

Review: Crop adaptation through introgression from wild relatives

Authors: Garrett M. Janzen¹, Li Wang¹, and Matthew B. Hufford^{1,*}

¹Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

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

The traditional paradigm in crop domestication has been origin from a wild relative within one or more geographically-defined centers followed by expansion to the modern-day extent of cultivation. Absent from this paradigm are the effects of hybridization between a diffusing nascent crop species and closely-related, locally-adapted wild relatives outside the center of origin. New methods have recently been employed to detect genome-wide patterns of introgression in a number of species. In this review, we will: 1) briefly describe these methods 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 applications

The recent availability of genome-wide resequencing and reduced-representation genotyping (*e.g.*, GBS and RAD-Seq) data combined with new analytical methods have facilitated comprehensive study of introgression across a number of species (**Table 1**).

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 to identify introgression across taxa [1, 2, 3, 4, 5, 6].

The methods introduced here do not include those marginally estimating introgression/migration rate as a component in the complete demographic history, such as Approximate Bayesian Computation (ABC) [7], diffusion approximations for demographic inference (dadi) [8], sorts of isolation with migration model [9] and the multiple sequentially Markovian coalescent (MSMC) [10]. The collection of methods described different layers of genetic variation: diversity and divergence, haplotype variation and phylogenetic relationships. We focus here on analytical prospect and limitations and introduce only a few representatives of each category.

First, genomic regions of introgression in admixed populations were expected to show a low level of divergence from source population. F_{st} and d_{XY} as well as its derivatives, such as G_{min} [6] and RND_{min} [5], a

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

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

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

where d_{out} equals to $(d_{XO} + d_{YO})/2$, the average sequence distance between each species and the out-group (O). In addition, RND_{min} is sensitive to variable mutation rate, which reflects on the branch length to the outgroup.

These statistics were recently developed into several new statistics, such as $U_{A,B,C(w,x,y)}$ by adding the divergence between non-admixed populations (A) and source population (C) as a comparison to that between admixed populations (B) and source population (C) [11]. This statistic summarizes number of sites, where one particular allele at frequency y in source population has a frequency bigger than x in admixed population (B) and a frequency smaller than w in non-admixed population (C). Similarly, $Q95_{A,B,C(w,y)}$ just set a hard cutoff as 95th percentile of allele frequencies in panel B [11]. These two statistics were also developed into incorporating two source populations (see details in [11]).

Second, local ancestry reconvolution (also known as chromosome painting) can identify the genomic regions that come from different source populations [12]. Such method usually takes phased haplotypes as input and an exception is MULTIMIX [13], which can take both phased and unphased haplotypes as input, and unveil the genomic ancestry from multiple source populations.

Third, ABBA-BABA statistics (also known as D-statistics) and its derivatives are widely applied to search for signs of introgression in genomic patterns of shared derived variants between populations or species. Based on D-statistics, \hat{f}_d [1] and five-taxon D statistics [4] are developed to localize the 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 enabled us 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 [14], which provided an estimate of fraction of each individual's genome coming from each population. Many approaches in this respect have been developed. Two of them, fineStructure [3] and Globetrotter [15], 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 at tracing back to an unsampled source population and at dating the admixture events.

2. 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 [16]), introgression has conferred Müllerian mimicry across butterfly species ([17]; introgression has spread insecticide resistance across mosquito species [18], and introgression across *Mimulus* (*i.e.*, monkeyflower) species has resulted in adaptation to pollinator preference and contributed to speciation [19].

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 [17].

Crop adaptation through introgression

Over the last few years, several high-profile publications based on genome-wide data have documented introgression between crops and wild relatives outside putative domestication centers. A history of introgression during diffusion appears to be the rule for crops rather than the exception.

1. Theory suggests that colonizing species will overwhelmingly be recipients of introgression from locally-adapted native species [20]. Crops, given their frequent history of diffusion from defined centers of origin, are therefore potential recipients of adaptive introgression.
2. Recent empirical studies have revealed that introgression has occurred in many of the world’s most important crops (**Table 2**).
 - (a) Maize: As maize spread from a lowland center of origin into the Mexican highlands, it received introgression from a highland-adapted wild relative. Introgression occurred predominantly at loci previously shown to underlie highland adaptation traits, strongly suggesting adaptive introgression [21].
 - (b) Cassava: Substantial introgression from a closely-related wild species has been found in cassava, occurring both naturally and through a targeted breeding program to introduce disease resistance [22].
 - (c) Rice: Introgression appears to have played an important role in the history of rice. 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 [23]. Whether this wild introgression conferred local adaptation to indica has yet to be shown.
 - (d) Barley: Landraces of barley show shared ancestry with geographically-proximate populations of wild relatives, suggesting introgression but not demonstrating conclusively that this was adaptive [24].
 - (e) Olive: Hybridization between wild and domesticated olive has been prevalent throughout the Mediterranean Basin. [25].
 - (f) Sunflower: Domesticated sunflower has received substantial introgression from wild relatives through breeding efforts to reintroduce a branching morphology. Natural introgression appears prominent in this system as well [26].
3. Predictions regarding the likelihood of adaptive introgression in additional crops can be made through comparison of their centers of origin to both their current extent of cultivation and the distribution of wild relatives across environmental gradients (**Figure 2**).

Table 2: Overview of literature evaluating potential for crop-wild introgression in the world’s 15 most important crops including a column of references and a column summarizing major findings.

Figure 2: Multi-panel figure showing maps of crop centers of domestication, distributions of wild relatives, and current extent of cultivation across gradients of (a) temperature, (b) precipitation, and (c) elevation. This figure will illustrate the adaptive potential of wild-to-crop introgression.

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. A demography incorporating introgression from additional

sources appears to be more correct for many crops. With this in mind, certain aspects of crop evolution must be re-evaluated:

1. 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.
2. Estimates of the timing of domestication based on levels of sequence divergence may be affected when introgressed haplotypes are included.
3. 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.

Future studies in crop-wild introgression

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

1. Basic questions:
 - (a) To what extent does the level of introgression across taxa depend on divergence time between donor and recipient taxa?
 - (b) 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?
 - (c) Can colonizing species served as bridges for gene flow between previously allopatric taxa?
2. Applications:
 - (a) Loci underlying the domesticated phenotype which may be beneficial targets for crop improvement can be more clearly identified by removing the confounding population genetic signal of introgression.
 - (b) Adaptive introgression that is clearly tied to a specific environment may include beneficial alleles that can be utilized in crop breeding.

References

- [1] S. H. Martin, J. W. Davey, and C. D. Jiggins, "Evaluating the use of abababab statistics to locate introgressed loci," *Molecular Biology and Evolution*, vol. 32, no. 1, pp. 244–257, 2015.
- [2] 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.

- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] M. A. Beaumont, W. Zhang, and D. J. Balding, “Approximate bayesian computation in population genetics,” *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] J. G. Schraiber and J. M. Akey, “Methods and models for unravelling human evolutionary history,” *Nature Reviews Genetics*, 2015.
- [13] C. Churchhouse and J. Marchini, “Multiway admixture deconvolution using phased or unphased ancestral panels,” *Genetic epidemiology*, vol. 37, no. 1, pp. 1–12, 2013.
- [14] 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.
- [15] 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.
- [16] F. Racimo, S. Sankararaman, R. Nielsen, and E. Huerta-Sanchez, “Evidence for archaic adaptive introgression in humans,” *Nat Rev Genet*, vol. 16, no. 6, pp. 359–371, 2015.
- [17] T. H. G. Consortium, “Butterfly genome reveals promiscuous exchange of mimicry adaptations among species,” *Nature*, vol. 487, no. 7405, pp. 94–98, 2012.
- [18] 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.

- [19] 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.
- [20] M. Currat, M. Ruedi, R. J. Petit, and L. Excoffier, “The hidden side of invasions: massive introgression by local genes,” *Evolution*, vol. 62, no. 8, pp. 1908–1920, 2008.
- [21] M. Hufford, P. Lubinsky, T. Pyhäjärvi, M. Devengeno, N. Ellstrand, and J. Ross-Ibarra, “The genomic signature of crop-wild introgression in maize,” *PLoS Genetics*, vol. 9, no. 5, p. e1003477, 2013.
- [22] 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, P. Nauluvula, V. Lebot, J. Ndunguru, G. Mkamilo, R. S. Bart, T. L. Setter, R. M. Gleadow, P. Kulakow, M. E. Ferguson, S. Rounsley, and D. S. Rokhsar, “Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity,” *Nat Biotech*, vol. 34, no. 5, pp. 562–570, 2016.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.