form of *O. perennis* complex" [9]. Asian rices readily hybridize

both with other domesticated subspecies and with wild relatives (of which there are about twenty [11]). The high genetic diversity within *O. sativa* is likely due to introgression with wild relatives both during domestication and upon the dispersal of *O. sativa* into new environments and sympatry with new relatives [13].

In an attempt to categorize the varieties of Asian rice, [14] performed isozyme polymorphism analysis (11 enzymes, 20 loci, 79 alleles) on 25,519 rice varieties. The vast majority of the rice varieties fell into the group indica (73.45%) or japonica (23.02%). Japonica rice cultivars were likely domesticated first from wild rice populations in southern China, and indica cultivars later developed through hybridization of ancient japonica with new wild rice populations in south and southeast Asia [15? ]. However, about 1.26% belonged to four other groups (and 2.27% were indeterminate). Some of the varieties from these other four groups most likely arose from hybrids of cultivated rice and *O. rufipogon*. Introgression of alleles between cultivated rices, as well as between cultivated and wild rices, is common [9, 13, 16] and natural introgression from wild to domesticate is suspected ([16] calls for research into this possibility). These rices often have blight resistance gene, likely imparted from the wild relative parent. Several resistance genes (grassy stunt virus, bacterial blight, brown planthopper, blast) are known to have been introgressed from wild relatives into *O. sativa* by researchers [17, 18]. Some aus and rayada varieties retain characteristics of the putative wild parent. Aus exhibits a sprawling growth pattern and easy-threshing grain. Rayada is adapted to areas with longer periods of flooding, with heightened seed dormancy and photoperiod sensitivity (permitting harvest during times without standing water). Aswina variety, adapted to growth in deepwater conditions with a short period of flooding, seems to be the result of hybridization between *O. sativa* and deepwater-adapted *O. rufipogon* populations.

In addition to investigative experiments, gene flow from wild relatives has been utilized to produce agronomic rice varieties. Yatsen No. 1, for example, showed resistance to pests and diseases and adapted well to environmental conditions [19]. Several lines were derived from Yatsen No. 1, and went on to be utilized extensively in parts of China.

### 3. Barley:

Domesticated and wild barleys belong to the same species, *Hordeum vulgare*, and are capable of producing viable offspring via hybridization [20]. Barley (*Hordeum vulgare* subsp. *vulgare*) is believed to have been domesticated at least twice, once from wild subsp. *spontaneum* in the Fertile Crescent and once from subsp. *spontaneum* var. *agriocrithon* in Tibet roughly 10,000 years ago [9, 21, 22, 23, 24]. However, many of the details of barley domestication are still disputed.

There has been little genetic investigation into spontaneous barley/*spontaneum* hybrids [25]. Barley/*spontaneum* hybrids are fertile, and morphologically intermediate (putatively hybrid) barleys are found when wild and domesticated barleys are grown in sympatry, but hybrids of other wild relatives generally exhibit greatly diminished fertility [25, 26]. Even when the two are not grown immediately adjacent to one another, introgression from wild to domesticate has been shown to happen over distances of more than a kilometer [27].

The barley domestication process has reduced the number of alleles in the domesticate to only 40% of that found in wild barley, though there remains a great deal of phenotypic diversity among the wild barleys [28]. The authors of [?] used STRUCTURE to look for patterns of introgression from wild relatives in a dataset of 803 landraces, and found a high amount of variability in the amount of contribution from wild relatives, as well as its location in the genome, within barley populations. This is indicative of contribution from numerous wild

populations. Furthermore, the authors found that wild introgression contribution is generally greatest from geographically-proximate populations, and that introgressed regions might be combined from geographically-separate wild populations. Low linkage disequilibrium and small blocks of identity by state indicate that these introgressed regions are old, perhaps dating back to the beginning of barley domestication. As landraces and nearby wild relatives share similar genomic sequences, the introgressed regions that are exclusive to that landrace are more likely to contain adaptive alleles. Such alleles were not identified specifically, though wild-domesticate breeding experiments have shown that wild barleys have alleles for several important agronomic phenotypes, including powdery mildew resistance, brittleness, flowering time, plant height, lodging, and yield [29, 30? ].

#### 4. Sunflower:

The common sunflower (*Helianthus annuus*) shows evidence of domestication in eastern United States [31, 32] with additional evidence of a possible second origin of domestication in Mexico [33]. Pre-Columbian *H. annuus* distribution spanned much of the Great Plains, from what is now north-central Texas up to and through Montana and North Dakota (see figure 1 of [? ]).

Domesticated sunflower has long lived in sympatry with wild relatives like *H. petiolaris* and *H. bolanderi* and forms stable hybrid populations [34, 35, 36]. Many wild sunflowers are locally-adapted, and weedy hybrid populations share these adaptations [37]. However, the most striking example of adaptive introgression within *Helianthus* is that of the cucumberleaf sunflower, *H. debilis* ssp. *cucumerifolius*. 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 [38]. These introgressed hybrids formed a new lineage of sunflower (*H. annuus* ssp. *texanus*, *H. a. texanus* hereafter) which displays *H. debilis*-like traits adaptive to south-central Texas climate and ecology. These adaptive *debilis*-like traits include resistance to herbivorous pests and an increased branching plant architecture, as well as higher overall fitness than *H. annuus* (as measured by higher seed production [? ]). Although *H. annuus* and *H. a. texanus* are interfertile, *H. a. texanus* displays persistent phenotypic differences from *H. annuus* [? ].

The genome of the common sunflower has been greatly influenced by introgression from wild relatives, due to both natural outcrossing events and concerted breeding efforts in crop improvement. *Helianthus* has several genes for downy mildew resistance, and each imparts resistance to one or more races of *Plasmopara halstedii*, one of the most agronomically important diseases in sunflower cultivation [39]. 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* [40]. 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 [41] while others (Pl1-Pl11) are effective for one or more types [42]. Silverleaf sunflower has also been the focus of drought resistance breeding efforts [43] and *Phomopsis* resistance breeding efforts [44]. *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 [45].

Recent investigations into the history of *Helianthus* introgression have implemented genomic methods. [? ] analyzed transcriptome sequence variation on cultivated and wild *H. annuus*, *H. petiolaris*, and *H. argophyllus*. Using STRUCTURE, these authors found that introgressions from wild relatives exist on every chromosome in at least one modern line, covering over

Table 1: Programs for identifying genomic regions under introgression

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

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.

## Re-evaluating concepts of domestication

A framework in which crops were domesticated from a single population or even a single species is, in several instances, an oversimplification. An history of introgression during diffusion appears to be the rule for crops rather than the exception. Theory suggests that colonizing species will overwhelmingly be recipients of introgression from locally-adapted native species [? ]. Crops, given their frequent history of diffusion from defined centers of origin, are therefore potential recipients of adaptive introgression.

With this in mind, certain aspects of crop evolution must be re-evaluated:

\* Estimates of the initial domestication bottleneck may be skewed when introgression is not considered. Chromosomal regions experiencing introgression may have an altered effective population size ( $N_e$ ) relative to non-introgressed regions depending on diversity within the donor taxon. For example, introgression from wild taxa with historically high  $N_e$  will lead to underestimates of the strength of the domestication bottleneck. Conversely, if the donor population has a relatively

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	[46]
Asian Rice ( <i>Oryza sativa</i> )	<i>O. rufipogon</i>	X	X	X	[? ]
Barley ( <i>Hordeum vulgare</i> )	<i>H. v.</i> subsp. <i>spontaneum</i>	X	X	X	[? ]
Sunflower ( <i>Helianthus annuus</i> )	<i>H. argophyllus</i> , <i>H. bolanderi</i> , <i>H. debilis</i> , <i>H. petiolaris</i>	X	X	X	[? ]
Cassava ( <i>Manihot esculenta</i> )	<i>M. glaziovii</i>	X	X	X	[? ]
Potato ( <i>Solanum tuberosum</i> )	many	X	X	X	[47]
Tomato ( <i>Solanum lycopersicum</i> )	<i>S. pimpinellifolium</i>	X	X	X	[48]
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		[? ]
Soybeans ( <i>Glycine max</i> )	<i>G. soja</i>	X	X		[49]
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		[50]
Grapes ( <i>Vitis vinifera</i> subsp. <i>vinifera</i> )	<i>V. v.</i> subsp. <i>sylvestris</i>	X	X		[? ]
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		[51]
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		[52]
Apple ( <i>Malus domestica</i> )	<i>M. sylvestris</i> , <i>M. orientalis</i> , <i>M. baccata</i> , <i>M. sieversii</i>	X	X		[? ]

small effective population size, the opposite bias may be imposed upon bottleneck estimations. In most cases, the effective population size of the domesticated crop will be lower than that of the wild progenitor and wild relative populations.

\* Estimates of the timing of domestication based on levels of sequence divergence may be affected when introgressed haplotypes are included. The directionality of this effect is likewise dependent on  $N_e$  of the donor population.

\* Loci under selection during domestication are often identified based on signatures of substantially-reduced nucleotide diversity in the domesticated taxon relative to the wild progenitor and high allele frequency differentiation between these taxa. Introgression may alter these signatures and confound detection of domestication loci.

\* When highly-introgressed or hybrid populations are selected for domestication (as in the case of potato and tomato), identification of original progenitor(s) and domestication centers is difficult. Determining whether a crop Crops arising from highly-interfertile complexes of wild relatives

## Future studies in crop-wild introgression

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

Additional study of introgression in agroecosystems could lead to advances in both basic and



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

## Conclusions

The study of crop domestication has been revolutionized by the advent and application of genomic tools. The genomes of crops and their wild relatives tell a story of give-and-take that extends well beyond the initial stages of domestication. Likewise, population genetic theory reinforces the proclivity of wild relatives to provide advantageous, locally-adapted alleles to crops as they disperse beyond their domestication centers into new geographies with new ecological pressures and niches.

## References

- [1] 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.
- [2] N. Barton, “The role of hybridization in evolution,” *Molecular Ecology*, vol. 10, no. 3, pp. 551–568, 2001.
- [3] 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.
- [4] 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.
- [5] D. Zamir, “Improving plant breeding with exotic genetic libraries,” *Nature reviews genetics*, vol. 2, no. 12, pp. 983–989, 2001.
- [6] S. D. Tanksley and S. R. McCouch, “Seed banks and molecular maps: unlocking genetic potential from the wild,” *Science*, vol. 277, no. 5329, pp. 1063–1066, 1997.
- [7] R. Hajjar and T. Hodgkin, “The use of wild relatives in crop improvement: a survey of developments over the last 20 years,” *Euphytica*, vol. 156, no. 1-2, pp. 1–13, 2007.
- [8] 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.
- [9] H.-I. Oka, *Origin of cultivated rice*. Elsevier, 2012.
- [10] 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.
- [11]

- [12] 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.
- [13] 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.
- [14] G. Khush, D. Brar, P. Virk, S. Tang, S. Malik, G. Busto, Y. Lee, R. McNally, L. Trinh, Y. Jiang, *et al.*, “Classifying rice germplasm by isozyme polymorphism and origin of cultivated rice,” *IRRI Discuss Pap*, vol. 46, p. 279, 2003.
- [15] J. P. Londo, Y.-C. Chiang, K.-H. Hung, T.-Y. Chiang, and B. A. Schaal, “Phylogeography of asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 25, pp. 9578–9583, 2006.
- [16] 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.
- [17] D. Brar and G. Khush, “Alien introgression in rice,” in *Oryza: From Molecule to Plant*, pp. 35–47, Springer, 1997.
- [18] G. S. KHUSH and K. Ling, “Inheritance of resistance to grassy stunt virus and its vector in rice,” *Journal of Heredity*, vol. 65, no. 3, pp. 135–136, 1974.
- [19] Y. Ting, “Wild rice of kwangtung and new variety bred from the hybrids of wild rice with cultivated rice,” *Coll. Agric. Sun Yatsen Univ., Agron. Bull*, vol. 3, pp. 1–22, 1933.
- [20] R. Von Bothmer, N. Jacobsen, C. Baden, R. Jørgensen, and I. Linde-Laursen, “An ecographical study of genus *Hordeum*,” 1995.
- [21] R. Takahashi, “The origin and evolution of cultivated barley,” *Advances in genetics*, vol. 7, pp. 227–266, 1955.
- [22] 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.
- [23] 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.
- [24] G. Haberer and K. F. Mayer, “Barley: From brittle to stable harvest,” *Cell*, vol. 162, no. 3, pp. 469–471, 2015.
- [25] N. C. Ellstrand, *Dangerous liaisons?: when cultivated plants mate with their wild relatives*. JHU Press, 2003.
- [26] J. R. Harlan, “The living fields,” *Our agricultural heritage.-271 S*, 1995.

- [27] 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.
- [28] R. Ellis, B. Forster, D. Robinson, L. Handley, D. Gordon, J. Russell, and W. Powell, “Wild barley: a source of genes for crop improvement in the 21st century?,” *Journal of experimental botany*, vol. 51, no. 342, pp. 9–17, 2000.
- [29] M. Von Korff, H. Wang, J. Léon, and K. Pillen, “Ab-ql 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] A. Schwarzbach and L. Rieseberg, “Likely multiple origins of a diploid hybrid sunflower species,” *Molecular Ecology*, vol. 11, no. 9, pp. 1703–1715, 2002.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] C. B. Heiser Jr, “Hybridization in the annual sunflowers: *Helianthus annuus*  $\times$  *h. debilis* var. *cucumerifolius*,” *Evolution*, pp. 42–51, 1951.
- [39] 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.
- [40] 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.
- [41] 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.

- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] D. Zohary, J. R. Harlan, and A. Vardi, “The wild diploid progenitors of wheat and their breeding value,” *Euphytica*, vol. 18, no. 1, pp. 58–65, 1969.

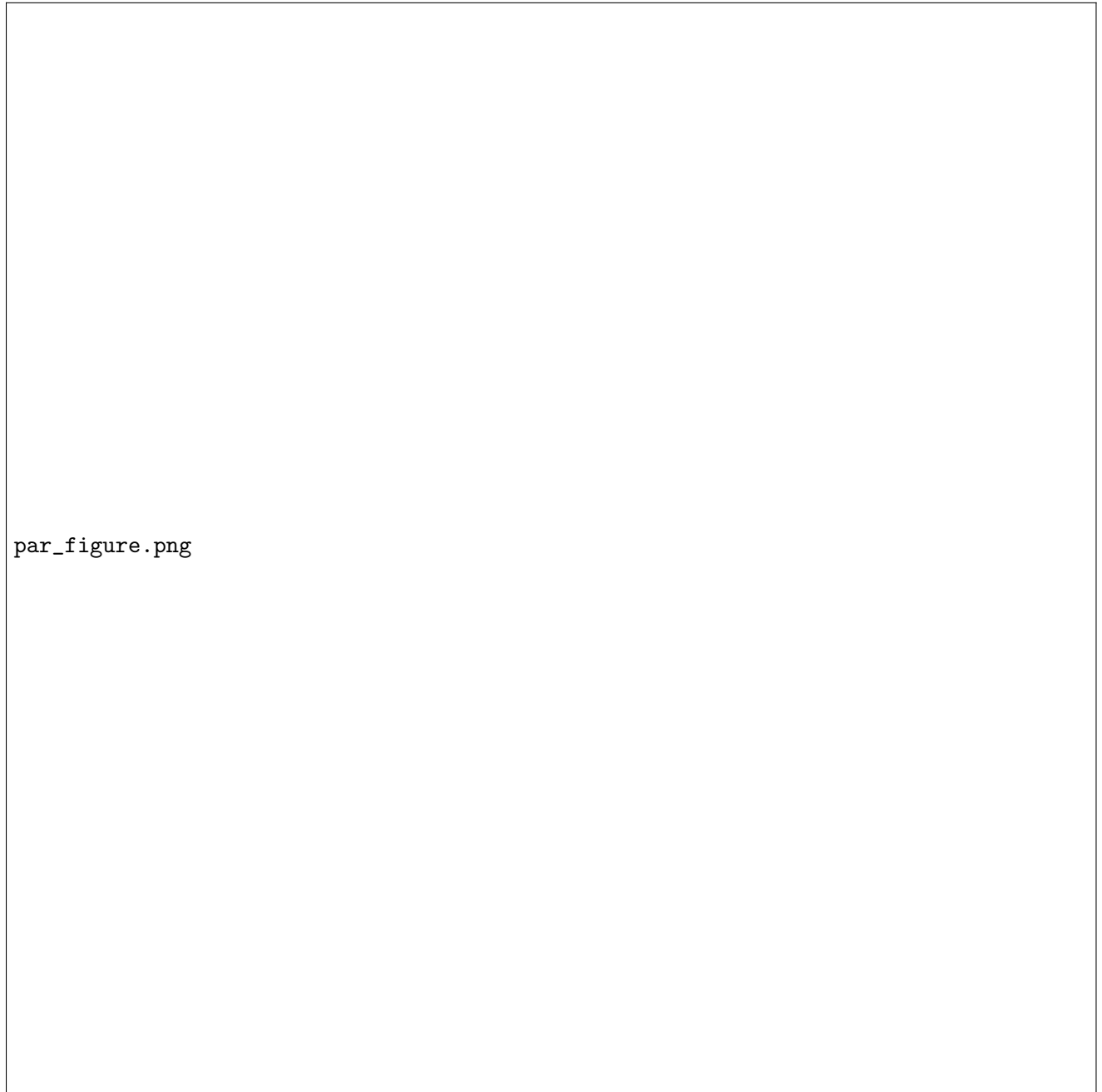


Figure 1: Map of the natural ranges of wild relatives of four domesticated crops, overlayed with average annual temperature.

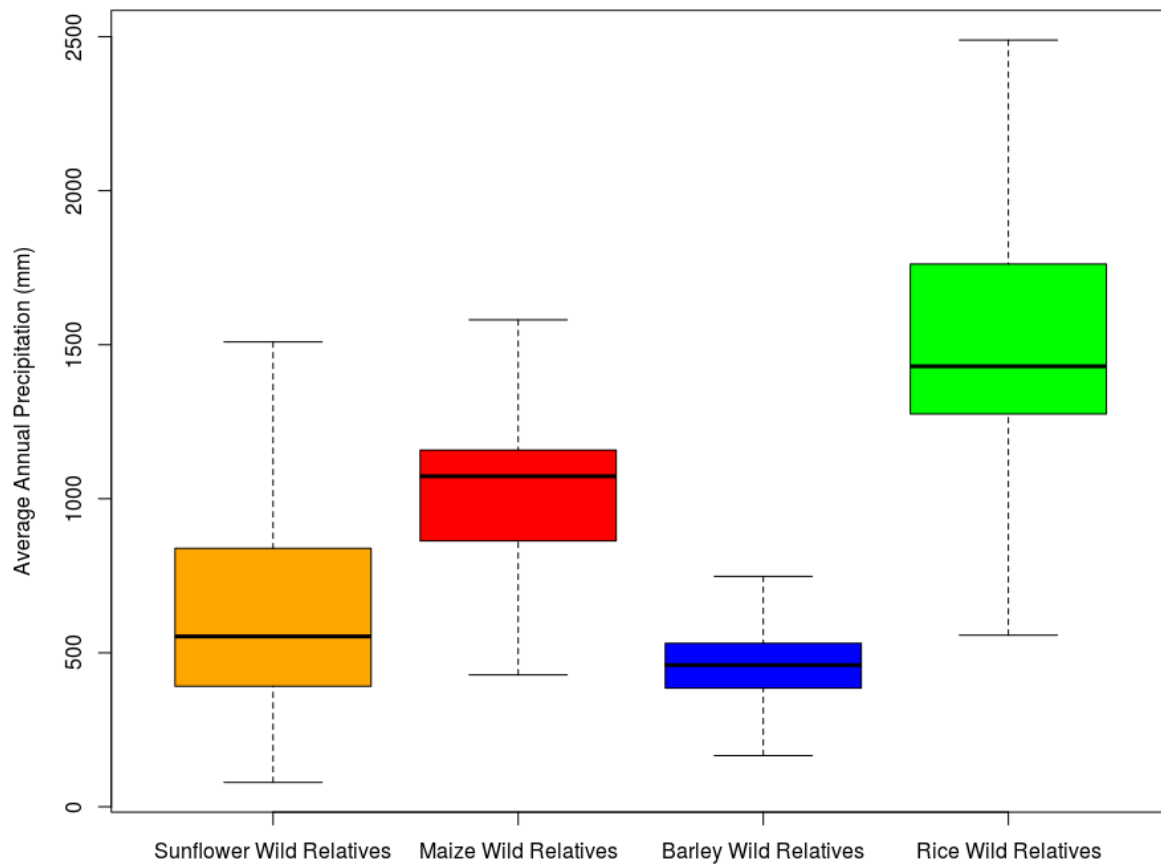
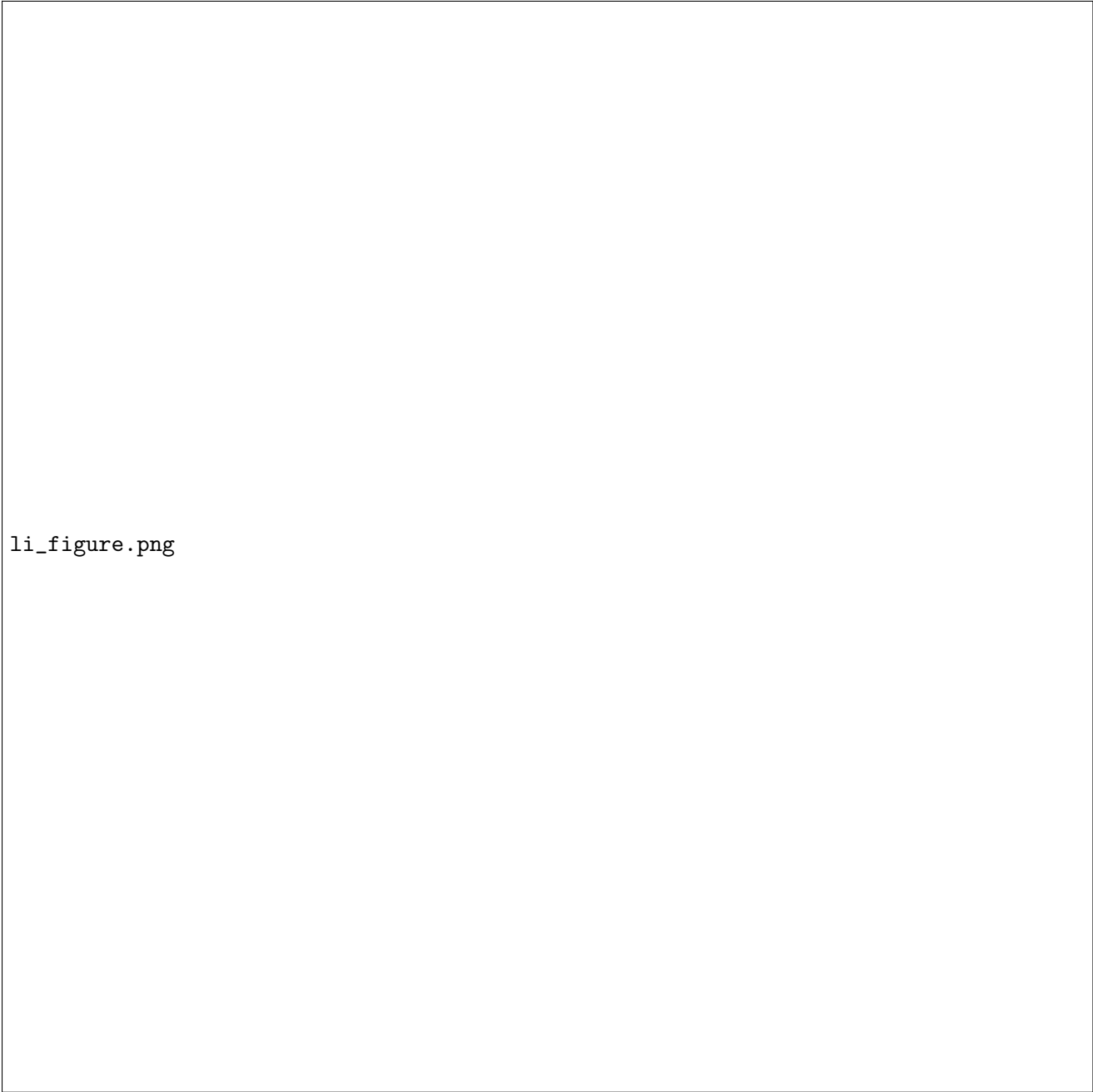


Figure 2: The distribution of average annual temperature experienced in the geographic home ranges of wild relatives interfertile with four crops



li\_figure.png

Figure 3: Li's caption here.