



**HAL**  
open science

## Crop domestication as a step towards reproductive isolation

M. I. Tenaillon, Ewen Burban, Stella Huynh, Arthur Wojcik, Anne-Céline Thuillet, Domenica Manicacci, Pierre R. Gérard, Karine Alix, Harry Belcram, A. Cornille, et al.

► **To cite this version:**

M. I. Tenaillon, Ewen Burban, Stella Huynh, Arthur Wojcik, Anne-Céline Thuillet, et al.. Crop domestication as a step towards reproductive isolation. *American Journal of Botany*, 2023, 110 (7), pp.e16173. 10.1002/ajb2.16173 . hal-04088575

**HAL Id: hal-04088575**

**<https://hal.science/hal-04088575v1>**

Submitted on 23 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title: Crop domestication as a step towards reproductive isolation

**Article type:** On the Nature of Things, OTNOT

**Authors:** M.I. Tenaillon<sup>\*1</sup>, E. Burban<sup>2</sup>, S. Huynh<sup>3</sup>, A. Wojcik<sup>1</sup>, A-C Thuillet<sup>3</sup>, D. Manicacci<sup>1</sup>, P. R. Gérard<sup>1</sup>, K. Alix<sup>1</sup>, H. Belcram<sup>1</sup>, A. Cornille<sup>1</sup>, M. Brault<sup>4</sup>, R. Stevens<sup>4</sup>, J. Lagnel<sup>4</sup>, C. Dogimont<sup>\*4</sup>, Y. Vigouroux<sup>\*3</sup>, S. Glémin<sup>\*2,5</sup>.

\*: Co-corresponding authors

**Affiliations:**

<sup>1</sup>: Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 12 route 128, Gif-sur-Yvette, France.

<sup>2</sup>: Laboratoire ECOBIO\_UMR 6553 UR1-CNRS, Campus de Beaulieu, Rennes, France.

<sup>3</sup>: UMR DIADE, Univ Montpellier, IRD, CIRAD, 911 avenue Agropolis, Montpellier, France.

<sup>4</sup>: GAFL - Génétique et Amélioration des Fruits et Légumes, INRAE, Allée des Chênes, Montfavet, France

<sup>5</sup>: Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Sweden.

**Keywords:** Speciation Gene flow, Adaptive introgression, Genomic divergence, Mating system, Genetic incompatibilities, Hybrid breakdown, Domestication load, Crop vulnerability.

**BACKGROUND**

Speciation, Darwin's mystery of mysteries, is a continuous process that results in genomic divergence accompanied by the gradual increment of reproductive barriers between lineages. Since the beginning of research on the genetics of speciation, sev-

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ajb2.16173.

eral questions have emerged such as: What are the genetic bases of incompatibilities? How many loci are necessary to prevent hybridization and how are they distributed along genomes? Can speciation occur despite gene flow and how common is ecological speciation? Early stages of divergence are key to understand the ecology and genetics of speciation, and semi-isolated species where hybrids can still be produced are particularly relevant. Here we argue that the recent divergence between wild and domesticated lineages is an excellent model to capture the very-first steps of reproductive barriers formation, and will bring novel insights into the speciation process.

### **WHY IS DOMESTICATION A GOOD MODEL TO STUDY SPECIATION?**

Domestication is the process of divergent selection between wild forms undergoing natural selection in their habitats, and domesticates evolving under combined natural and human-mediated selection. It has been increasingly recognized that evolution of domesticated species shows many similarities with evolution in the wild: it results primarily from changing environmental conditions and involves unconscious selection under a protracted process (Purugganan 2019) with selection intensities of the same magnitude or even smaller (Yang *et al.* 2019). Thus, domestication has been considered as a choice example to study adaptation. Here, we argue that it also offers an excellent opportunity to catch the very-first processes at work in ecological speciation, where adaptive divergence between nascent lineages triggers the onset of reproductive isolation (RI).

Allele differentiation resulting from divergent selection can be measured by  $F_{ST}$ .  $F_{ST}$  between wild and domestic pairs range between 0.05 in sweet cherry and 0.51 in Tomato (Appendix S1, and references herein; see the Supplementary Data with this article), which cover a wide range of divergence within a “grey zone of speciation” in which barriers to gene flow exist but are not complete (Roux *et al.* 2016). Interesting-

ly, within this continuum, self-fertilizing taxa display greater genetic differentiation than outcrossers (Figure 1). As mating systems are predicted to affect the speciation process, domestication also offers the opportunity to address this question (Marie-Orleach *et al.* 2022). In contrast, life span seems to have no significant effect on divergence (Figure 1), although annuals and perennials experience contrasted domestication dynamics in many respects (Gaut *et al.* 2015).

The existence of reproductive barriers between wild and domesticated plants has been repeatedly documented. Despite the occurrence of wild-cultivated gene flow, the establishment of wild alleles into domesticated populations and reciprocally – introgressions – is rare (Ellstrand *et al.* 2013). Perhaps the best documented examples come from maize, where the introgression from the *mexicana* teosinte subspecies has contributed to highland adaptation of maize landraces (Calfee *et al.* 2021); and conversely, introgression from locally-adapted maize has contributed to teosinte adaptation in Europe (Le Corre *et al.* 2020). Interestingly, introgressions in the two directions are removed by selection around domestication genes (Le Corre *et al.* 2020; Calfee *et al.* 2021). This points to a prominent role of pre- and post-zygotic genetic barriers in the divergence of wild and domesticated lineages, and some genes involved in reproductive barriers have been identified such as the *Tcb1* locus in maize that governs pollen rejection by teosinte (Lu *et al.* 2019).

Whether partial isolation between wild and domesticated forms will ultimately result in full speciation is unknown. But clearly, partial RI does occur and has contributed to the maintenance of the distinct features between wild and domesticated forms, the so-called domestication syndrome. RI therefore stands as a major component of the domestication syndrome, but has been so far largely ignored (Dempewolf *et al.* 2012). It is even possible that reinforcement played a role in the establishment

of the domestication syndrome, which involves the evolution of stronger RI due to the costs associated with producing low-fitness hybrids (Rushworth *et al.* 2022).

### THE GENETIC BASES OF REPRODUCTIVE ISOLATION

The establishment of reproductive barriers can occur through various mechanisms. Selection leading to the fixation of advantageous alleles in different environments, resulting in local adaptation, can cause hybrid offspring to have lower fitness in parental environments, which strengthens isolation as populations adapt to differing conditions. This process may contribute to RI between wild and domesticated forms, and some crops may already be considered as independently evolving lineage once human-mediated cessation of gene flow is complete. Loci involved in such adaptation, those governing domestication traits, display a high degree of differentiation between wild and domesticated forms as well as a pattern of positive selection within forms compared with neutral loci (Figure 2). They contribute to limiting effective gene flow at nearby loci, leading to the progressive buildup of the so-called genomic islands of divergence (Wolf and Ellegren 2017).

RI may also be promoted by the buildup of intrinsic barriers from the differential fixation of alleles that are incompatible at two or more interacting loci (Bateson-Dobzhansky-Muller Incompatibilities – BDMIs). Such BDMIs can evolve as a by-product of local adaptation to contrasting environments or through non-adaptive processes (Wolf and Ellegren 2017). If selection favors distinct mutational steps at several loci in each population, deleterious side effect interactions may arise when brought together in hybrids. These interactions may in turn provoke detrimental symptoms and/or Transmission Ratio Distortions (TRDs) at F2 generation for recessive alleles, contributing to intrinsic post-zygotic isolation.

In domesticated forms, the accumulation of deleterious mutations through domestication bottlenecks and linked selection may have accelerated the evolution of BDMIs between wild and domesticated forms. The loci underlying BDMIs should display genomic fingerprints that can be similar to those left by selection for habitat adaptation in domestic or wild populations (Figure 2). However, in the absence of intra-form selection, we expect increased divergence between forms while the level of polymorphism is not affected by the cessation of gene flow (Figure 2). This illustrates how the use of different statistics helps to clarify the mechanisms at work in RI (Cruickshank and Hahn 2014). There is of course a continuum of scenarios between those presented above: the barrier loci that limit gene flow between wild and domesticates can be directly targeted by selection within one of the two forms.

A particular kind of negative epistatic interaction can emerge as a by-product of coevolution between nuclear and cytoplasmic genomes. If different combinations have coevolved in divergent lineages, this may result in organelle dysfunction and hybrid breakdown when inter-lineage crosses occur (Burton *et al.* 2013). Such negative cytonuclear conflicts often result in asymmetrical reproductive barriers, which can be revealed by reciprocal crosses between wild and domesticated lineages as observed in *Citrus* (Wang *et al.* 2022).

Finally, reproductive barriers may result from parental conflicts generating allelic dosage perturbations. If species have evolved contrasting levels of parental conflicts, it can translate into paternal or maternal excess of gene expression in a hybrid context (Florez-Rueda *et al.* 2016). As evidenced by transcriptomic comparisons of wild and domestic forms (e.g., common bean, Bellucci *et al.* 2014; tomato, Sauvage *et al.* 2017) and simulations (Burban *et al.* 2022), domestication led to a profound reorchestration of coexpression networks, which can then cause disruptions in allelic dos-

age between wild and domestic forms resulting in fitness decline in wild x domesticated crosses.

### **HYPOTHESES TESTING & CHALLENGES**

Do the number of generations since domestication correlate with hybrid defects? Does the mutation load depend on the domestication history and the strength of RI? Do “stronger” domestication syndromes and/or higher genome-wide neutral divergence and/or extent of islands of differentiation induce stronger isolation?

Answering these questions will bring unique insights into the very-first steps of reproductive barriers formation, but detecting barrier genes is a daunting task. Divergent selection and BDIMs among loci create patterns of strong allelic differentiation relative to the genomic background (Figure 2) that together with linked loci, form genomic islands of differentiation. Their detection requires overcoming confounding effects such as local variation in recombination rates and effective population size (Wolf and Ellegren 2017). At the species level, parameters such as mating system, intensity of domestication and changes in effective population size (e.g., due to domestication bottlenecks) determine the extent of selection, genetic drift and linkage disequilibrium, and in turn the expected size and depth of islands of differentiation. Ultimately, interpretations of genomic differentiation patterns need to be guided by modelling in order to properly estimate the fraction of the genome recalcitrant to gene flow and identify the corresponding regions, which can then be combined to experimental results.

### **CONCLUSION**

The alterations of habitats due to human activities are precious laboratories to explore the mechanisms involved in adaptive divergence and the initial phases of speciation (Thompson *et al.* 2018, Touchard *et al.* 2023). The establishment of RI be-

tween wild and domestic forms is a crucial aspect of domestication that has received little attention. In addition to providing basic information about the processes at work in the early stages of speciation, testing for cross-compatibility between cultivated plants and their wild relatives and detecting the underlying barrier loci are essential for overcoming them. Crops wild relatives have faced continuous environmental challenges in their natural environment and often exhibit greater genetic diversity than their domesticated relatives, so they are a valuable reservoir of adaptive alleles that transferred to crops could help mitigate their vulnerability.

### **Acknowledgements section**

We are grateful to the editor in chief and two anonymous reviewers for helping us to improve the manuscript.

### **Authors contribution & Funding**

This essay is the result of discussions and close collaboration between all participants in the DomIsol project (ANR-19-CE32-0009-02).

### **Data Availability Statement**

No new data were generated for this manuscript.

### **Literature Cited**

**Bellucci E, Bitocchi E, Ferrarini A, et al. 2014.** Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell* **26**: 1901–1912.

**Burban E, Tenailon MI, Le Rouzic A. 2022.** Gene network simulations provide testable predictions for the molecular domestication syndrome. *Genetics* **220**: iyab214.

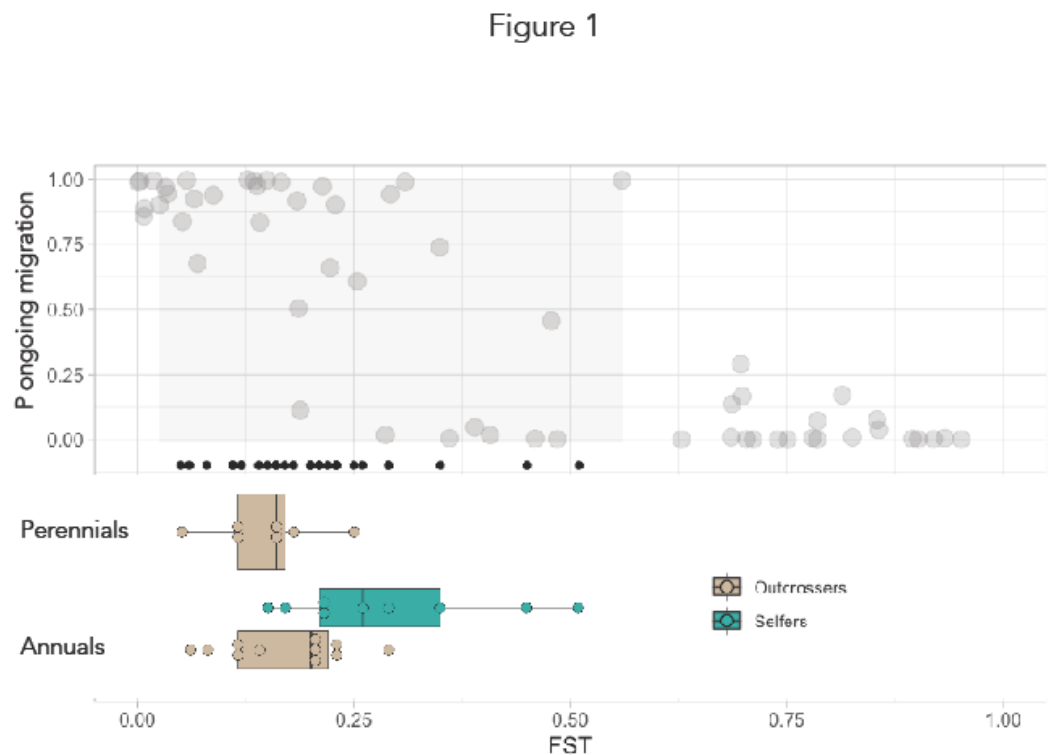


- Burton RS, Pereira RJ, Barreto FS. 2013.** Cytonuclear genomic interactions and hybrid breakdown In: Futuyma DJ, ed. *Annual Review of Ecology Evolution and Systematics. Annual Review of Ecology, Evolution, and Systematics* **44**: 281–302.
- Calfee E, Gates D, Lorant A, Perkins MT, Coop G, Ross-Ibarra J. 2021.** Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *PLOS Genetics* **17**: e1009810.
- Cruickshank TE, Hahn MW. 2014.** Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.
- Dempewolf H, Hodgins KA, Rummell SE, Ellstrand NC, Rieseberg LH. 2012.** Reproductive isolation during domestication. *Plant Cell* **24**: 2710–2717.
- Ellstrand NC, Meirmans P, Rong J, et al. 2013.** Introgression of crop alleles into wild or weedy populations In: Futuyma DJ, ed. *Annual Review of Ecology Evolution and Systematics. Annual Review of Ecology, Evolution, and Systematics* **44**: 325–345.
- Florez-Rueda AM, Paris M, Schmidt A, Widmer A, Grossniklaus U, Stadler T. 2016.** Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Molecular Biology and Evolution* **33**: 2935–2946.
- Gaut BS, Díez CM, Morrell PL. 2015.** Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics* **31**: 709–719.
- Hejase HA, Salman-Minkov A, Campagna L, et al. 2020.** Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proceedings of the National Academy of Sciences* **117**: 30554–30565.
- Le Corre V, Siol M, Vigouroux Y, Tenailon MI, Délye C. 2020.** Adaptive introgression from maize has facilitated the establishment of a teosinte as a noxious weed in Europe. *Proc. Proceedings of the National Academy of Sciences* **117**: 25618–25627.

- Lu Y, Hokin SA, Kermicle JL, Hartwig T, Evans MMS. 2019.** A pistil-expressed pectin methylesterase confers cross-incompatibility between strains of *Zea mays*. *Nature Communications* **10**: 2304.
- Marie-Orleach L, Brochmann C, Glémin S. 2022.** Mating system and speciation I: Accumulation of genetic incompatibilities in allopatry. *PLOS Genetics* **18**: e1010353.
- Purugganan MD. 2019.** Evolutionary Insights into the Nature of Plant Domestication. *Current Biology* **29**: R705–R714.
- Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016.** Shedding light on the grey zone of Speciation along a continuum of genomic divergence. *PLOS Biology* **14**: e2000234.
- Rushworth CA, Wardlaw AM, Ross-Ibarra J, Brandvain Y. 2022.** Conflict over fertilization underlies the transient evolution of reinforcement. *PLOS Biology* **20**: e3001814.
- Sauvage C, Rau A, Aichholz C, et al. 2017.** Domestication rewired gene expression and nucleotide diversity patterns in tomato. *Plant Journal* **91**: 631–645.
- Thompson KA, Rieseberg LH, Schluter D. 2018.** Speciation and the City. *Trends in Ecology & Evolution* **33**: 815–826.
- Touchard F, Simon A, Bierne N, Viard F. 2023.** Urban rendezvous along the sea-shore: Ports as Darwinian field labs for studying marine evolution in the Anthropocene. *Evolutionary Applications* **16**: 560–579.
- Wang N, Li C, Kuang L, et al. 2022.** Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. *Proceedings of the National Academy of Sciences* **119**: e2206076119.
- Wolf JBW, Ellegren H. 2017.** Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* **18**: 87–100.

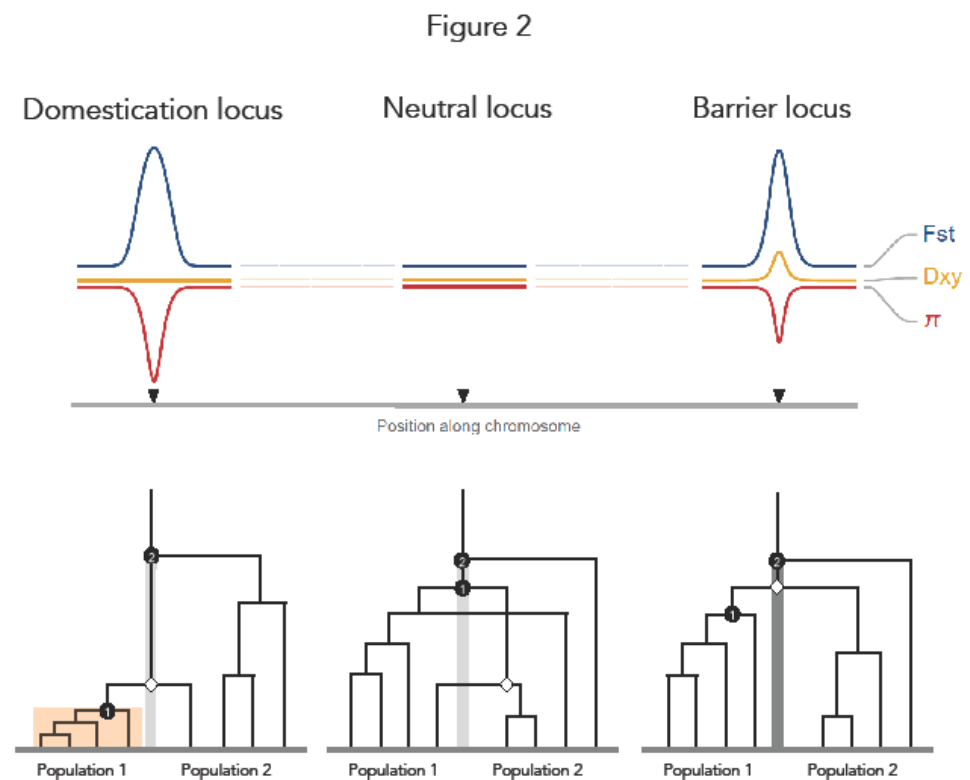
Yang CJ, Samayoa LF, Bradbury PJ, *et al.* 2019. The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proceedings of the National Academy of Sciences* **116**: 5643–5652.

### Figure legends



**Figure 1.** The grey zone of speciation as defined by Roux *et al.* 2016 encompasses the range of allele differentiation between wild and domestic forms across 27 plant species. Upper panel: Data illustrating the grey zone of speciation are taken from (Roux *et al.* 2016).  $F_{ST}$  values (x-axis) were computed for 61 pairs of animal populations/species across sequenced loci (natural divergence/speciation). The posterior probability of ongoing migration (y-axis) for a given pair reflects the capacity of demographic models that allow for ongoing exchange of migrants between diverging

lineages to predict the observed data compared to models where gene flow has stopped. The light grey rectangle spans the range of  $F_{ST}$  values in which both currently isolated and currently connected pairs are found, and therefore defines the co-called grey zone of speciation. Lower panel: Black dots along the x-axis correspond to  $F_{ST}$  values obtained for 27 wild/domestic plant species (Appendix S1 and references herein, see the supplementary data with this article).  $F_{ST}$  values for plant species were used to compute boxplots for annual (/biannual) species and perennial species. Boxplots are colored according to mating system.



**Figure 2. Theoretical expectations of summary statistics under divergence with gene flow in wild and domestic populations.** Patterns of allele differentiation ( $F_{ST}$ ), divergence ( $D_{xy}$ ) and diversity within the domesticated populations ( $\pi$ ) are displayed along the chromosome around three loci (black arrows) evolving under distinct sce-

narios: selective sweep at a locus involved in environmental adaptation and/or governing a domestication trait (domestication locus), neutrality (neutral locus), gene flow arrest at a locus that contributes to RI between populations (barrier locus). Representative genealogies of eight individuals from two divergent populations, a domestic population 1 and a wild population 2 are displayed (adapted from Hejase *et al.* (2020)). At a neutral locus and under continuous gene flow (light grey vertical bar), no allelic differentiation ( $F_{ST}=0$ ) is observed between populations that behave as a single population. Allelic differentiation ( $F_{ST}>0$ ) can be initiated either because the time to the most recent common ancestor – TMRCA is represented by black circles for population 1 and 2 – is reduced by a selective sweep (light orange rectangle) in one of the two populations, in this case the domestic population 1; or because the time to the first cross-coalescence between the populations (diamond) is increased by selection against gene flow (barrier locus, solid vertical bar). Note that in all graphs, the TMRCA of the population 2 is also the TMRCA of wild and domestic populations.