Gene network simulations provide testable predictions for the molecular domestication syndrome

Ewen Burban[*†], Maud I Tenaillon[‡], Arnaud Le Rouzic[*]

* Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198, Gif-sur-Yvette, France.

† CNRS, Univ. Rennes, ECOBIO - UMR 6553, F-35000 Rennes, France

‡ Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France.

Running title:

Network evolution during domestication

Keywords:

Plant domestication
Gene expression regulation
Population bottleneck
Phenotypic plasticity
Environmental change
Network topology
Artificial selection
Individual-based simulations
Population genetics
Genetic correlations

Corresponding author:

Arnaud Le Rouzic
UMR EGCE, Avenue de la Terrasse, Bâtiment 13,
91190 Gif-sur-Yvette, France
lerouzic@egce.cnrs-gif.fr
+33 (0)1 69 82 36 65

Abstract

The domestication of plant and animal species lead to repeatable morphological evolution, often referred to as the *phenotypic domestication syndrome*. Domestication is also associated with important genomic changes, such as the loss of genetic diversity and modifications of gene expression patterns. Here, we explored theoretically the effect of domestication at the genomic level by characterizing the impact of a domestication-like scenario on gene regulatory networks. We ran population genetics simulations in which individuals were featured by their genotype (an interaction matrix encoding a gene regulatory network) and their gene expressions, representing the phenotypic level. Our domestication scenario included a population bottleneck and a selection switch (change in the optimal gene expression level) mimicking canalizing selection, i.e. evolution towards more stable expression to parallel enhanced environmental stability in man-made habitat. We showed that domestication profoundly alters genetic architectures. Based on the well-documented example of the maize (*Zea mays ssp. mays*) domestication, our simulations predicted (i) a drop in neutral allelic diversity, (ii) a change in gene expression variance that depended upon the domestication scenario, (iii) transient maladaptive plasticity, (iv) a deep rewiring of the gene regulatory networks, with a trend towards gain of regulatory interactions between genes, and (v) a global increase in the genetic correlations among gene expressions, with a loss of modularity in the resulting coexpression patterns and in the underlying networks. Extending the range of parameters, we provide empirically testable predictions on the differences of genetic architectures between wild and domesticated and forms. The characterization of such systematic evolutionary changes in the genetic architecture of traits contributes to define a *molecular domestication syndrome*.

Introduction

Domestication of plant and animal species that occurred during the Neolithic revolution, 13,000 to 10,000 years ago, is often regarded as the most significant step in recent human history. It has laid the foundations for major transitions in lifestyle, culture, and civilization with the adoption of sedentarism, the emergence of socio-economic organization, the urbanization and anthropization of natural ecosystems, and the boost of technological advances (Diamond 2002). From the point of view of an evolutionary biologist, domestication is perceived as a process of rapid evolution over successive generations of anthropogenic selection, that has led to environmental adaptation to habitats created by humans and acquisition of profitable traits for them. Such innovations originate from genetic and plastic variation sustaining phenotypic shifts in domesticates compared to their wild counterparts (Fuller *et al.* 2010). In plants, traits targeted by those shifts alter architecture (more compact morphology), life-history (loss or partial loss of seed dispersal and seed dormancy, increased synchronicity of germination and ripening), as well as production- and usage-related traits (taste, increase of harvestable organs). They are often associated with convergent phenotypic changes across species (Larson *et al.* 2014), and collectively referred to as the *phenotypic domestication syndrome*.

The discovery of the genetic bases underlying variation of domesticated traits has been the focus of ample empirical work. Dozens of domestication genes have been discovered, most of which are transcription factors (Martínez-Ainsworth and Tenaillon 2016; Fernie and Yan 2019) embedded into complex gene regulatory networks (GRNs). Perhaps the most emblematic example is provided by the *Tb1* gene, which together with other genes controls maize branching architecture via hormone and sugar signaling (Doebley *et al.* 1997; Whipple *et al.* 2011; Dong *et al.* 2017, 2019). It is responsible for the strong apical dominance phenotype, i.e. repression of axillary bud outgrowth (Clark *et al.* 2006). Interestingly, in contrast to the maize allele, the *Tb1* allele from its wild ancestor (teosinte) confers a responsiveness to light when introgressed into a maize background (Lukens and Doebley 1999). It therefore appears that domestication has triggered the selection of a constitutive shade avoidance phenotype in maize (Studer *et al.* 2017), that has translated into a loss of phenotypic plasticity. Along this line, recent results indicate that reduced Genotype-by-Environment (GxE) interactions may be a general consequence for traits targeted by human selection. For example, genomic regions displaying footprints of selection explain less variability for yield GxE than “neutral” regions (Gage *et al.* 2017). Decreased phenotypic plasticity during domestication likely results both from the stability of human-made compared with wild habitats and from selection for stable crop performance across environments; it has yet to be characterized in other crops and for a broader range of traits.

In addition to the phenotypic syndrome, genome-wide sequencing data have revealed the outlines of a *molecular domestication syndrome*. This molecular syndrome includes a loss of genetic diversity through linked selection and constriction of population size due to sampling effects (Yamasaki *et al.* 2005). Severity of those genetic bottlenecks as estimated by nucleotide diversity loss, ranges from 17% to 49% in annuals while often no loss is observed in perennial fruit crops (reviewed in Gaut *et al.* 2015). The combined effect of bottlenecks, increased inbreeding (Glémin and Bataillon 2009) and linked selection in domesticates translates into shrink in effective population size, which in turn reduces the efficacy of selection against deleterious mutations (Moyers *et al.* 2018). Although increased recombination rate in domesticates compared with their wild relatives may partially compensate this effect (Ross-Ibarra 2004), fixation of deleterious mutations in domesticates and a resulting genetic load is often observed as exemplified in African rice (Nabholz *et al.* 2014), grapevine (Zhou *et al.* 2017), and maize (Wang *et al.* 2017).

Regarding molecular phenotypes assessed by transcriptomic surveys, data are still scarce and emerging patterns not as clear. Measures of variation of gene expression in domesticates relative to their wild counterparts either reveal a significant loss as in rice, cotton (Liu *et al.* 2019), beans (Bellucci *et al.* 2014); or a significant gain as in tomato (Sauvage *et al.* 2017); or no substantial change as in soybean (Liu *et al.* 2019), olives (Gros-Balthazard *et al.* 2019), and maize (Swanson-Wagner *et al.* 2012). In the latter, however, reduced variation in expression was observed at domestication candidates, indicating that selection primarily acts on *cis*-acting regulatory variants (Hufford *et al.* 2012). This result was further confirmed in F1 hybrids from maize x teosinte crosses where large differences in expression were primarily caused by *cis*-divergence, and correlated with genes targeted by selection during domestication (Lemmon *et al.* 2014).

Beyond quantitative measures of gene expression, domestication is also associated with gene network rewiring. A pioneer work in maize indeed indicates that 6% of all genes display altered co-expression profiles among which, genes targeted by selection during domestication and/or breeding are over-represented (Swanson-Wagner *et al.* 2012). Interestingly, networks encompassing domestication targets display greater connectivity in wild than in domesticated forms as if selection had triggered connection loss to/from these genes. In beans, coexpression networks at the genome level revealed a global excess of strong correlations in domesticates compared with wild, the latter being sparser with more isolated nodes and smallest connected components than the former (Bellucci *et al.* 2014). In contrast to maize, little qualitative difference was reported as for networks surrounding selected and neutral contigs.

While population genetic tools have been broadly used to estimate domestication bottlenecks and associated genetic load in plants (Eyre-Walker *et al.* 1998; Tenaillon

et al. 2004; Wright *et al.* 2005; Gaut *et al.* 2015; Kono *et al.* 2016; Liu *et al.* 2017; Wang *et al.* 2017), a theoretical framework that considers *molecular domestication syndrome* as a whole allowing to make predictions beyond verbal models is still in its infancy (Stetter *et al.* 2018). Here, we propose to simulate the evolution of gene regulatory networks in a population submitted to domestication-like pressures. We used a modified version of a classical gene network model (the 'Wagner' model, after Wagner 1994, 1996) to represent the complex genetic architecture of gene expression regulation, and tracked the evolution of genetic diversity, of gene expression plasticity, and of network topology in scenarios featuring (i) a temporary drop in the population size (bottleneck), and (ii) a substantial change in the selection regime. The demographic scenario was defined based on maize, an outcrosser crop with a relatively simple domestication (a single origin for the crop with a moderate domestication bottleneck); our simulations aim at providing experimentally testable predictions, and a general framework to explore more complex scenarios and other mating systems.

Materials and methods

Gene network model

The gene network model was directly inspired from Wagner (1996), with minor changes detailed below. Individual genotypes were stored as $n \times n$ interaction matrices **W**, representing the strength and the direction of regulatory interactions between n transcription factors or regulatory genes. Each element of the matrix W_{ij} stands for the effect of gene j on the expression of gene i; interactions can be positive (transcription activation), negative (inhibition), or zero (no direct regulation). Each line of the **W** matrix can be interpreted as an allele (the set of *cis*-regulatory sites of the transcription factor). The model considered discrete regulatory time steps, and the expression of the n genes, stored in a vector **P**, changes as $\mathbf{P}_{t+1} = F(\mathbf{W}\,\mathbf{P}_t)$, where $F(x_1, \ldots, x_n)$ applies a sigmoid scaling function $f(x)$ to all elements to ensure that gene expression ranges between 0 (no expression) and 1 (full expression). We used an asymmetric scaling function as in Rünneburger and Le Rouzic (2016); Odorico *et al.* (2018): $f(x) = 1/(1 + \lambda\, e^{-\mu x})$, with $\lambda = (1-a)/a$ and $\mu = 1/a(1-a)$. This function is defined such that $a = 0.2$ stands for the constitutive expression (in absence of regulation, all genes are expressed to 20% of their maximal expression).

The kinetics of the gene network was simulated for 24 time steps in each individual, starting from $P_0 = (a, \ldots, a)$. The simulation program reports, for each gene i, the mean $\overline{p_i}$ and the variance V_i of its expression level over the four last time steps. A non-null variance indicates an unstable gene network.

In addition to this traditional framework, we considered that one of the network genes was a "sensor" gene influenced by the environment. This makes it possible for the network to react to an environmental signal, and evolve expression plasticity. In practice, the environmental signal at generation g was drawn in a uniform distribution $e_g \sim U(0,1)$ and the value of the sensor gene was e_g at each time step (the sensor gene had no regulator and was not influenced by the internal state of the network).

Population model

The gene network model was coupled with a traditional individual-based population genetics model. Individuals were diploid and hermaphrodite, and generations were non-overlapping. Reproduction consisted in drawing, for each of the N offspring, two parents randomly with a probability proportional to their fitness. Each parent gave a gamete, i.e. a random allele at each of the n loci (free recombination). There was no recombinations between *cis*-regulatory sites at a given locus. The genotype of an individual (the **W** matrix from which the expression phenotype was calculated) was obtained by averaging out maternal and paternal haplotypes. Even if regulatory effects were additive, non-linearities in the gene network model are expected to generate a substantial amount of interactions (including dominance, epistasis, and pleiotropy) at the phenotypic (gene expression) level.

Individual fitness w was calculated as the product of two components, $w = w_U \times w_S$. The first term w_U corresponds to the penalty for unstable networks, $w_U = \prod_{i=1}^{n} exp(-s' V_i)$, s' being the strength of selection on unstable networks. The second term w_S corresponds to a Gaussian stabilizing selection component, which depends on the distance between the expression phenotype and a selection target θ: $w_S = \prod_{i=1}^{n} exp[-s_i(\overline{p_i} - \theta_i)^2]$, s_i standing for the strength of stabilizing selection on gene *i*. As detailed below, some genes were not selected (in which case $s_i = 0$), some genes were selected for a stable optimum θ_i ("stable" genes), while a last set of genes were selected for optima that changed every generation g ("plastic genes"), half of them being selected for $\theta_{ig} = e_g$, and the other half for $\theta_{ig} = 1 - e_g$ .Selection was moderate (s=10) for most simulations, albeit stronger selection (s=50) was also tested (Figure S2, S8C, and S9C).

Mutations occurred during gametogenesis with a rate m, expressed as the mutation probability per haploid genome. A mutation consists in replacing a random element of the **W** matrix by a new value drawn in a Gaussian distribution centered on the

former value $W_{ij}' \sim N(W_{ij}, \sigma_m)$, where σ_m is the standard deviation of mutational effects.

Domestication scenario and parameterization

Domestication was associated with two independent changes in the simulation parameters: a temporary demographic bottleneck (decrease in population size), and a change in the gene expression optima (modification of the selection pressure). In order to calibrate simulations with realistic parameters, we used the well-known maize domestication scenario as a reference. This scenario includes a protracted model of domestication involving a moderate bottleneck starting about 9,000 years ago with a ratio of the size of the bottlenecked population to the duration of the bottleneck equals to *k* = 2.45 (Wright *et al.* 2005). Simulations were thus split in three stages: (i) a long ($T_a = 12,000$ generations) "burn-in" stage aiming to simulate pre-domestication conditions, after which the large population size $N_a = 20,000$ (the largest population size that was computationally tractable) "ancestral" species is expected to harbor a genotype adapted to wild conditions (selection optima θ_a , drawn in a uniform $U(0,1)$ distribution at the beginning of each simulation for “stable” genes, fluctuating optima for “plastic” genes), (ii) a bottleneck of $T_b = 2,800$ generations (Eyre-Walker *et al.* 1998), during which the population size was reduced to $N_b = 3,500$ individuals, and selection optima switched to θ_b, and (iii) $T_c = 6,200$ generations of expansion of the domesticated species (population size back to $N_a = 20,000$), while the selection optima remained to the "domestication" conditions θ_b (Figure S1).

For computational feasibility, the regulation network size was limited to 24 genes (+1 environmental signal), from which 12 were under direct selection (Figure S3). Before domestication, the network encompassed 12 unselected, 6 stable, and 6 plastic genes. At the onset of domestication, we modified the selection regime to mimic increased environmental stability and, in turn decreased plasticity (12 unselected, 10 stable and 2 plastic genes). The mutation rate was set to $m = 10^{-3}$ / gamete/ generation, which, given the estimated mutational target of 24 genes of 1kb (average estimated length of enhancers from (Oka *et al.* 2017; Ricci *et al.* 2019)) roughly corresponded to a per-base mutation rate of 3×10^{-8} par generation, close to the maize estimate (Clark *et al.* 2004).

In addition to the default full domestication scenario described above, we explored control simulations to disentangle the contribution of the bottleneck and the selection switch in emerging patterns: a scenario with no bottleneck, a scenario with no selection switch, and a scenario with no stabilizing selection. We assessed the sensitivity of our results for the full domestication scenario to independent changes in parameters values by (1) increasing the number of genes of the GRN, from 24 to 48,

and doubling the number of selected genes and the mutation rate per genome accordingly; (2) setting the mutation rate to 0 at the time of domestication to evaluate selection response from standing variation only; (3) modulating selection intensity both through a decrease in selected genes count (by two-fold), and through a modification of the fitness function to simulate stronger selection;(4) dissociating selection switch from a loss of plasticity, either by maintaining the selection for plasticity over genes during domestication, or by keeping the same number of plastic gene before and after domestication ; (5) testing the effect of a harsher bottleneck with 350 individuals instead of 3500. All scenarios were replicated 1000 times.

Implementation

The simulation model was implemented in C++ and compiled with gcc v-7.5.0. Simulation runs were automated via bash scripts, replicates being launched in parallel, and simulation results were analyzed with R version 4.0 (R Core Team 2020).

Model output and descriptive statistics

For each simulation run, summary statistics were computed every 100 generations. The output includes the population mean and variance of (i) the absolute fitness w, (ii) gene expressions $\overline{p_i}$, (iii) gene regulations W_{ij} for all pairs of genes. In addition, the environmental index e_g and all selection optima θ_g were recorded.
Effective population sizes were estimated as $N_e = N/(1+4V_{\overline{w}})$ (Walsh & Lynch 2018), where $V_{\overline{w}}$ stands for the variance in the relative fitness ($V_{\overline{w}} = V_w/E_w{}^2$, V_w and E_w being the population variance and the population mean of the absolute fitness, respectively).

A proxy for neutral molecular variance around gene *i* was obtained by reporting the average population variance of the w_{ij} for a subset of genes *j* which expression was very low ($\overline{p_j} < 0.1$ over the whole simulation), as *cis*-regulatory sites sensitive to non-expressed transcription factors are expected to evolve neutrally.

Environmental reaction norms (gene expression plasticity) were estimated for each gene i by regressing the average expression $\bar{p}_i$ over the environmental index e_g, taken over a sliding window of 10 consecutive measurements (1000 generations).

The effect of gene regulations W_{ij} being quantitative (and thus, never exactly 0), the presence/absence of a connection in the network was determined by the following procedure: the expression phenotypes **P** and $\mathbf{P}^0_{ij}$ were calculated both from the full **W** matrix, and from each of the n^2 possible $\mathbf{W}^0_{ij}$ matrices in which W_{ij} was replaced

by 0. The regulation W_{ij} was considered as a meaningful connection when the Euclidean distance d(**P**, $\mathbf{P}^0_{ij}$) exceeded an arbitrary threshold of 0.1. Using other thresholds shifted the number of connections upward or downward, but did not affect the results qualitatively.

Genetic correlation matrices were estimated directly from the population covariances in gene expressions (hereafter called **G** matrices, although they reflect here all genetic components and not only additive (co)variances). The evolution of **G** matrices was tracked by computing the distance between consecutive matrices $\mathbf{G}_g$ and $\mathbf{G}_{g+500}$ in the simulation output. In practice, genetic covariances were turned into genetic correlation matrices, and then into genetic distance matrices $d = \sqrt{2(1-r)};$ the difference between both genetic distance matrices was calculated from their element-wise correlation, as in a Mantel test (function mantel.rtest in the R package ade4, Dray and Dufour 2007). Network topological features, including the number of clusters used as an index of modularity, were measured with the package igraph (Csardi and Nepusz 2006).

Data availability

The C++ simulation software is available at https://github.com/lerouzic/simevolv. All scripts (simulation launcher, data analysis, and figure generation) are available at https://github.com/lerouzic/domestication.

RESULTS

We used a gene network model encompassing 24 transcription factors to simulate the *molecular domestication syndrome* and provide testable predictions regarding (i) the evolution of molecular and expression variance, (ii) the evolution of plasticity and (iii) the extent of network rewiring. Regulation strength between genes was a quantitative variable directly affected by mutation at *cis*-regulatory sites, so that individual genotypes were stored in a matrix of interactions among all genes. Our simulations featured an outcrosser crop undergoing a rather classic protracted domestication scenario with a single moderate bottleneck. We modeled the selection switch associated with domestication both as a change in the gene expression optima and a change in the need of plastic response. This default domestication scenario was compared with simulations without bottleneck (albeit a selection switch), and simulations without selection switch (albeit a bottleneck).

Adaptation during habitat shift

The strong selection switch resulted both in a brief reduction of the N_e/N ratio (<1) (Figure S1), and in an immediate change in absolute fitness which dropped to < 0.1%, mimicking transient fitness loss of wild plants during habitat shift -- a wild individual would have a probability < 0.001 to be selected by a breeder over a modern crop strain (Figure 1). Fitness was slowly regained as domesticated plants adapted to their new cultivated habitat. Fitness recovery was slower in the scenario including a bottleneck, the end of which was featured by an increase in the rate of fitness gain. With the parameter set chosen for the simulations, the populations had entirely recovered their initial fitness roughly 9,000 generations after the selection switch (i.e. present), the process being 2000 generations faster in absence of a bottleneck (Figure 1). Most of the evolutionary change was due to new mutations, as simulations without mutations from the beginning of domestication i.e., adapting from the standing genetic variation only, did show a very limited response to selection (Figure S8B).

We simulated the loss of plasticity during domestication as a change in selection regime for four plastic genes (out of 6) towards stable selection or neutrality (Figure S3). The speed at which the gene network evolved increased by a factor of 10 to 20 —depending on gene type — when the selection regime shifted (Figure 2A). The genes for which the *cis*-regulatory regions evolved the most were genes which expression was plastic before the bottleneck. These genes were also affected by a second wave of accelerated evolution by the end of the bottleneck. The selection switch translated into an abrupt change in reaction norm for genes that became selected for a flat reaction norm (plastic → stable in Figure 2B). We indeed observed a rapid loss of plasticity, showing that it was an evolvable feature that responded to selection. More surprisingly, however, the loss of plasticity also affected genes that (i) were no longer under direct selection (Neutral), and (ii) were supposed to remain plastic, albeit to a lower extent. This short-term maladaptive evolution highlighted the genetic constraints during the rewiring of the network caused by the selection switch. Immediately after it, plastic genes were still tightly connected to genes that were selected to evolve a flat reaction norm, and the first stage of this evolutionary change involved a maladaptive trade-off. It was slowly resolved by rewiring the connections across genes. The bottleneck retarded the evolutionary change, as adaptive plasticity was recovered faster in constant population-size simulations. Maladaptive plasticity did not evolve in simulations where plastic genes were under the same selection regime before and after domestication (figure S11D and S11E), showing that underlying constraints resulted from selection-triggered changes in reaction norms.

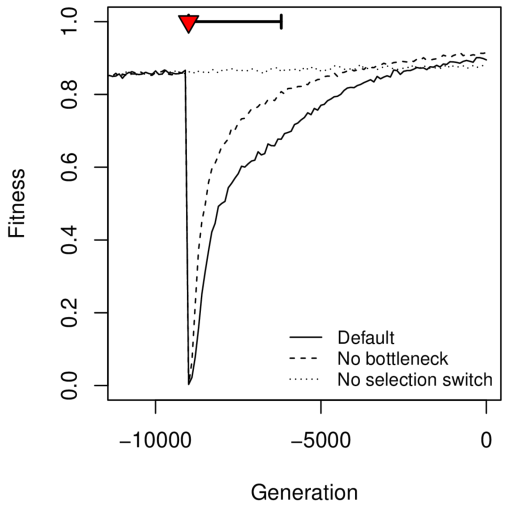

Figure 1: Effect of domestication on the average population fitness. The fitness drop corresponds to the switch in the optimal gene expressions at the onset of domestication (selection switch, red triangle). The population bottleneck is indicated as a thick horizontal segment. Three scenarios were considered : a full domestication scenario with selection switch and bottleneck (default); a scenario without bottleneck; a scenario with a bottleneck but constant selection (no switch). The figure shows the average over 1000 simulations for each scenario.

Molecular variation is affected by both demography and selection

Regulation strength was modeled as a quantitative variable directly affected by mutation. For any given gene in the network, we defined its molecular variance among individuals of the population as the average variance of the regulation strength at *cis*-regulatory sites. Hence, molecular variance is an analogous measure of nucleotide genetic diversity at *cis*-regulatory sites of the network. We focused here on the subset of *cis*-regulatory sites that were the targets of low-expressed genes in the network (average expression < 0.1 over the whole time series), considering these sites as neutral (but in total linkage disequilibrium with non-neutral *cis*-regulatory sites). Based on empirical evidence, the first signal that we expected was a loss of neutral genetic diversity. The variance indeed dropped sharply at the beginning of the domestication (Figure 3A). Such variance drop was due to two overlapping mechanisms: (i) genetic drift, that increased during the bottleneck, and (ii) selection, which generated selective sweeps leading to the loss of the genetic diversity at linked neutral sites during the bottleneck, which was too short to reach a drift-mutation-selection equilibrium. The maximum observed drop in genetic diversity was ~60% loss (from ~1 $x10^{-4}$ to ~$0.4x10^{-4}$) during the bottleneck for the default scenario. Recovery was slow and still ongoing at the end of the simulations.

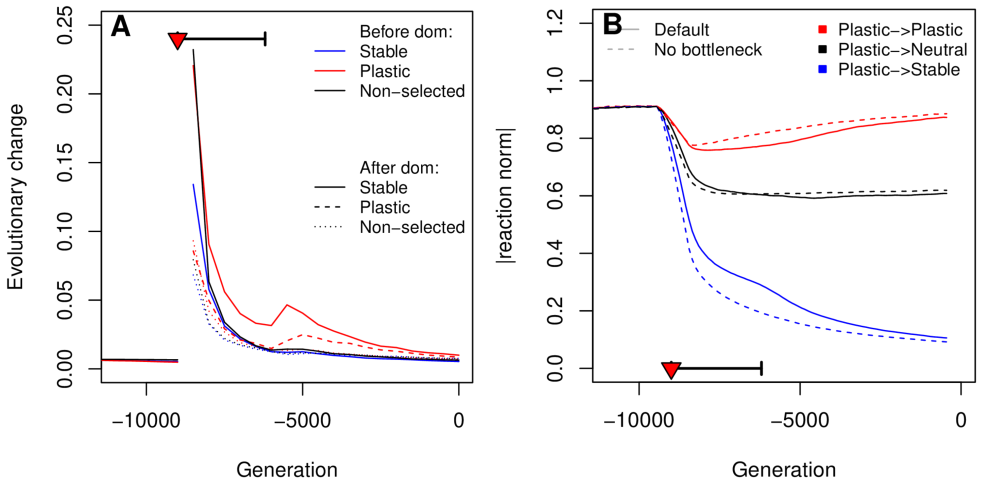

Figure 2: Changes in the rate of evolution and in plastic reaction norms upon domestication. A: Speed of evolutionary change as measured by the Euclidean distance between the regulation strengths of consecutive (100 generations apart) regulation matrices for various gene categories. Colors correspond to the status of the gene before domestication, line type to the status of the gene after domestication. B: Evolution of the average reaction norm for genes that were selected to be plastic before domestication. The selection switch and population bottleneck are indicated as a red triangle and a thick horizontal segment, respectively. The figure shows the average over 1000 simulations for each scenario.

In addition to change in molecular variance, we investigated the evolution of phenotypic (expression) variance during the domestication. Our results showed that in contrast to the former, phenotypic variance increased during domestication before reaching progressively its original level about 6000 generations after the end of the bottleneck (Figure 3B). Note that the recovery was faster than for diversity because the expression level was directly targeted by stabilizing selection. Expression variance bursts, absent from the simulations without selection switch, can be associated with ongoing adaptation: they corresponded to the segregation of selected variants that brought the phenotype closer to the new optimum (Figure S5). In the default scenario, adaptation proceeded in two waves: one immediately following the selection switch, and one, albeit much reduced, at the end of the bottleneck. The second wave was absent in simulations without bottleneck, confirming that the bottleneck slowed down the expression response to anthropogenic selection. In sum, the domestication, as we simulated it, was associated with an increase in the gene expression variance, as a result of the balance between the selection switch (which increased temporarily the variance) and the bottleneck (which slightly reduced the variance). In case of a stronger bottleneck, however, the expression diversity was reduced showing that the net effect on phenotypic diversity strongly depended on the details of the domestication scenario (Figure S8D).

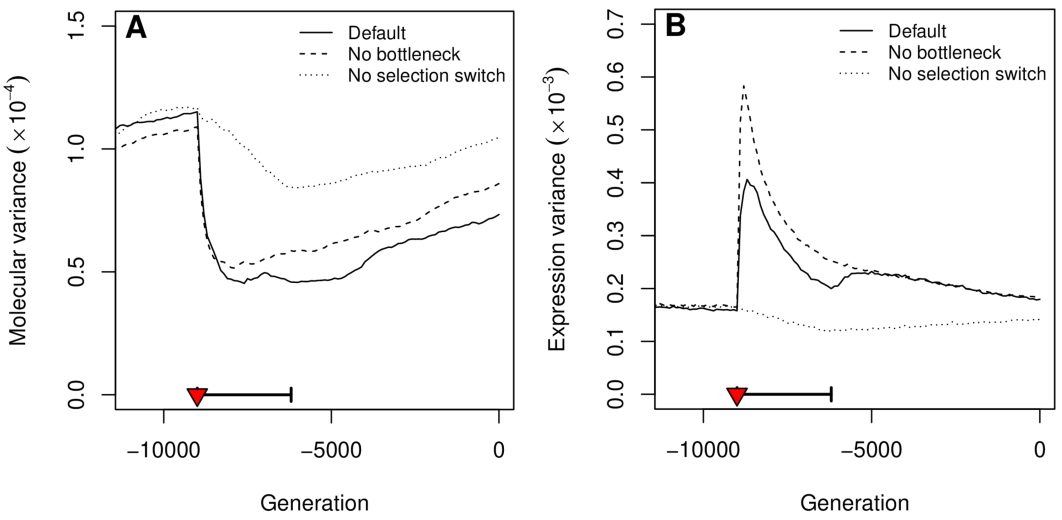

Figure 3 : Evolution of neutral molecular variance and gene expression variance through time. The population neutral molecular variance (A) was estimated from the regulation strength to low-expressed transcription factors, which measures the genetic diversity at neutral loci that are in complete linkage disequilibrium with selected genes. The population expression variance (B) stands for the within-population “phenotypic” variance of gene expressions, averaged over all genes. The selection switch and population bottleneck are indicated as a red triangle and a thick horizontal segment, respectively. The figure shows the average over 1000 simulations for each scenario.

Domestication is associated with the rewiring of gene networks

Genetic correlation matrices (**G** matrices) were estimated from the population covariances in gene expressions. Genetic correlations evolved rapidly after the domestication, and this evolution was driven both by the change in the selection regime and by the bottleneck (Figure 4A, Figure S7A). Domestication resulted in (i) a slight increase in the average coexpression (Figure 4A), and (ii) a redistribution of genetic correlations, with a loss of observable clusters of correlations (Figure 4B). The slight trend towards larger coexpressions hide a wide diversity of evolutionary change depending on the pair of genes considered (Figure S7B). Overall, strong correlations weakened during domestication, while many weak coexpressions increased.

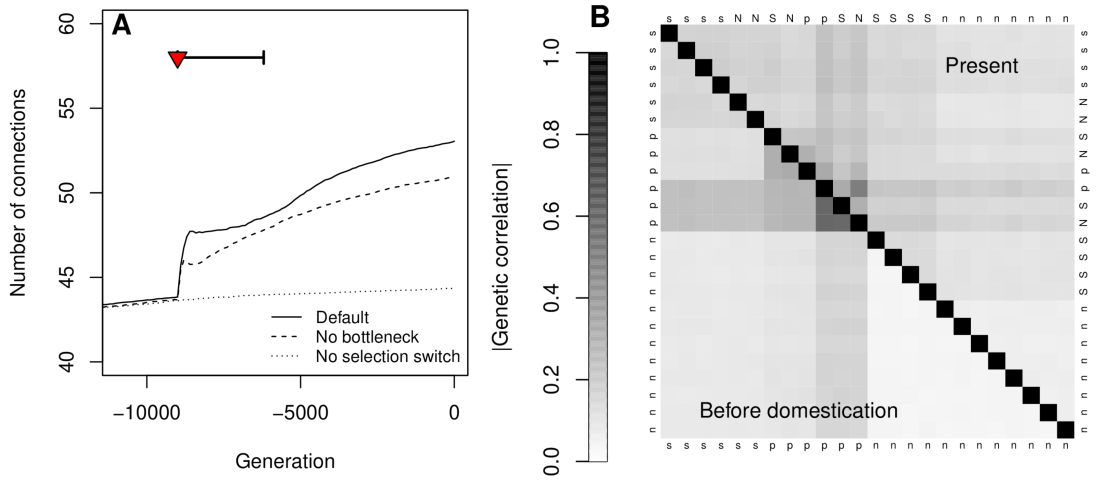

Figure 4: Consequences of domestication on gene coexpression. (A): Evolution of the average absolute value of within-population genetic correlations. The selection switch and population bottleneck are indicated as a red triangle and a thick horizontal segment, respectively. (B) Average genetic correlation for each pair of genes, at generation -9000 (just before the onset of domestication) below the diagonal, and at generation 0 (last generation of the simulations), above the diagonal. Gene selection status is indicated (n: non-selected, s: stable, p:plastic); capital letters indicate genes whose selection status changed during domestication.

We explored the evolution of the GRN topology during domestication, by tracking the evolution of the number of connections. We observed a strong signal of network rewiring during the first stage of domestication, with an increase (by a factor > 10) of the rates of both gained and lost connections, immediately after the selection switch (Figures 5A and S4A). This rewiring was solely due to the selection switch, as there was no effect of the bottleneck alone on the network evolution. The rewiring was associated with a systematic excess of gained connections over lost connections, i.e. domestication caused an increase in the total number of connections (Figure 5A). As a consequence of the gain of new connections, the number of clusters decreased (some connections appeared between previously independent modules, Figure 5B). New connections appeared to be distributed evenly across the network (Figure S4B and S6).

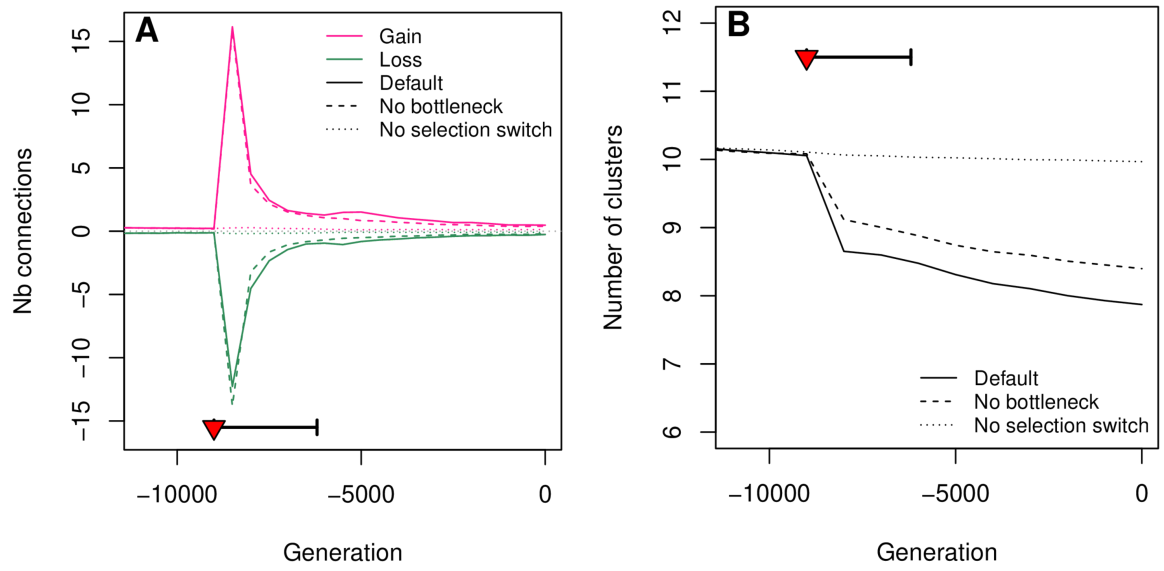

Figure 5: Evolution of gene network properties. The presence/absence of a regulatory connection (A) was determined based on its effect on gene expression (see methods). Connection gains and losses were counted over windows of 500 generations. The drop in the number of clusters (B) corresponds to new connections among existing clusters. The selection switch and population bottleneck are indicated as a red triangle and a thick horizontal segment, respectively. The figure shows the average over 1000 simulations for each scenario.

Discussion

Domestication is a complex process, involving deep modifications of the demographic, environmental, and selective context in which populations evolve. Here, we explored the consequences of domestication-like changes on the evolution of gene regulatory networks underlying domestication traits, combining a population bottleneck and phenotypic canalization, simulated as the evolution of selection pressure towards decreased plasticity paralleling environmental stability of phenotypes.

Adaptive dynamics under domestication

We observed that the bottleneck had a substantial effect on genetic diversity, including (i) a substantial loss of neutral genetic (molecular) diversity, (ii) a moderate loss of expression variance. These observations are in line with theoretical expectations. When the population size drops, genetic diversity is expected to be lost progressively, as the inbreeding coefficient increases by a factor $(1-1/2N_e)$ every generation. How much of the initial diversity of the species survives the bottleneck depends on the strength and the duration of the population size drop; in our

simulations, parameterized from the maize domestication scenario, about 60% of the initial neutral diversity survived the bottleneck. This estimate matched the 60% of mean pairwise diversity retained in “neutral” maize regions as defined as those located 5 kb away from genes, with π=0.00691 and 0.0115 in maize and teosintes, respectively (Beissinger *et al.* 2016).

Less expected perhaps was the fact that even such a mild bottleneck penalized substantially the response to anthropic selection (Figure 1). This may be due to less frequent occurrence of adaptive mutations during the bottleneck or a diminished efficiency of selection, or a combination of both. In the simulations, the end of the bottleneck was associated with a burst of segregating adaptive alleles (Figure 3), which suggests a two-stage domestication scenario: (i) during the bottleneck, the adaptive alleles that segregate (either from the standing genetic variation and/or from new mutations) increased the population fitness, but tend to have suboptimal effects (e.g. negative side effects on well-adapted genes are illustrated by plastic genes whose reaction norm diminishes while they are continuously selected to be plastic); (ii) after the end of the bottleneck, a new set of adaptive alleles can invade the population (because more mutations are available and selection is more efficient), fine-tuning genetic effects, e.g. on reaction norms (Figure 2). Hence, we expect mutations segregating during the first stage and surviving to drift to display greater effects than those segregating during the second stage. In line with this prediction, early work on maize domestication has identified several quantitative trait loci (QTLs) with large effects, some of which were fine-mapped down to individual genes such as *Tb1* (Doebley *et al.* 1997; Studer *et al.* 2011) and *Tga1* (Wang *et al.* 2005). Examples of early mutations with large effects have also been recovered in tomato (Frary *et al.* 2000), in wheat (Simons *et al.* 2006), in rice (Konishi *et al.* 2006; Li *et al.* 2006), in barley (Komatsuda *et al.* 2007) among others. These large QTLs that most likely encode early domestication targets stand as exceptions in the overall architecture of domestication traits dominated by small-effect QTLs as recently reported in maize (Chen *et al.* 2020).

Most phenotypic changes associated with domestication are controlled by mutations in transcription factors, and therefore involve a re-orchestration of gene networks (Martínez-Ainsworth and Tenaillon 2016) as described in cotton (Rapp *et al.* 2010), maize (Hufford *et al.* 2012), bean (Bellucci *et al.* 2014) and tomato (Sauvage *et al.* 2017). Consistently, in our simulations, the gene network was deeply rewired, as the rate of gain/loss connections increased by more than one order of magnitude. This effect was solely due to the shift in the selection regime. Before domestication, the population was well-adapted to an arbitrary wild type fitness landscape, involving genes which expression was constant and genes which expression was selected to track the environment. The structure of the underlying network evolved so that expressions of genes of the same type were genetically correlated, suggesting direct or indirect regulatory connections. When the fitness landscape changed, some genes that were previously correlated were forced to become independent. The

results suggest that this was easier to achieve by adding connections rather than removing them, illustrating evolution by genetic tinkering instead of re-engineering. Interestingly, there was no apparent cost to this additional complexity, as the fitness after domestication reached similar levels as before domestication.

Model approximations

Gene network models based on Wagner (1994) are built on a set of simplifying assumptions: the network dynamics is discretized and simplified (e.g. no distinction between RNA products and proteins), mutations can occur only in *cis* (transcription factors do not evolve), there are no interactions between transcription factors (their effects adds up), a given transcription factor can act both as an activator and a repressor, and a single mutation can switch the sign of the effect. Little is known about the potential effect of such details on the general dynamics of the network, but it can probably be safely assumed that complexifying the model is unlikely to affect qualitatively the outcome of the simulations.

For the sake of realism, and to connect the model results to quantitative genetics theory, we proposed several changes to the original framework. We adopted the setting used in e.g. Siegal and Bergman (2002), in which gene expression was considered as a quantitative character, with a continuous scaling function between 0 (no expression) and 1 (maximal expression), instead of the traditional on/off binary setting (Wagner 1996; Ciliberti *et al.* 2007). We used an asymmetric sigmoid scaling (as in Rünneburger and Le Rouzic 2016) to ensure that a non-regulated gene has a low expression (here, 20% of the max). The major improvement of the model was the possibility to evolve a plastic response. We added a perfect environmental cue as an input of the network through a sensor gene, which expression was reflecting the environmental index during the whole network dynamics. The literature provides alternative settings to introduce plasticity in the Wagner model, such as the introduction of the environmental cue as the starting state of the network, mimicking developmental plasticity (Masel 2004), or trans-generational plasticity (Odorico *et al.* 2018).

Computational constraints limited the population size to a maximum of N=20,000. Estimates of the effective population sizes of both maize and teosinte vary roughly between 10^5 and 10^6 (Eyre-Walker 1998; Tenaillon et al. 2004; Beissinger et al. 2016; Wang et al. 2017) depending on data/methods/models, suggesting that genetic drift before and after the bottleneck was substantially larger in the simulations than expected in a realistic domestication scenario. The domestication scenario was also greatly simplified, with a single bottleneck. Refinements of this initial scenario include multiple expansion waves of semi-domesticated forms, as well as rapid population growth and gene flow with wild relatives post-domestication (Beissinger *et al.* 2016; Kistler *et al.* 2018). Larger population size would raise the

neutral diversity, but is unlikely to impact general outcomes. As for the exploration of more complex scenarios, we provide a model that could be modified to refine predictions for a broader range of scenarios and species. Likewise, network size also had to be limited to n=24 genes, as the complexity of the gene network algorithm increases with the square of the number of genes. Defining a realistic size for a gene network remains problematic, as, in fine, most genes are connected through correlated regulations. Nevertheless, we considered here only transcription factors (or TF-like regulators, such as regulatory RNAs), which have the potential to affect the expression of other genes.

Finally, how selection affects the expression level of such TFs remains quite arbitrary. In the default scenario, most of the response to selection was due to new mutations, as the standing genetic variation alone could not explain more than about 20% of the (log) fitness recovery. The contribution of standing genetic variation to the response to selection is a complex function of the mutational variance, the strength of stabilizing selection before domestication, and the strength of directional selection during domestication (Stetter *et al.* 2018). The simulations thus correspond to a harsh domestication scenario in this respect, where the number of selected traits and the phenotypic changes induced by domestication were both large compared to the phenotypic diversity of the wild ancestor. We also considered that the expression level of only half of the network genes was under direct selection pressure -- this would happen if half of the TFs were regulating directly key enzymes or growth factors. Simulating twice less selected genes did not affect the qualitative outcomes of the results (Figures S9B and S11B).

The molecular syndrome of domestication

Simulations confirm that the domestication process is expected to be associated with several characteristic signatures (S) at the molecular level; S1: a decrease of allelic diversity, S2: a change in gene expression variance, S3: the rewiring of the gene regulatory networks, and S4: less modularity of coexpression patterns.

The loss of genetic diversity (S1) was both due to the bottleneck (genetic drift removed rare alleles from the population) and to the selection shift (selective sweeps decreased the genetic diversity at linked loci), it is thus expected to be a general signature of domestication. Empirically, a loss of genetic diversity is indeed always associated with domestication, although its amplitude may vary (reviewed in Gaut *et al.* 2015).

The direction and magnitude of the evolution of gene expression variance (S2) depends on the balance between selection and drift; bottlenecks tend to reduce diversity, while a shift in the selection regime tends to increase it transiently (segregation of adaptive variants). Given our simulation parameters, inspired from

the maize domestication scenario featuring a mild bottleneck, expression variance increased. This was not necessarily the case with all parameter combinations, as a stronger bottleneck led to a decrease in both molecular and expression variance. The strength and the pattern of selection also affect the speed and the nature (soft vs. hard) of the selective sweeps, which may differ across species. As a consequence, domestication is not expected to be associated with a systematic evolution of gene expression variance: it may increase when the bottleneck is moderate, as in maize, or decrease in species where the bottleneck was drastic and/or associated with an autogamous mating system, such as rice, cotton (Liu *et al.* 2019), and beans (Bellucci *et al.* 2014). .

Genetic networks were rewired (S3) and evolved towards less modularity, as a consequence of swapping the selection pattern among genes (shift in the optimal expression for stable genes, and loss of plasticity for others). New connections occurred among previously isolated modules, but former connections were not all eliminated. As a result, the rewiring of regulatory connections lead to a moderate increase in gene coexpressions (S4), associated with a loss of structure in the coexpression network (uncorrelated genes became correlated, and strongly correlated genes became more independent). This illustrates a realistic evolutionary scenario towards non-adaptive complexity, where the final network structure is not the more efficient one, but rather results from the accumulation of successive beneficial mutations in an existing, constrained genetic background. Empirically, we therefore predict that connections involving genes targeted by domestication should increase rather than decrease, in line with observations in beans where coexpression networks revealed a global excess of strong correlations in domesticates compared with wild (Bellucci *et al.* 2014). Global increase in genetic correlations should translate into greater constraints and pleiotropy, and less independent modules. Interestingly, the general increase in genetic correlations was associated with a trend towards homogenization, i.e. strong correlations tended to weaken whereas uncorrelated genes became slightly correlated. Empirical comparisons at 18 domestication-related traits between two independent populations of offspring generated by the intermating of multiple parents from a teosinte population and from a maize landrace, revealed several interesting features in line with our observations: only a subset of genetic correlations (33 out of 153) were conserved between teosinte and maize, teosinte correlations were more structured among trait groups (Yang *et al.* 2019).

The limited genetic diversity available in modern cultivated species is often considered as a limitation to further response to artificial selection. Controlling recombination has been proposed as crucial for plant breeders to engineer novel allele combinations and reintroduce diversity from wild crop relatives (reviewed in Taagen *et al.* 2020). Yet, if the domestication syndrome was also associated with changes in the pleiotropy of the genetic architecture, genetic progress might also be limited by undesirable genetic correlations among traits of interest (Yang *et al.* 2019).

Understanding how genetic constraints evolved under anthropic selection and whether it is possible to avoid or revert them requires a better understanding of the complex non-linear mapping between domestication genes and phenotypes.

Acknowledgements

We are grateful to Sylvain Glémin for insightful comments on the manuscript. We thank Clémentine Vitte for useful literature suggestions on enhancers. Simulations were performed on the core cluster of the Institut Français de Bioinformatique (https://www.france-bioinformatique.fr/ifb-core/).

Funding information

This work received financial support from two grants overseen by the French National Research Agency (ANR): one from the LabEx BASC -- Biodiversité, Agroécosystèmes, Société, Climat (ANR-11-LABX-0034) to ALR, and the DomIsol project (ANR-19-CE32-0009) to MIT. We thank the GDR 3765 “Approche Interdisciplinaire de l'Évolution Moléculaire” for travel support to EB. EGCE and GQE-Le Moulon benefit from the support the Institut Diversité, Écologie et Évolution du Vivant (IDEEV), and GQE-Le Moulon from Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Literature cited

Beissinger T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, *et al.*, 2016 Recent demography drives changes in linked selection across the maize genome. Nat. Plants 2: 1–7.

Bellucci E., E. Bitocchi, D. Rau, M. Rodriguez, E. Biagetti, *et al.*, 2014 Genomics of origin, domestication and evolution of Phaseolus vulgaris, pp. 483–507 in *Genomics of plant genetic resources*, Springer.

Chen Q., L. F. Samayoa, C. J. Yang, P. J. Bradbury, B. A. Olukolu, *et al.*, 2020 The genetic architecture of the maize progenitor, teosinte, and how it was altered during maize domestication. PLoS Genet. 16: e1008791.

Ciliberti S., O. C. Martin, and A. Wagner, 2007 Innovation and robustness in complex regulatory gene networks. Proc. Natl. Acad. Sci. 104: 13591–13596.

Clark R. M., E. Linton, J. Messing, and J. F. Doebley, 2004 Pattern of diversity in the genomic region near the maize domestication gene tb1. Proc. Natl. Acad. Sci. 101: 700–707.

Clark R. M., T. N. Wagler, P. Quijada, and J. Doebley, 2006 A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nat. Genet. 38: 594–597.

Csardi G., and T. Nepusz, 2006 The igraph software package for complex network research. InterJournal Complex Systems: 1695.

Diamond J., 2002 Evolution, consequences and future of plant and animal domestication. Nature 418: 700–707.

Doebley J., A. Stec, and L. Hubbard, 1997 The evolution of apical dominance in maize. Nature 386: 485–488.

Dong Z., W. Li, E. Unger-Wallace, J. Yang, E. Vollbrecht, *et al.*, 2017 Ideal crop plant architecture is mediated by tassels replace upper ears1, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. Proc. Natl. Acad. Sci. 114: E8656–E8664.

Dong Z., Y. Xiao, R. Govindarajulu, R. Feil, M. L. Siddoway, *et al.*, 2019 The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. Nat. Commun. 10: 1–15.

Dray S., and A.-B. Dufour, 2007 The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22: 1–20.

Eyre-Walker A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. 95: 4441–4446.

Fernie A. R., and J. Yan, 2019 De novo domestication: an alternative route toward new crops for the future. Mol. Plant 12: 615–631.

Frary A., T. C. Nesbitt, A. Frary, S. Grandillo, E. Van Der Knaap, *et al.*, 2000 fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85–88.

Fuller D. Q., R. G. Allaby, and C. Stevens, 2010 Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. World Archaeol. 42: 13–28.

Gage J. L., D. Jarquin, C. Romay, A. Lorenz, E. S. Buckler, *et al.*, 2017 The effect of artificial selection on phenotypic plasticity in maize. Nat. Commun. 8: 1–11.

Gaut B. S., C. M. Díez, and P. L. Morrell, 2015 Genomics and the contrasting dynamics of annual and perennial domestication. Trends Genet. 31: 709–719.

Glémin S., and T. Bataillon, 2009 A comparative view of the evolution of grasses under domestication. New Phytol. 183: 273–290.

Gros-Balthazard M., G. Besnard, G. Sarah, Y. Holtz, J. Leclercq, *et al.*, 2019 Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. Plant J. 100: 143–157.

Hufford M. B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J.-M. Chia, *et al.*, 2012 Comparative population genomics of maize domestication and improvement. Nat. Genet. 44: 808–811.

Kistler L., S. Y. Maezumi, J. G. De Souza, N. A. Przelomska, F. M. Costa, *et al.*, 2018 Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science 362: 1309–1313.

Komatsuda T., M. Pourkheirandish, C. He, P. Azhaguvel, H. Kanamori, *et al.*, 2007 Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc. Natl. Acad. Sci. 104: 1424–1429.

Konishi S., T. Izawa, S. Y. Lin, K. Ebana, Y. Fukuta, *et al.*, 2006 An SNP caused loss of seed shattering during rice domestication. Science 312: 1392–1396.

Kono T. J., F. Fu, M. Mohammadi, P. J. Hoffman, C. Liu, *et al.*, 2016 The role of deleterious substitutions in crop genomes. Mol. Biol. Evol. 33: 2307–2317.

Larson G., D. R. Piperno, R. G. Allaby, M. D. Purugganan, L. Andersson, *et al.*, 2014 Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. 111: 6139–6146.

Lemmon Z. H., R. Bukowski, Q. Sun, and J. F. Doebley, 2014 The role of cis regulatory evolution in maize domestication. PLoS Genet 10: e1004745.

Li C., A. Zhou, and T. Sang, 2006 Rice domestication by reducing shattering. Science 311: 1936–1939.

Liu Q., Y. Zhou, P. L. Morrell, and B. S. Gaut, 2017 Deleterious variants in Asian rice and the potential cost of domestication. Mol. Biol. Evol. 34: 908–924.

Liu W., L. Chen, S. Zhang, F. Hu, Z. Wang, *et al.*, 2019 Decrease of gene expression diversity during domestication of animals and plants. BMC Evol. Biol. 19: 19.

Lukens L. N., and J. Doebley, 1999 Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. Genet. Res. 74: 291–302.

Martínez-Ainsworth N. E., and M. I. Tenaillon, 2016 Superheroes and masterminds of plant domestication. C. R. Biol. 339: 268–273.

Masel J., 2004 Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. J. Evol. Biol. 17: 1106–1110.

Moyers B. T., P. L. Morrell, and J. K. McKay, 2018 Genetic costs of domestication and improvement. J. Hered. 109: 103–116.

Nabholz B., G. Sarah, F. Sabot, M. Ruiz, H. Adam, *et al.*, 2014 Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima). Mol. Ecol. 23: 2210–2227.

Odorico A., E. Rünneburger, and A. Le Rouzic, 2018 Modelling the influence of parental effects on gene-network evolution. J. Evol. Biol. 31: 687–700.

Oka R., J. Zicola, B. Weber, S. N. Anderson, C. Hodgman, *et al.*, 2017 Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. 18: 137.

R Core Team, 2020 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rapp R. A., C. H. Haigler, L. Flagel, R. H. Hovav, J. A. Udall, *et al.*, 2010 Gene expression in

developing fibres of Upland cotton (Gossypium hirsutum L.) was massively altered by domestication. BMC Biol. 8: 139.

Ricci W. A., Z. Lu, L. Ji, A. P. Marand, C. L. Ethridge, *et al.*, 2019 Widespread long-range cis-regulatory elements in the maize genome. Nat. Plants 5: 1237–1249.

Ross-Ibarra J., 2004 The evolution of recombination under domestication: a test of two hypotheses. Am. Nat. 163: 105–112.

Rünneburger E., and A. Le Rouzic, 2016 Why and how genetic canalization evolves in gene regulatory networks. BMC Evol. Biol. 16: 239.

Sauvage C., A. Rau, C. Aichholz, J. Chadoeuf, G. Sarah, *et al.*, 2017 Domestication rewired gene expression and nucleotide diversity patterns in tomato. Plant J. 91: 631–645.

Siegal M. L., and A. Bergman, 2002 Waddington's canalization revisited: developmental stability and evolution. Proc. Natl. Acad. Sci. 99: 10528–10532.

Simons K. J., J. P. Fellers, H. N. Trick, Z. Zhang, Y.-S. Tai, *et al.*, 2006 Molecular characterization of the major wheat domestication gene Q. Genetics 172: 547–555.

Stetter M. G., K. Thornton, and J. Ross-Ibarra, 2018 Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLoS Genet. 14: e1007794.

Studer A., Q. Zhao, J. Ross-Ibarra, and J. Doebley, 2011 Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. 43: 1160–1163.

Studer A. J., H. Wang, and J. F. Doebley, 2017 Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. Genetics 207: 755–765.

Swanson-Wagner R., R. Briskine, R. Schaefer, M. B. Hufford, J. Ross-Ibarra, *et al.*, 2012 Reshaping of the maize transcriptome by domestication. Proc. Natl. Acad. Sci. 109: 11878–11883.

Taagen E., A. J. Bogdanove, and M. E. Sorrells, 2020 Counting on crossovers: controlled recombination for plant breeding. Trends Plant Sci. 25: 455–465.

Tenaillon M. I., J. U'Ren, O. Tenaillon, and B. S. Gaut, 2004 Selection versus demography:

a multilocus investigation of the domestication process in maize. Mol. Biol. Evol. 21: 1214–1225.

Wagner A., 1994 Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. Proc. Natl. Acad. Sci. 91: 4387–4391.

Wagner A., 1996 Does Evolutionary Plasticity Evolve? Evolution 50: 1008–1023.

Wang H., T. Nussbaum-Wagler, B. Li, Q. Zhao, Y. Vigouroux, *et al.*, 2005 The origin of the naked grains of maize. Nature 436: 714–719.

Wang L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay of demography and selection during maize domestication and expansion. Genome Biol. 18: 1–13.

Whipple C. J., T. H. Kebrom, A. L. Weber, F. Yang, D. Hall, *et al.*, 2011 grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. Proc. Natl. Acad. Sci. 108: E506–E512.

Wright S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, *et al.*, 2005 The effects of artificial selection on the maize genome. Science 308: 1310–1314.

Yamasaki M., M. I. Tenaillon, I. V. Bi, S. G. Schroeder, H. Sanchez-Villeda, *et al.*, 2005 A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17: 2859–2872.

Yang C. J., L. F. Samayoa, P. J. Bradbury, B. A. Olukolu, W. Xue, *et al.*, 2019 The genetic architecture of teosinte catalyzed and constrained maize domestication. Proc. Natl. Acad. Sci. 116: 5643–5652.

Zhou Y., M. Massonnet, J. S. Sanjak, D. Cantu, and B. S. Gaut, 2017 Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. Proc. Natl. Acad. Sci. 114: 11715–11720.

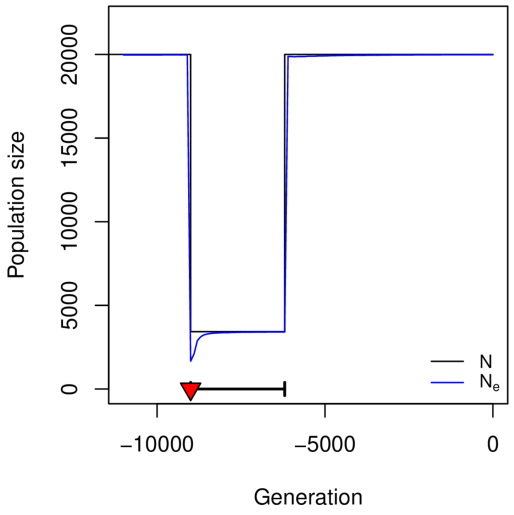

Figure S1: Dynamics of the population size during the domestication process. The variation of N was a parameter of the model (drop of the population size from N=20,000 to N=3,500 during 2,800 generations); the effective population size N_e was estimated from the variance in fitness (averaged over 1000 simulation replicates); its variation reflects the intensity of selection.

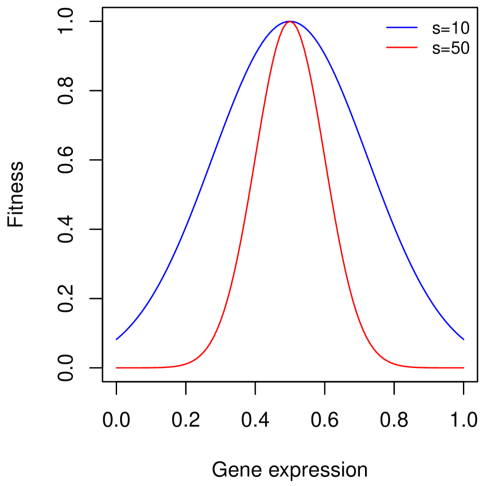

Figure S2: Illustration of the fitness function around an optimal expression of $\theta = 0.5$, for two strengths of selection (s=10, the default simulations, and s=50, the strong selection setting).

	Before domestication	After domestication		
		Plastic	Stable	Non-selected
Plastic (p)	6	2	2	2
Stable (s)	6	0	4	4
Non-selected (n)	12	0	4	8
Total	24	2	10	12

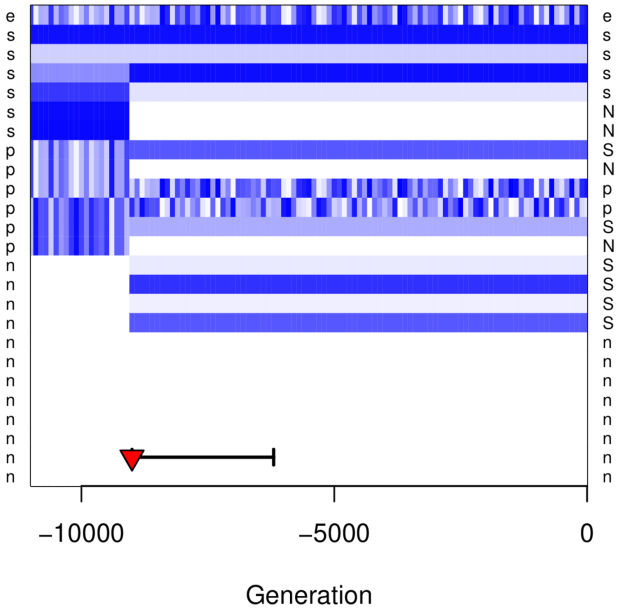

Figure S3: Description of the selection switch.Top: summary table of the selection switch. Three categories of genes were considered: “stable” (s), which optimum does not change throughout the simulation (except at the onset of domestication for half of them), “plastic” (p), which optimum tracks the environmental signal (with a +1 correlation for half the genes, and a -1 correlation for the other half), and “non-selected” genes (n), which expression is not part of the fitness function. Capital letters on the right indicate genes which status changed during the domestication process. The total number of selected genes (12 out of 24) remains the same before and after domestication. Bottom: illustration of the variation of gene expression optima from a single simulation. The figure indicates the expression of the sensor gene (first line, e), and the optimal expression for all other genes (light: low expression, dark: strong expression). White stands for unselected genes, and capital letters for genes which status changed.

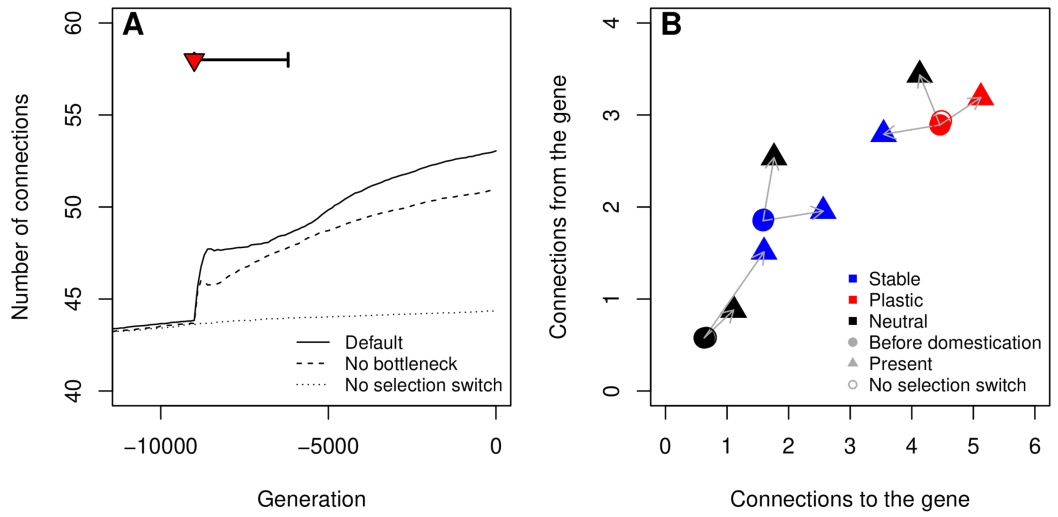

Figure S4: Evolution of the number of strong connections in the GRN through time. A: Evolution of the average number of strong connections in different evolutionary scenarios. B: Evolution of the numbers of connections to (number of regulators of the target gene) and from (number of genes regulated by the target gene) each type of gene during domestication in the default simulation (bottleneck and selection switch). The “no selection switch” scenario (open circles) is indicated as a control.

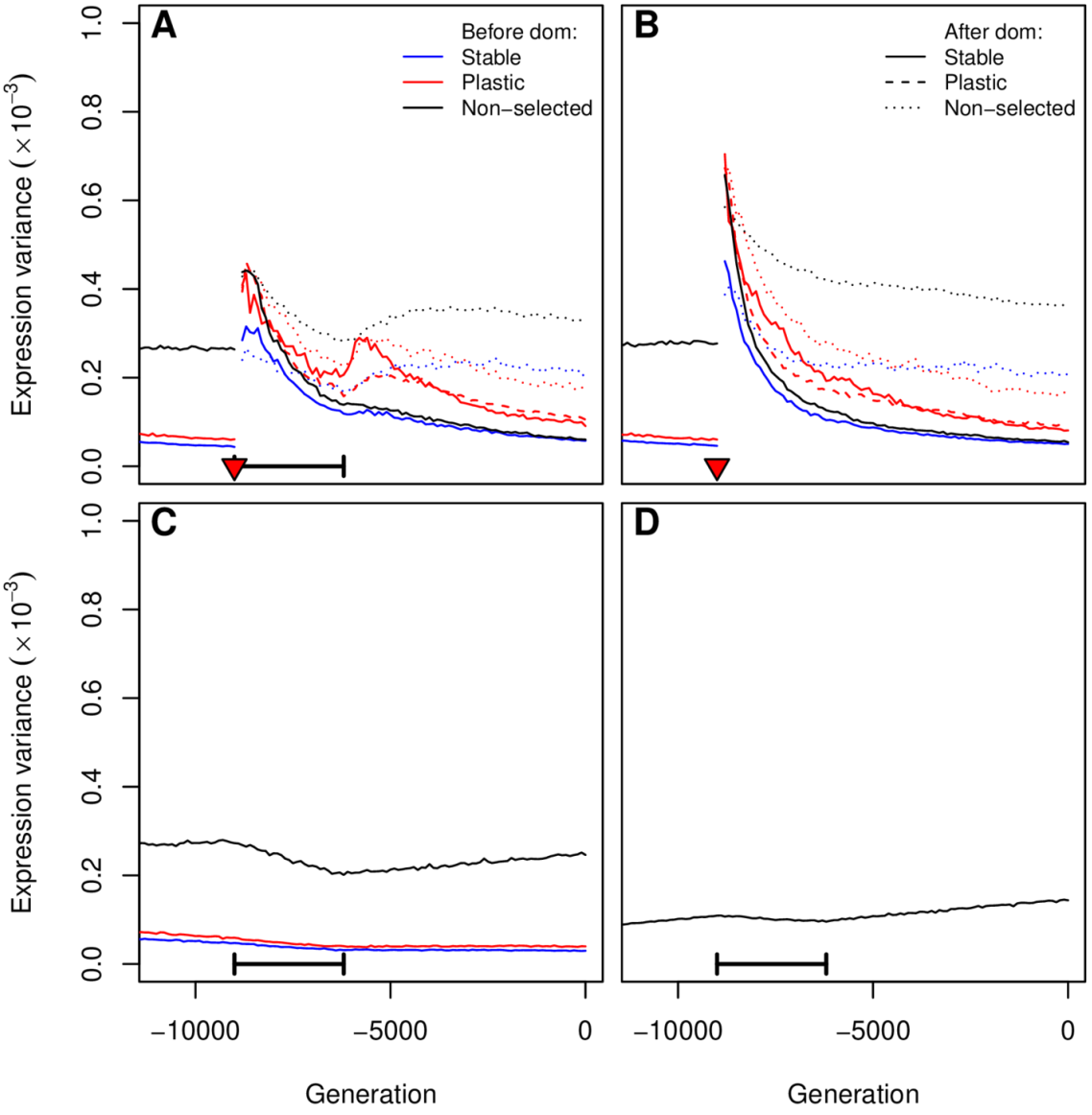

Figure S5: Evolution of expression variance through time. Genetic expression variance was averaged across 1000 simulations for every type of gene. A: Default simulations (bottleneck and selection switch), B: No bottleneck, C: No selection switch, D: no selection on the optimum. Colors indicate the gene status before domestication (constant and stable genes were merged for clarity), and the line type indicates the gene status after domestication.

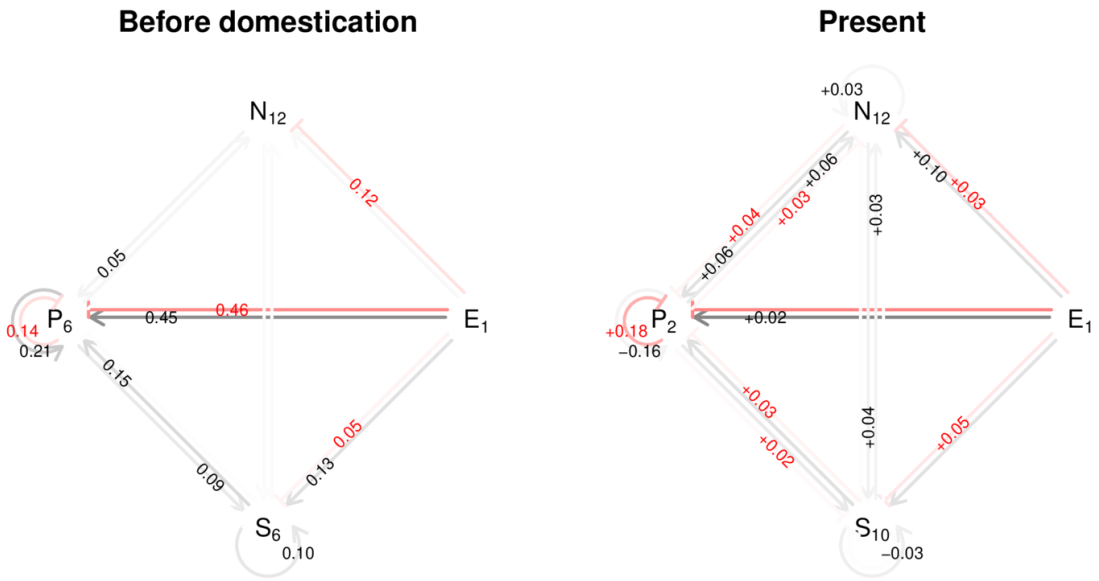

Figure S6: Summary of the topological changes in the network. The figure displays the location of the "strong" connections in the network, based on the effect of regulation on gene expression (details in the "methods" section). Three categories of genes are represented: non-selected (N), plastic (P), stable (S), and (E) stands for the gene that is directly reflecting the environment. Subscripts indicate the number of genes for each category. The left panel indicates the frequency of connections between each gene category (i.e. the number of strong connections divided by the theoretical maximum) before domestication. Right: Network at the end of the simulations. Numbers stand for the differences compared to the left panel. In both panels, connections between genes from the same category are indicated as curved arrows. Red arrows stand for negative (inhibition) effects, black arrows for positive (activation) effects. Connection frequencies less than 5% are omitted for clarity.

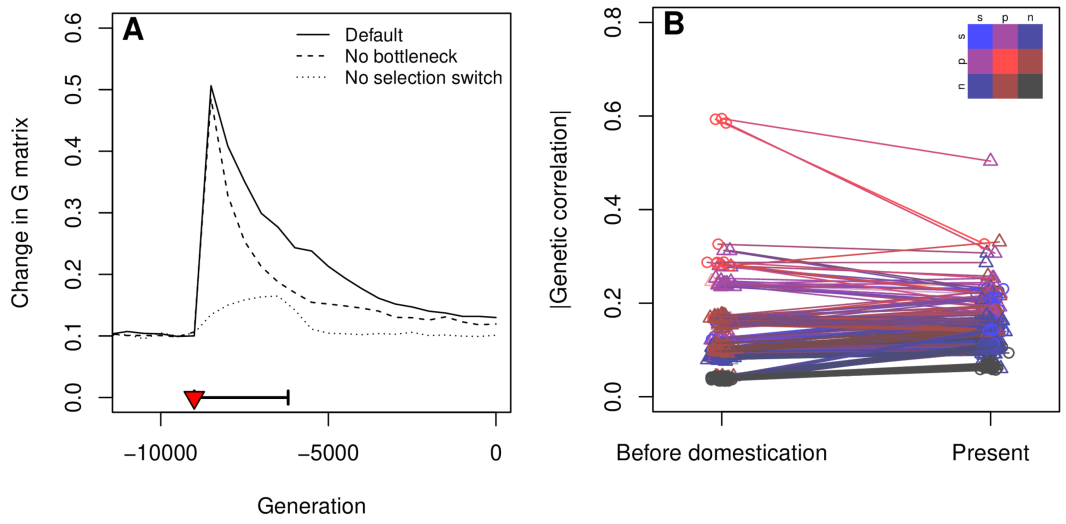

Figure S7: Evolution of the genetic coexpression structure. (A) The evolution of the genetic (G) covariance matrix of all gene expressions was tracked by measuring the distance between G every 500 generations (details in the methods section). (B) Genetic correlations before (generation -9000) and after (generation 0) domestication were compared between all pairs of genes. The color code indicates the category of the genes compared:blue for stable/constant, red for plastic, black for non-selected, and intermediate colors for comparisons between genes of different categories. Circles denotes correlations between genes from the same category, triangles from different categories. The x-axis was slightly jittered for clarity.

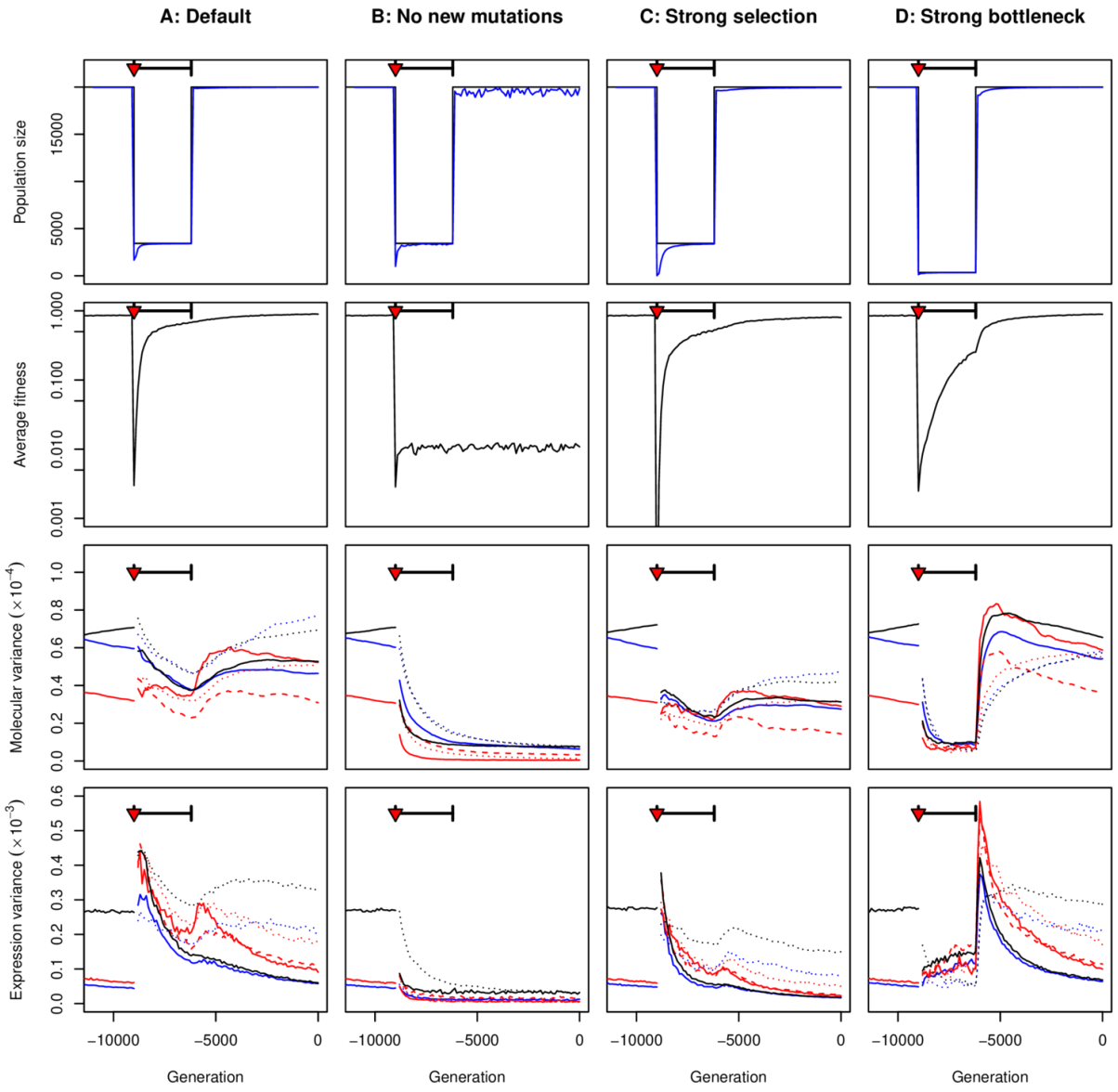

Figure S8: Sensitivity to variations in the model parameters (1 / 4). The left column (A) stands for the default parameter set. B: “No new mutations”: the response to domestication relied only on the standing genetic variation (mutation rate set to 0 at the onset of domestication); C: “Strong selection”: the selection coefficient was set to 50 instead of 10 (as illustrated in Figure S2); D: “Strong bottleneck”: the strength of the bottleneck was 10 times stronger (350 individuals instead of 3500). First row; census and effective population sizes (same caption as fig S1), second row: Average fitness (same caption as fig 1), third row: Molecular variance (same caption as fig 3A, same colors as fig S5), fourth row: expression variance (same caption as fig S5).

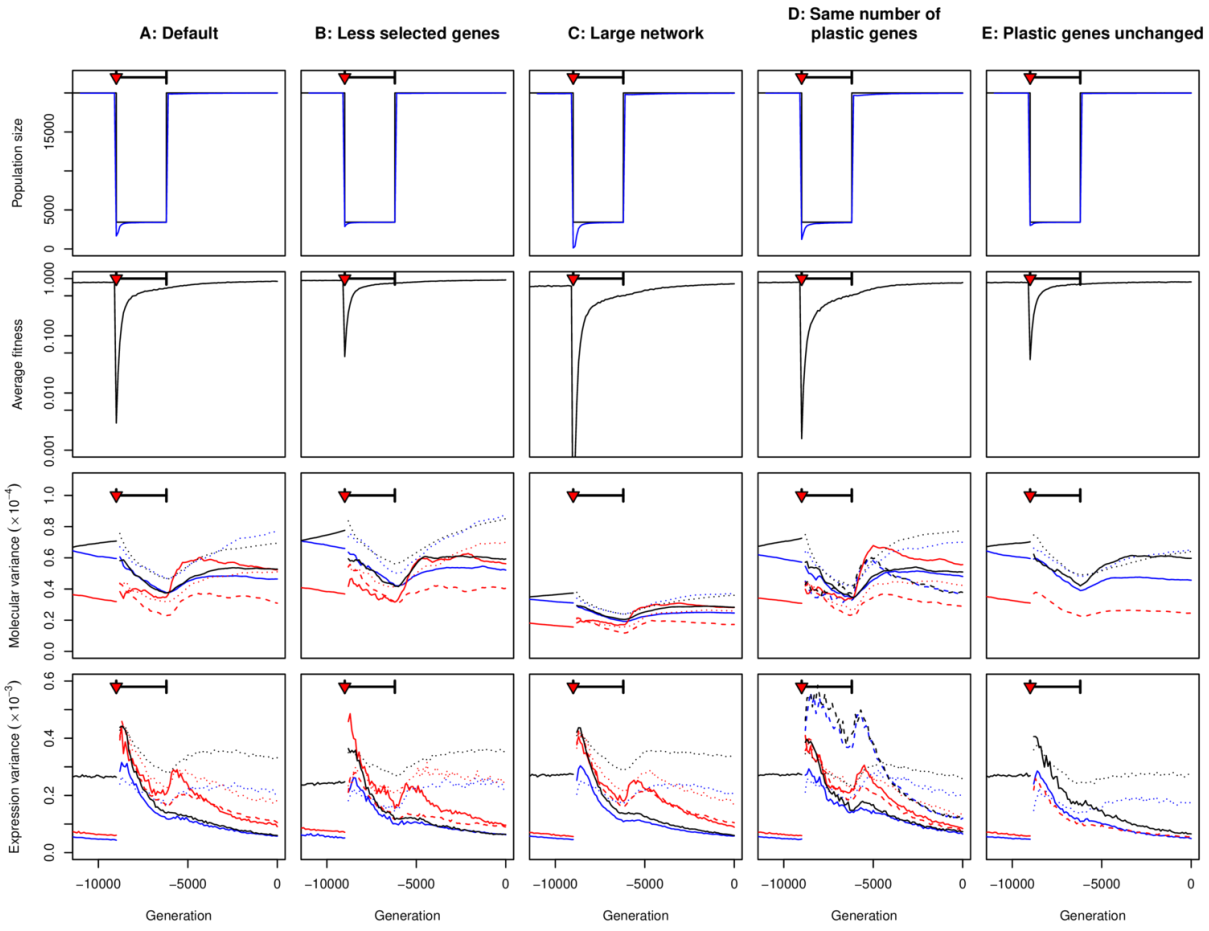

Figure S9: Sensitivity to variations in the model parameters (2 / 4).The left column (A) represents the default parameter set. B: “Less selected genes”: the network still contained 24 genes, but only 6 (instead of 12) evolved under selection, with the same proportion of plastic, stable, and non-selected genes; C: “Large network”: 48 genes (+ environment) were considered in the simulation, with 24 genes under selection. D: “Same number of plastic genes”: Domestication was not associated with less selection on plasticity; 6 genes are still selected for plasticity during the domestication process, but these were not the same as the plastic genes before domestication. E: “Plastic genes unchanged”: the six plastic genes after domestication were the same as the ones before domestication. Rows stand for the same parameters as in fig S7.

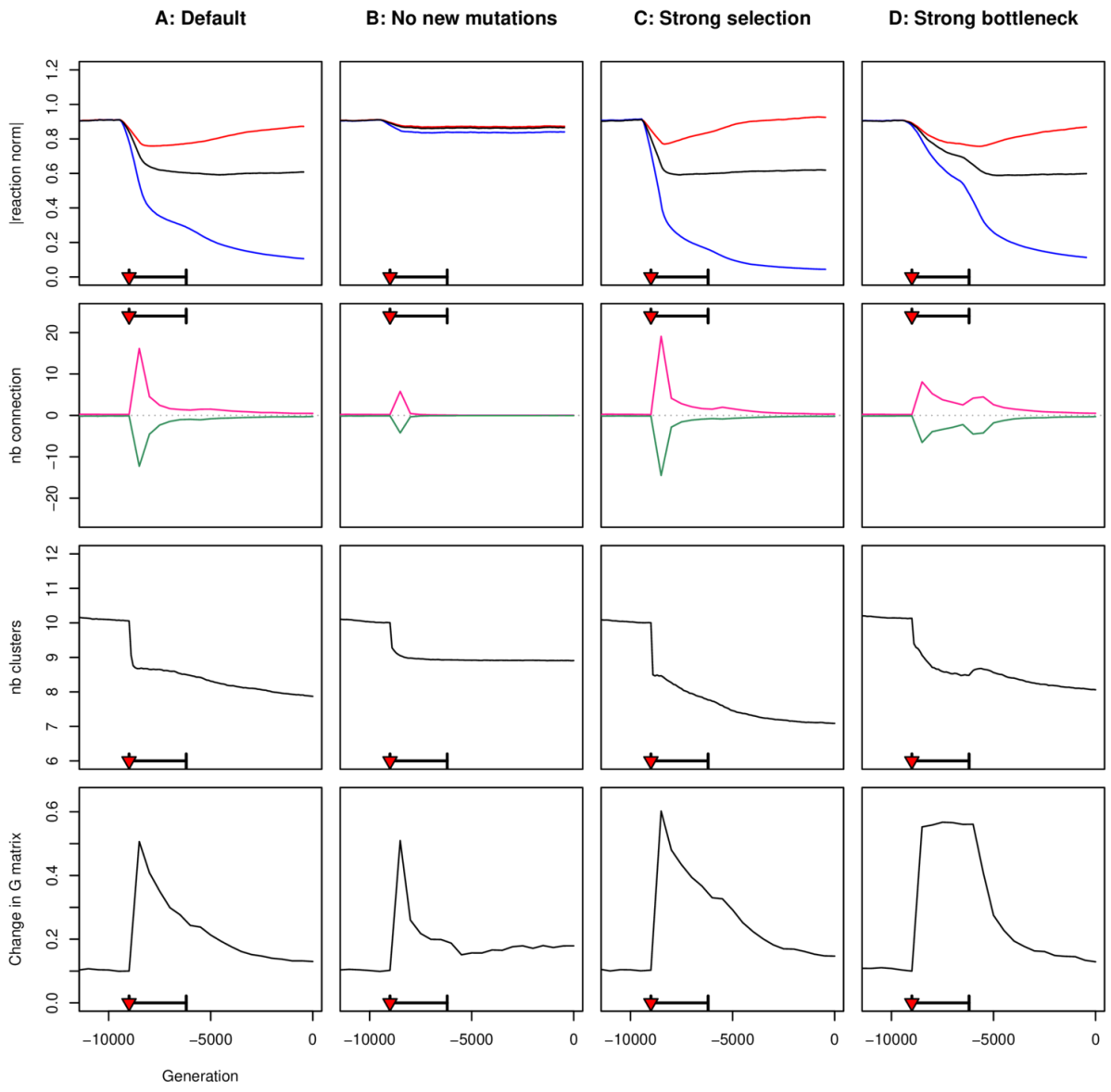

Figure S10: Sensitivity to variations in the model parameters (3 / 4). Columns are the same as in Figure S8. Rows stand for, respectively, the average absolute reaction norm (as in Figure 2B), The number of gained vs loss connections (as in Figure 5A), the average number of clusters (as in Figure 5B), and the dynamics of the change in the genetic covariance matrix (as in Figure S7A).

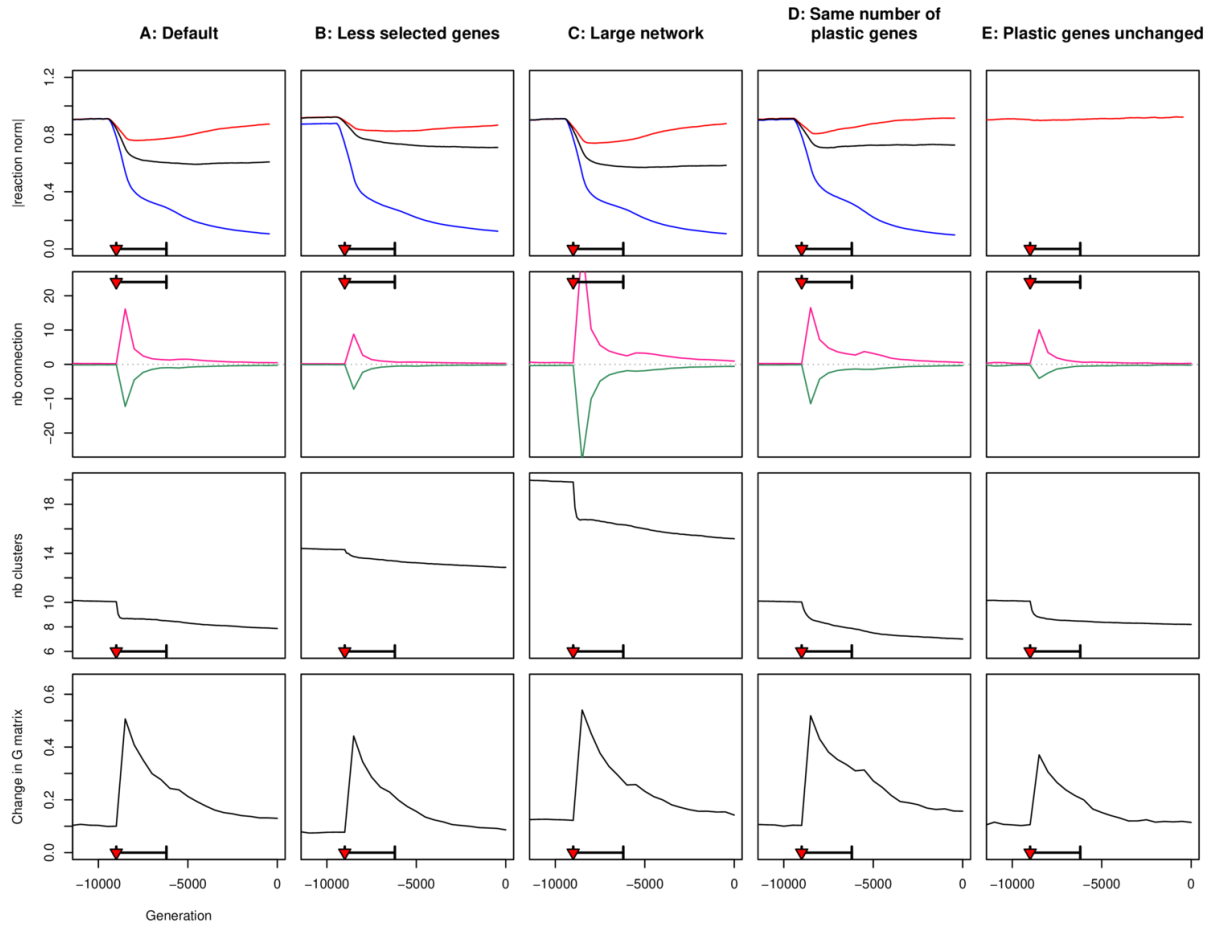

Figure S11: Sensitivity to variations in the model parameters (4 / 4). Columns are the same as in Fig S9, Rows are the same as in Fig S10.