

Crop adaptation with gene flow from wild relatives

author1¹, author2¹, and author3^{1,2}

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

²Corresponding Author: Matthew B. Hufford; 339A Bessey Hall, Iowa State University, Ames, IA, USA; phone: 1-515-294-8511; email: mhufford@iastate.edu

1 Perspectives and key issues

- ways to detect gene flow

compromise between sample size of the source population and the SNP density across the genome

- ways to detect selection

The methods to detect both gene flow and selection are summarized in Racimo et al. (2015) (archaic adaptive introgression in humans), probably we should not write the first two items.

- What is the proportion of immigrants showing the signatures of adaptation across the genome
- Most studies targeted quite long introgression region (in Mb); it is difficult to pinpoint the loci which are introgressed and caused the linkage.
- The extent to which adapted immigrants are physically linked on chromosomes and reside within structural features restricting recombination, such as inversions
- the importance of genome structures on local adaptation

An antagonism exists between selection for combination of locally adapted genes, and migration and recombination tending to break them down. Therefore, genome structures, such as inversion and centromeres, which constrains the recombination, can boost the effectiveness of selection by extending the linkage disequilibrium.

2 Theory of adaptative introgression

3 Empirical Data

3.1 maize

Maize is a classical model in crop species to study the issue of adaptive introgression from wild relatives, as it is sympatric with a lot of its wild relatives with a range of different divergence time [?]. *Zea mays* ssp. *mexicana* (referred to *mexicana* hereafter) and Mexican highland maize is a well-studied pair. Introgression between *mexicana* and Mexican highland maize has been

reported based on evidence from both morphological data [?, ?, ?] and molecular analysis [?, ?, ?, ?, ?]. However, hufford2013 for the first time revealed the evidence of adaptive introgression from *mexicana* to Mexican highland maize. The authors identified nine genomic regions, which showed evidence of introgression from *mexicana* to maize in both the HAMPMIX and the linkage model of STRUCTURE analyses with over seven sympatric population pairs among the total nine pairs sampled. Among the nine regions, three spans 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 chromosome 4 (associated with QTL controlling pigment density and macrohairs) and 9 (overlapped with QTL for macrohairs) exhibited more macrohairs and greater pigmentation under the highland environmental settings than the populations with absence of introgression from *mexicana*. To put it together, 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 and theoretical support for parallel highland adaptation in the Mexican and South American highland regions, which was explained partly owing to the difference in the 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 Mexican highland population were more likely also showing signatures of local adaptation in *mexicana* and *pulvillumis* populations than those South American highland population. To sum up, the adaptive introgression from wild relatives may play a significant role in patterning the genetic differences of maize highland populations.

However, has the introgression from *mexicana* spread to multiple highland populations? How does the adaptation 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) martin2015 were calculated to locate the introgressed loci. Figure 1 showed the loess regression of \hat{f}_d in 10kb nonoverlapping windows across two chromosomes (chr4 and chr5) in multiple comparisons. On chromosome 4 (Figure 1), Mexican highland (MexHigh) and Guatemalan highland (GuaHigh) exhibited strong evidence of introgression from *mexicana*, and the peaks of distribution corresponded to the region identified in Hufford et al. (2013) hufford2013. 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 South-western US Highland (SW_US), but not the other two. More details on the other chromosomes could be referred to [?]. Overall, both analyses revealed that gene flow exist from *mexicana* 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. Even though it is clear that the absence of introgression from wild relatives provided less genetic resource for the highland adaptation and thus made the highland adaptation unique in the Andes, it is yet unknown that out of reach of wild relatives is a reason for reduced fitness in the Andes population. And the question whether convergent evolution occurs between populations with and without introgression from wild relatives is a key topic for future studies in maize.

3.2 barley

3.3 wheat

3.4 sunflower

4 Concluding remarks and future directions

[width=.45]FigsAndTables/fdTrackChr4AllPop [width=.45]FigsAndTables/fdTrackChr5AllPop

Figure 1: Loess regression of \hat{f}_d in 10kb nonoverlapping windows across the genome in five independent comparisons: MexLowMexHighmexicanaoutgroup, MexLowGuaHighmexicanaoutgroup, MexLowSW_USmexicanaoutgroup, MexLowAndesmexicanaoutgroup, MexLowSA_{Lowmexicana}-outgroup. *A* : Chromosome4; *B* : Chromosome5.