

# 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* **HEAD** However, archaeological and genetic evidence are beginning to reveal that, in many cases, domestication has been temporally protracted and more diffuse [1, 2, 3, 4]. *===== the paper* "Domestication rates in wild-type wheats and barley under primitive cultivation" states that domestication could have taken place in as little as 20 or 30 years. *I'm still reviewing other papers to find the best source(s) here.* 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 purugganan2011archaeological suggests that the rate of phenotypic change in domestication is slower than expected, nearly that of natural selection in wild populations. See the first paragraph of the Discussion of that paper for a list of older papers supporting rapid domestication. Also, which Gaut grape paper are you referring to?* 581e3099c7cdd17ac1f83e12d6fc4b38aacf6384 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 [5, 6]. 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. To date, literature on crop-wild gene flow has focused on the risk of introgression of transgenes from domesticated crops into wild relatives (for a review, [7]) and on modern plant breeding efforts to introgress desired traits from wild relatives (for a review, [8, 9, 10]). *cite one paper here...the best, most recent and comprehensive review* The history of natural introgression of wild alleles into domesticated crops over evolutionary timescales has received considerably less attention. However, 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 [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 ( $\delta a \delta i$ ) [18], isolation with migration models [19], and a series of method utilizing sequentially Markovian coalescent (PSMC, MSMC and SMC++) [20, 21, 22]). Rather, we focus on methods that explicitly identify introgressed genomic segments based on patterns of nucleotide/haplotype diversity, level of 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 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.

*abrupt transition here to Gmin equation...can you smooth this into the text? I am not much a fan of Gmin, as we applied this method and it is not good. I am wondering if we just give the equation of RNDmin, in order to focus readers attention on the later.*

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

where  $d_{min}$  is the minimum sequence distance between haplotypes in species X and Y;  $d_{out}$  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 two source populations (see details in [23]).

Second, local ancestry deconvolution (also known as chromosome painting) assigns genomic regions to various source populations [24]. One category of such methods evaluates genetic ancestry of genomes in question as a mixture of reference genomes using the coalescent with recombination based on hidden Markov model (HMM) across different parts of genome, such as HAPMIX [12]. Another category of such methods clusters the admixed populations with the reference samples by breaking genomes into sliding windows, such as PCAdmix [25] and LAMP [26]. Such methods typically require phased haplotypes as input (but see [27]).

Third, the ABBA-BABA statistic (also known as the D-statistic) and its derivatives are widely applied to introgression detection and make inferences based on genomic patterns of shared derived variants between populations or species in a phylogenetic context. D-statistic evaluates the possibility of introgression across the genome between species/populations by estimating the asymmetry in the frequencies of the two non-concordant gene trees in a four-taxon phylogeny. Elaborations of

the D-statistic include  $\hat{f}_d$  [11] and the five-taxon D-statistic [14]. The former uses allele frequencies from each population/species and thus has stronger power to detect genomic regions under introgression, and the latter detects introgression based on the localized phylogenetic pattern but distinct at identifying the direction of introgression.

Fourth, approaches that assign genome-wide genetic ancestry based on the popular software STRUCTURE [28] are another category to detect introgression. For example, fineSTRUCTURE [13] calculates the probability that genomic regions belong to genetic groups by patterns of haplotype similarity. GLOBETROTTER [29] is a further development of fineSTRUCTURE that allows for identification of an unsampled source population and for dating of admixture events. *not sure if we should have this category. In the Schraiber 2015 review paper, he put fineSTRUCTURE and GLOBETROTTER in the chromosome painting category.*

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 [30]); introgression has conferred Müllerian mimicry *I would explain a bit more here...wing coloration loci, protects against predation...* across butterfly species [31]; introgression has spread insecticide resistance across mosquito species [32], and introgression across *Mimulus* (*i.e.*, monkeyflower) species has resulted in adaptation to pollinator preference and contributed to speciation [33].

## 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*) approximately 9,000 years ago in the lowlands of the Balsas River Valley in Mexico [34]. From this domestication center, maize spread into the highlands of the Mexican Central Plateau, where it came into sympatry with wild *mexicana*. Introgression from *mexicana* to maize in the highlands of Mexico has been reported based on evidence from both morphological data [35, 36, 37] and molecular analyses [38, 39, 40, 41, 42]. However, [43] 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 showed evidence of *mexicana* introgression based on local ancestry deconvolution using HAPMIX and results from 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 [36], traits known to be important in adaptation to 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 settings 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 additional regions. A recent study [3] 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. *on the fence about including deleterious allele results from introgressed regions*

## 2. Asian Rice:

The story of the domestication of asian rice is still under debate. A recent model utilizing molecular clocks points to single domestication of asian rice that occurred 8,200-13,500 BP in the Yangtze Basin in China from the wild species *O. rufipogon*, with later divergence of the two prominent varietal groups indica and japonica lines [44]. Sequence data from indica, japonica, and wild *O. rufipogon* accessions points at japonica domestication in the Pearl Valley in China, and indica being formed from hybridization between japonica and rufipogon in Southern and South-eastern Asia [? ]. On the other hand, genetic and archaeobotanical evidence points towards independent domestications of the indica and japonica subspecies in the Yangtze Basin and the Ganges plains in India [45].

The authors of [? ] review the evidence for both hypotheses and support the single domestication event hypothesis, contributing the contrary evidence to the effects of a history of introgression, selection (primarily that of the non-shattering trait allele sh4), and diversification. In their defense of the single-domestication hypothesis, the authors review the case that, during the spread of domesticated rice into new environments, gene flow from wild relatives would have imparted adaptive traits via introgression. The wild relatives rufipogon and nivara are both phenotypically diverse and show evidence of local adaptation. Rufipogon reproduction can be either primarily sexual or vegetative. Whereas rufipogon is adapted to forested wetland environments, nivara is adapted to dryer conditions and has life cycle adaptations to survive grazing pressure. Likewise, cultivated rice varieties display phenotypic differences important for fitness. In particular, two of the cultivated rice deepwater varieties (rayada and ashwina) are said to be selected for the environment along the Ganges river. Gene flow between these wild and cultivated rices is common, producing nuisance weedy hybrids in when the two are grown in sympatry. However, during early domestication, introgression may have been even more prevalent 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. 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. Introgression into indica is expected to be more likely than japonica, due to the higher degree of sympatry between indica and its wild relatives. The high genetic diversity within domesticated rice is likely due to introgression from wild relatives both within the domestication center(s) and in new environments where rice has dispersed [46].

## 3. Barley:

Domesticated and wild barley belong to the same species, *Hordeum vulgare*, and are capable of producing viable offspring via hybridization [47]. Barley (*Hordeum vulgare* subsp. *vulgare*) is believed to have been domesticated at least twice roughly 10,000 BP, once from wild subsp. *spontaneum* in the Fertile Crescent and once from subsp. *spontaneum* var. *agriocrithon* in Tibet [48, 49, 50, 51, 52]. However, details of barley domestication are still disputed.

There has been little genetic investigation into spontaneous barley/*spontaneum* hybrids [53]. 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 [53, 54]. 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 [55].

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 [56]. 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 [57, 58? ].

#### 4. Sunflower:

The common sunflower (*Helianthus annuus*) shows evidence of domestication in eastern United States [59, 60] with additional evidence of a possible second origin of domestication in Mexico [61]. 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 [62, 63, 64]. Many wild sunflowers are locally-adapted, and weedy hybrid populations share these adaptations [65]. 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 [66]. 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 [67]. Some of these downy mildew resistance genes were

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

found in wild relatives (including *H. argophyllus*, *H. tuberosus*, and *H. praecox*) and have been successfully bred into modern *H. annuus* [68]. 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 [69] while others (Pl1-Pl11) are effective for one or more types [70]. Silverleaf sunflower has also been the focus of drought resistance breeding efforts [71] and *Phomopsis* resistance breeding efforts [72]. *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 [73].

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 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.

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	[74]
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	[75]
Tomato ( <i>Solanum lycopersicum</i> )	<i>S. pimpinellifolium</i>	X	X	X	[76]
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		[77]
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		[78]
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		[79]
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		[80]
Apple ( <i>Malus domestica</i> )	<i>M. sylvestris</i> , <i>M. orientalis</i> , <i>M. baccata</i> , <i>M. sieversii</i>	X	X		[? ]

## 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 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] R. S. Meyer, J. Y. Choi, M. Sanches, A. Plessis, J. M. Flowers, J. Amas, K. Dorph, A. Barretto, B. Gross, D. Q. Fuller, I. K. Bimpong, M.-N. Ndjiondjop, K. M. Hazzouri, G. B. Gregorio, and M. D. Purugganan, “Domestication history and geographical adaptation inferred from a snp map of african rice,” *Nat Genet*, vol. 48, pp. 1083–1088, 09 2016.
- [3] L. Wang, T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, and M. Hufford, “The interplay of demography and selection during maize domestication and expansion,” *bioRxiv*, p. 114579, 2017.
- [4] 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.
- [5] N. Barton, “The role of hybridization in evolution,” *Molecular Ecology*, vol. 10, no. 3, pp. 551–568, 2001.



- [6] 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.
- [7] 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.
- [8] D. Zamir, “Improving plant breeding with exotic genetic libraries,” *Nature reviews genetics*, vol. 2, no. 12, pp. 983–989, 2001.
- [9] 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.
- [10] 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.
- [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] C. Churchhouse and J. Marchini, “Multiway admixture deconvolution using phased or unphased ancestral panels,” *Genetic epidemiology*, vol. 37, no. 1, pp. 1–12, 2013.
- [28] 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.
- [29] G. Hellenthal, G. B. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush, and S. Myers, “A genetic atlas of human admixture history,” *Science*, vol. 343, no. 6172, pp. 747–751, 2014.
- [30] 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.
- [31] H. G. Consortium *et al.*, “Butterfly genome reveals promiscuous exchange of mimicry adaptations among species,” *Nature*, vol. 487, no. 7405, pp. 94–98, 2012.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] H. Wilkes, “Hybridization of maize and teosinte, in mexico and guatemala and the improvement of maize,” *Economic Botany*, pp. 254–293, 1977.
- [36] 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.
- [37] J. F. Doebley, “Maize introgression into teosinte-a reappraisal,” *Annals of the Missouri Botanical Garden*, pp. 1100–1113, 1984.

- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] R. Von Bothmer, N. Jacobsen, C. Baden, R. Jørgensen, and I. Linde-Laursen, "An ecographical study of genus *hordeum*," 1995.
- [48] R. Takahashi, "The origin and evolution of cultivated barley," *Advances in genetics*, vol. 7, pp. 227–266, 1955.
- [49] 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.
- [50] H.-I. Oka, *Origin of cultivated rice*. Elsevier, 2012.
- [51] 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.

- [52] G. Haberer and K. F. Mayer, “Barley: From brittle to stable harvest,” *Cell*, vol. 162, no. 3, pp. 469–471, 2015.
- [53] N. C. Ellstrand, *Dangerous liaisons?: when cultivated plants mate with their wild relatives*. JHU Press, 2003.
- [54] J. R. Harlan, “The living fields,” *Our agricultural heritage.-271 S*, 1995.
- [55] 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.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] A. Schwarzbach and L. Rieseberg, “Likely multiple origins of a diploid hybrid sunflower species,” *Molecular Ecology*, vol. 11, no. 9, pp. 1703–1715, 2002.
- [63] 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.
- [64] 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.
- [65] 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.
- [66] C. B. Heiser Jr, “Hybridization in the annual sunflowers: *Helianthus annuus*  $\times$  *h. debilis* var. *cucumerifolius*,” *Evolution*, pp. 42–51, 1951.
- [67] 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.

- [68] 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.
- [69] 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.
- [70] 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.
- [71] 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.
- [72] 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.
- [73] 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.
- [74] 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.
- [75] 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.
- [76] 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.
- [77] 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.
- [78] 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.
- [79] 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.
- [80] 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.

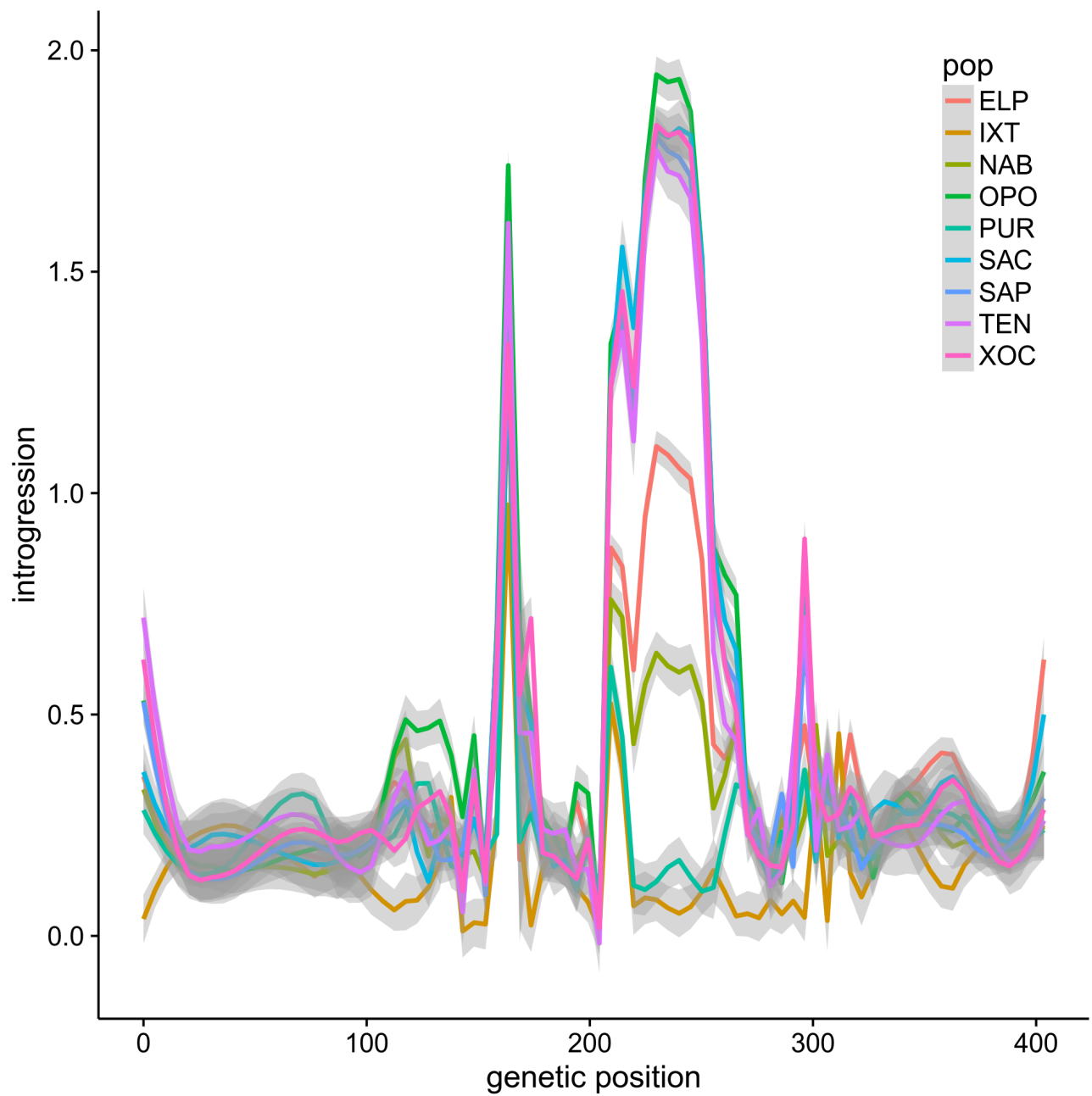


Figure 1: Li's caption here.

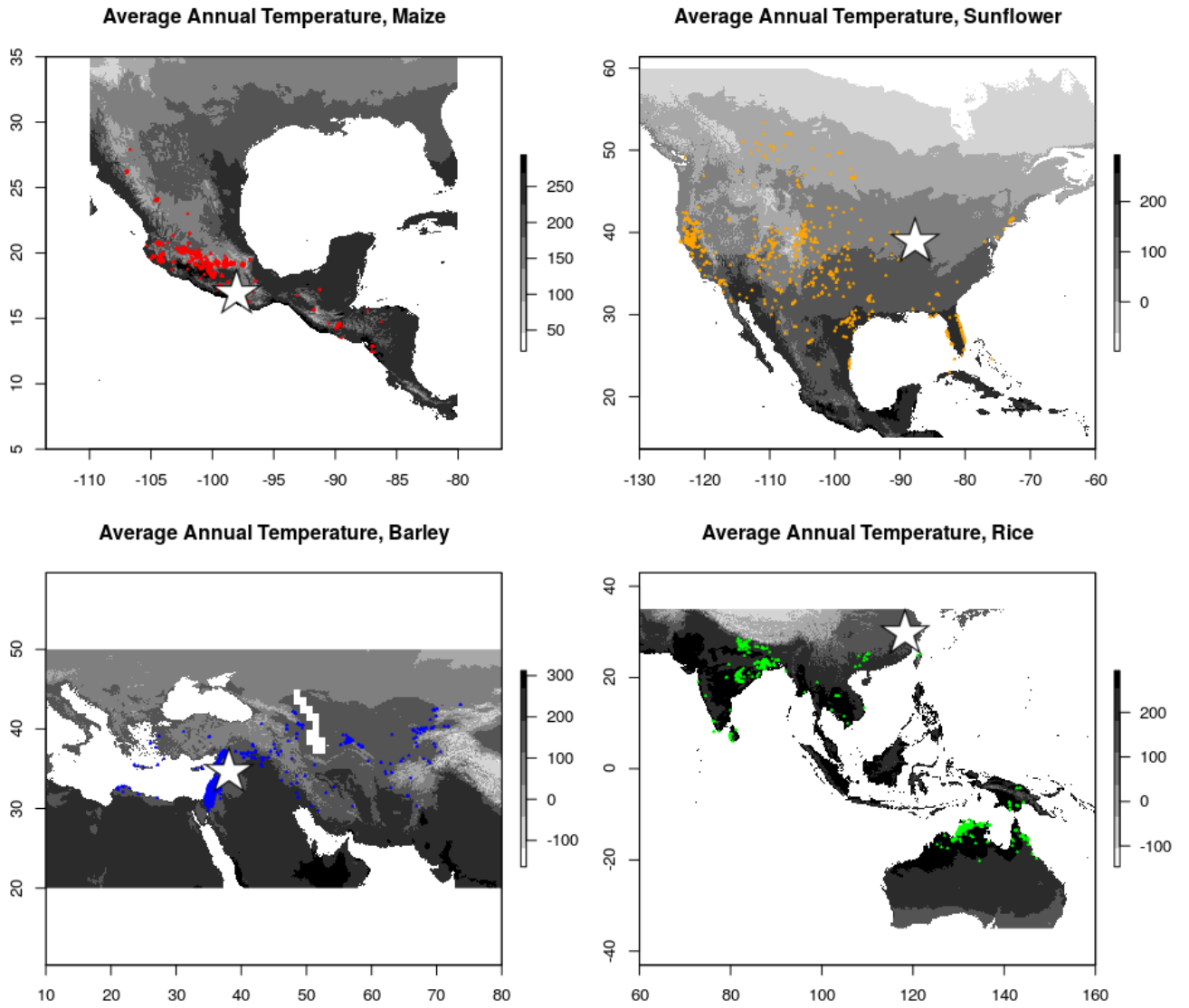


Figure 2: Map of the natural ranges of wild relatives of four domesticated crops, overlaid with average annual temperature.

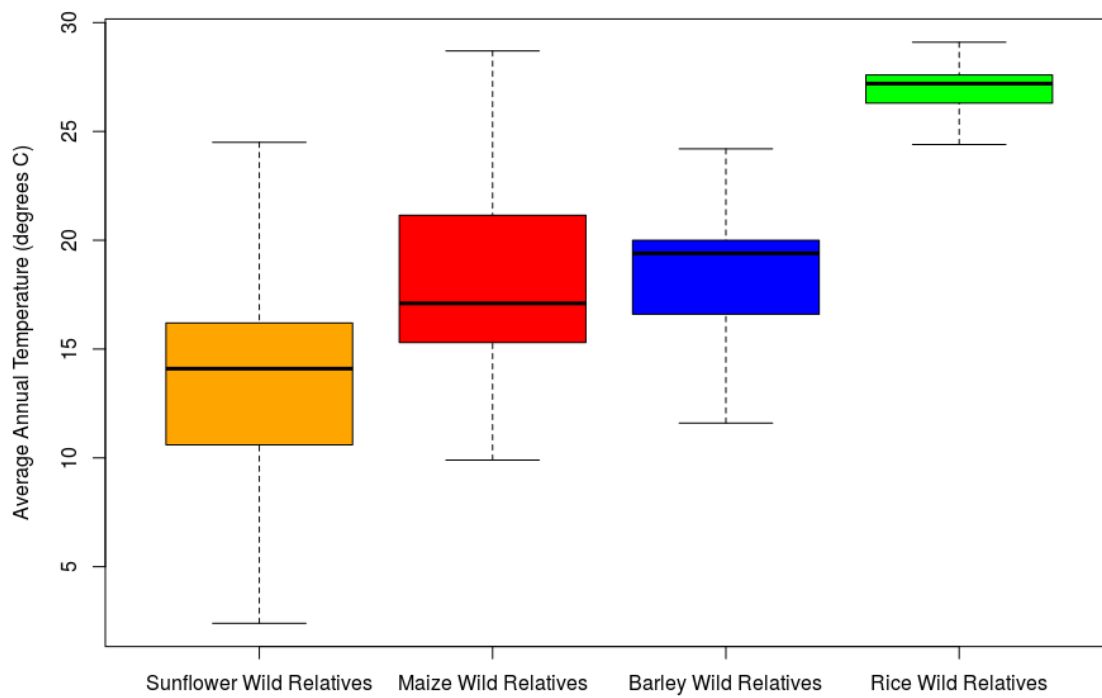


Figure 3: The distribution of average annual temperature experienced in the geographic home ranges of wild relatives interfertile with four crops