Association between gene expression plasticity and regulatory network topology

Apolline J. R. Petit[1,*], Anne Genissel[2], and Arnaud Le Rouzic[1]

[1] *Université Paris-Saclay, CNRS, IRD, UMR EGCE, 91190, Gif-sur-Yvette, France*
[2] *Université Paris Saclay, INRAE, UR BIOGER, 91120, Palaiseau, France.*
[*] *Corresponding author: Université Paris-Saclay, CNRS, IRD, UMR EGCE, 91190, Gif-sur-Yvette, France.*
Email: apolline.petit@universite-paris-saclay.fr

Abstract

What drives the evolution of gene network topology? During the last decades, many looked into it to find structures responsible for gene expression patterns. Gene expression plasticity for instance has been associated with many network structures, but both theoretical results and empirical observations were often equivocal. Our objective was to decipher the regulatory causes of gene expression plasticity, and more specifically, which structures in regulatory networks were relevant for the sensitivity to environmental factors. We sought the common regulatory structures associated with gene expression plasticity between predictions from an evolutionary simulation model and the global regulatory network from *Escherichia coli*. The combination of empirical and theoretical approaches and their strikingly similar results confirmed that selection promotes more regulation towards plastic genes and, as a consequence, plastic genes were more often regulated by feedforward loops than non-plastic genes. Selection tends to bias the distribution of regulatory loop motifs towards positive feedforward and diamond loops, but this enrichment in specific structures was the same in plastic and non-plastic genes. The impossibility to predict gene expression plasticity from the network regulatory structure opens interesting questions about the nature of the missing information in current systems biology databases.

Significance statement

Organisms can respond to their environment, and some of their traits are said to be plastic, i.e. dependent on environmental signals. At the individual level, understanding how gene expression varies in response to the environment is of critical importance. Some argued that this sensitivity could be deduced from the structure and motifs of gene regulatory networks, but theoretical and empirical studies could not reach a consensus. Here, we aim at disentangling the relationship between regulatory structure and gene expression plasticity, combining empirical data of *Escherichia coli* and population evolution modelling. Both approaches agree on a larger amount of regulators for plastic genes compared to non-plastic genes. There were clear signals of selection enriching networks in specific structures, such as feedforward loops, but this selection was independent from gene expression plasticity.

Introduction

Understanding the mechanisms involved in the evolution of regulatory networks remains a major challenge in systems biology. It is widely acknowledged that biological gene networks are not random [1, 2], but the interplay of natural selection, mutation bias, functional constraints, and genetic drift in the evolution of network topologies is not clearly established. Some regulatory structures, such as feedforward loops (FFLs) and feedback loops (FBLs) have attracted attention because of their abundance in biological systems [3, 4]. This enrichment could be the consequence of selection, indirectly promoting motifs through their effect on gene expression patterns, including stability, plasticity, or oscillation.

Changes in genes expression is a molecular manifestation of phenotypic plasticity and it conditions the tempo and mode of the organism's response to environmental cues. It is now feasible to measure

these changes at a large scale, by identifying differentially-expressed genes using RNA sequencing. At the gene level, plasticity is contingent on signal-sensitive regulation of transcription. Natural or artificial selection for plasticity might indirectly affect the underlying regulatory network structure. Yet, both theoretical and empirical work falls short of consensus when linking gene expression patterns with specific network topologies. From the theoretical side, feedforward loops [1, 5, 6, 7, 8], diamond loops [7, 8], and feedback loops [9, 10] have been reported to be involved in plastic (environmentally-sensitive) gene expression. Counterintuitively, other theoretical studies associate those same loop categories with robustness-related functions (FFLs: [11, 12, 13], FBLs: [14]). From the empirical side, two studies on synthetic bacteria [5] and *Escherichia coli* [15] brought experimental evidence of a link between feedforward loops and sensitivity to environmental signals. However these studies were restricted to a few genes, and the genome-wide picture remains unknown [16]. Overall, there is no clear evidence that specific network structures are involved in the regulation of plastic genes.

The structure of gene regulation can be described at several levels, i.e., the number of regulators, the number and the shape of regulatory loops (feedback, feedforward, or diamond) pointing to a gene, or the loop motif (the arrangement of activation and inhibition patterns within a loop). In regulatory networks, which of these levels — if any — is relevant for gene expression plasticity? Our ambition here is to tackle this question with a dual approach, coupling a theoretical model with a database analysis in order to derive consistent predictions and observations. For this purpose, we wish to (i) quantify the number of regulators and regulatory loops targeting plastic and non-plastic genes, and (ii) determine whether loop categories and loop motifs are relevant to predict plastic gene expression. Plastic genes were collected from 11 transcriptomic studies in which *E. coli* populations were submitted to several environmental treatments. The local regulatory structure of these genes was extracted from a transcriptional regulation database. Theoretical predictions were derived from evolutionary simulations, in which the regulatory interactions were free to evolve under a mutation-drift-selection process in populations. Gene expression plasticity was then measured as the response of individual genes to a "sensor", environmentally-induced gene. Both simulated and empirical datasets were analysed with the same statistical pipeline, allowing for direct comparison.

Results

The *E. coli* regulatory network was built from Ecocyc database [17] (see Fig. suppl. 1 for a graphic representation of the network). Regulations were encoded as qualitative factors ("activating" or "inhibiting"). Independently from the Ecocyc regulation data, a list of environmentally-dependent ("plastic") and independent ("non-plastic") genes was established based on 11 studies reporting differentially-expressed genes in response to various environmental changes. Out of the 4638 genes annotated in the reference annotation (RefSeq: GCF_000005845.2), 2144 genes were never reported as differentially expressed in these 11 studies, and were thus considered as "non-plastic". 1074 genes were differentially expressed in at least two studies, and therefore considered as "plastic" (Suppl. Table 1). The 1420 genes reported only once as differentially expressed were excluded from the analysis.

Computer simulations were performed by coupling a dynamic gene network model for the genotype-phenotype map, and an individual-based model for the population genetics [18, 19]. Equilibrium gene expression was assimilated with a multivariate phenotypic trait, on which stabilizing selection was applied. The genotype was encoded as an interaction matrix across 36 genes: 1 sensor gene, informing the network about the state of the environment, 10 transcription factors, able to affect the transcription of all other genes, 5 neutral genes, 20 selected genes (10 stable genes, which optimal expression was constant across environments, and 10 plastic genes, which optimal expression depended on the environmental cue, fluctuating randomly every generation). Information about the environment was thus integrated into the gene regulatory network and could affect all gene expressions through cascading regulations. Populations of 20,000 clonal haploid individuals were submitted to 4,000 generations of selection, and the average regulatory network at the end of the simulations was used to reconstruct discrete gene networks, in which each gene-gene interaction could be "activating", "inhibiting", or "null". Genes were classified as "plastic" or "non-plastic" based on the amplitude of their response to the environmental cue; about 73% of the classified genes were plastic.

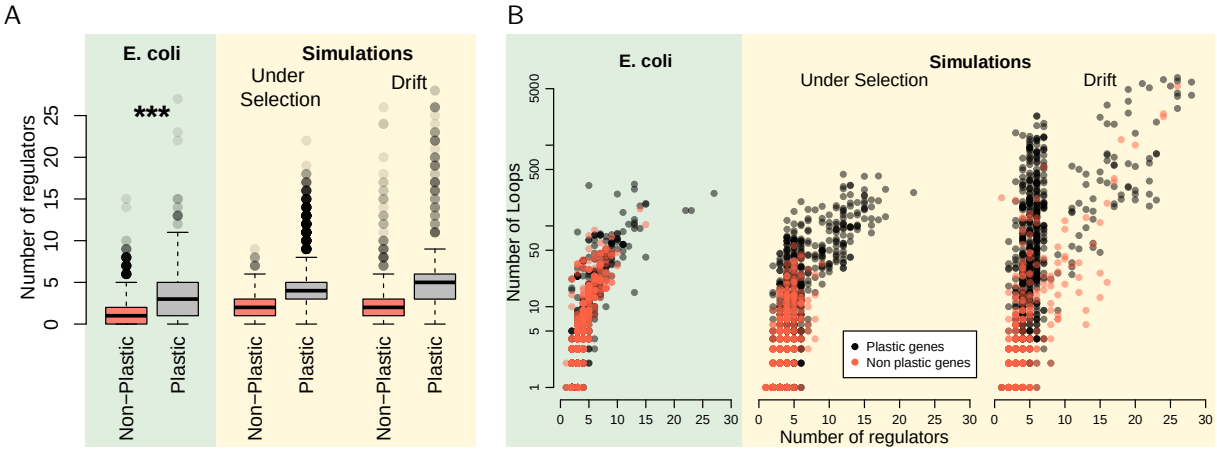


Figure 1: Regulation of plastic and non-plastic genes. A) Distribution of the number of regulations toward the genes. Stars indicate t-test p-values lower than 0.001 (*E. coli* data only). B) Number of feedforward, diamond, and feedback loops by which genes were regulated (only genes regulated by at least one loop are displayed). Simulation results have been downsampled so that the same amount of data points are displayed in all panels.

Regulation of plastic *vs.* non-plastic genes

We first considered the number of regulatory interactions towards plastic and non-plastic genes. Plastic genes were influenced by more regulators than non-plastic genes, both in *E. coli* and in simulations (Fig. 1A). In *E. coli*, plastic genes were regulated by 3.2 genes on average (4.6 in simulations), while non-plastic genes were regulated by 1.3 genes on average (2.4 in simulations). The number of feedforward, diamond, and feedback loops targeting plastic genes was also larger than to non-plastic genes in both *E. coli* and in simulations, including in control simulations where mutations could accumulate freely with genetic drift (Fig. 1B). In absence of selection (drift simulations), genes that happened to evolve a plastic response were also characterized by a larger number of regulators, suggesting that plasticity tends to be passively induced by a larger number of regulators. In these drift simulations, there were more regulatory loops for a specific number of regulators than in regular simulations, showing that there was no positive selection on the number of loops, which even appears to be selected against when too high.

Regulatory loops

In *E. coli*, 64% of plastic genes (Fig. 2A), and only 30% of non-plastic genes (Fig. 2C), were regulated by at least one of the loop categories considered (*i.e.* feedforward, feedback, and diamonds, corresponding to the three loops that have been associated with plasticity in the literature). The most common observed pattern for *E. coli* genes was the regulation by both FFLs and DMDs, with an almost even distribution of both loop categories. Feedback loops (FBLs) were very rare in *E. coli* (68 genes regulated by FBLs out of 3214). Overall, the proportions of genes regulated by FFLs, DMDs, FBLs, and any combinations of thereof were strikingly similar between plastic and non-plastic genes, and none of these loop categories appeared to be preferentially associated with plastic gene expression.

In simulated gene networks, plastic genes were also more often regulated by feedforward, feedback, or diamond loops than non-plastic genes. Most genes were targeted by a several categories of loops, and the proportion of FFLs *vs.* DMDs was even (Fig. 2B and D). Overall, there was no clear difference in the relative proportion of FFL and DMD loops involved in the regulation of plastic *vs.* non-plastic genes. Contrary to our findings on *E. coli*, feedback loops were common in simulated networks only for plastic genes, and always associated with FFL and DMD loops.

Loop motif frequencies

Loop categories remains a coarse-grained classification, as the precise regulatory function necessarily depends on the sign and the coherence of interactions. We have thus determined the motif frequencies

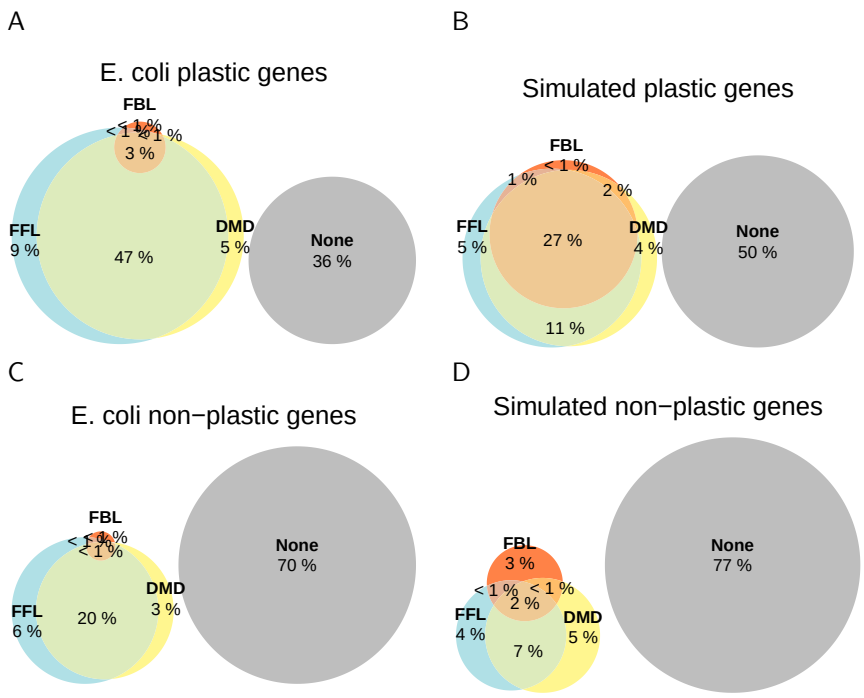


Figure 2: Genes classified by incoming regulation loops. A) Plastic genes in *E. coli* B) plastic genes in simulations, C) non-plastic genes in *E. coli*, D) non-plastic genes in simulations. FFL: FeedForward loops, FBL: FeedBack Loops, DMD: Diamond motifs. "None" stands for genes which regulators were unknown, or which were regulated by another motif (no loops, or large motifs involving more than four genes).

within FFL and DMD loops (8 motifs for FFLs, Table 1, and 10 motifs for DMD, Table 2), both for the *E. coli* dataset and for simulated networks. Feedback loops were virtually absent in *E. coli* networks, and thus not included in Fig. 3A.

In both *E. coli* and simulated gene networks, the most frequent feedforward motif was C1 (coherent, two genes activating), followed by I1 (incoherent, one gene activating and the other inhibiting) (Fig. 3A). Other motifs were found in similar proportions, with the exception of I3, frequent in simulations but almost absent in *E. coli*. In diamond loops, the most frequent motifs in both *E. coli* and simulations were PP (activation by both regulating genes) and PM (activation by one gene, inhibition by the other one). These structures were strikingly similar to the major feedforward motifs. The frequency of motif was similar through all studies used to identify plastic genes (Fig. supp. 3). Motifs involving negative regulations (NN, NM, and NP) tended to be more frequent in *E. coli* than in simulated networks; but overall, the motif distribution remained quite similar.

In contrast, the motif frequency distribution resulting from an evolutionary process, both in simulations and in *E. coli*, was substantially different from the randomly shuffled networks. In particular, the most frequent motifs (C1 and I3 for FFLs, PP and PM in DMDs) were enriched in evolved networks compared to the shuffled and drift controls. This enrichment was a consequence of selection in simulations, and it is reasonable to postulate that the same explanation holds for *E. coli*.

While there was a clear contrast between control and selected networks, differences in motif composition between plastic and non-plastic genes were minor for *E. coli*. Statistical tests in *E. coli* (Fisher's exact test) did not support differences between plastic and non-plastic genes. The contrast between plastic and non-plastic genes was slightly larger in simulations, particularly in the frequency of the FFL motif C1, C2, I3, and the DMD motifs PM. None of these differences matched those observed in *E. coli*. They are not coherent either with the differences between plastic and non-plastic genes observed in drift simulations.

In a Principal Component Analysis on the relative proportion of each motifs of the three loop categories (Fig. 3B), PC1 (55% of the variance) was associated with the contrast between selected and random networks, while PC2 (19% of the variance) was discriminating *E. coli* vs. simulations. Plastic

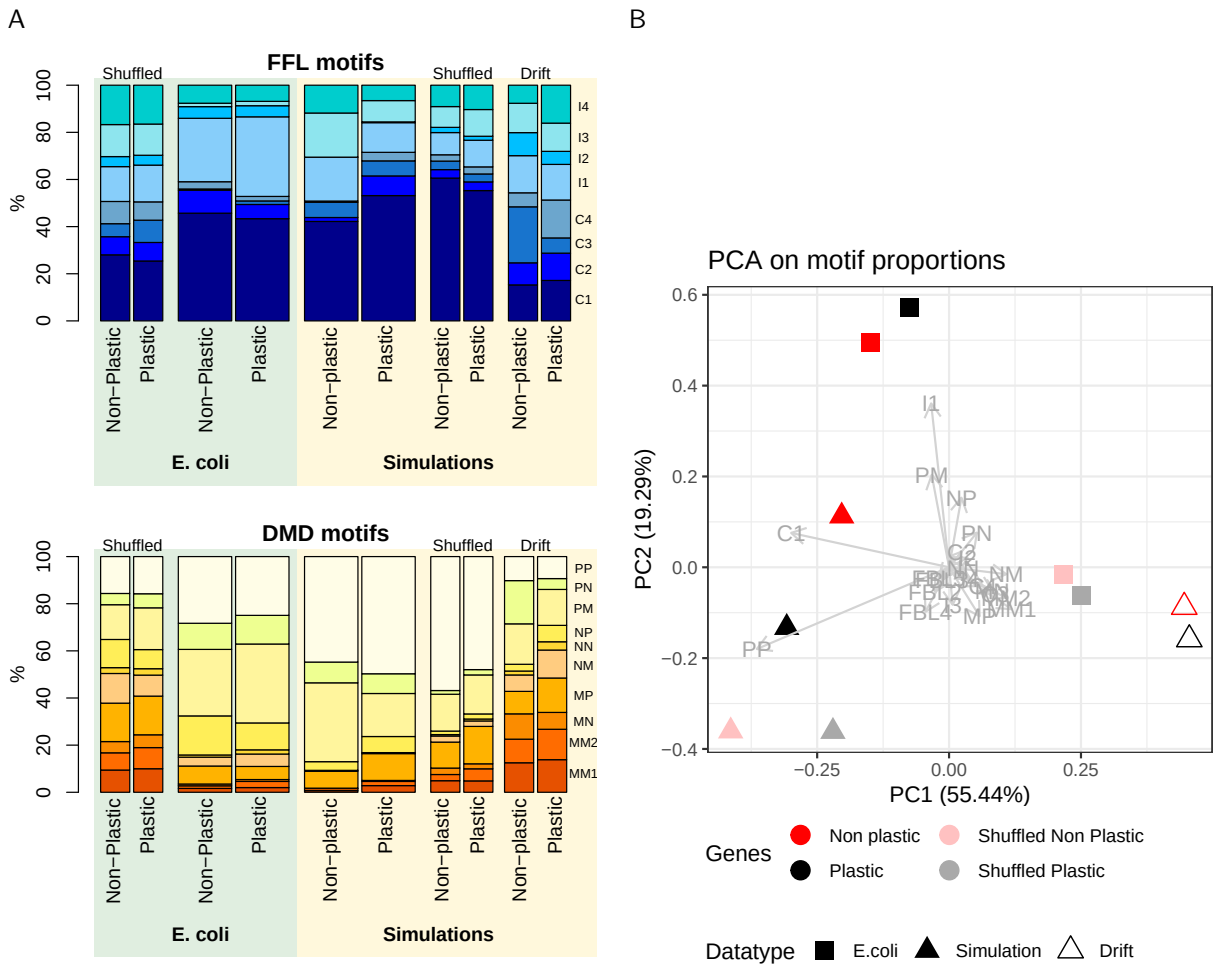


Figure 3: Loop motifs regulating plastic and non-plastic genes. A) Loop motif frequencies within feedforward (top) and diamond (bottom) loop categories. Motif nomenclature matches Tables 1 and 2. Motif frequencies were normalized per gene (i.e., all genes have the same weight in the figure, irrespective of the number of regulatory motifs targeting them). Shuffled controls were obtained by permuting randomly all non-null interactions. Simulations runs in absence of selection are indicated as "drift". B) Principal Component Analysis on loop motif frequencies (FFL, DMD and FBL).

and non-plastic genes from the same conditions were found closer to each other than any other points, reinforcing that the contribution of plasticity to the distribution of motif frequency was minor.

Discussion

Our objective was to identify the gene regulatory structures involved in gene expression plasticity, from both evolutionary simulations and empirical regulation networks. We found that plastic genes indeed had specific regulatory characteristics: genes which expression depends on the environment had more regulators, and were the target of more regulatory loops. The analysis of the frequency of different kinds of regulatory loops and regulatory motifs showed that both real and simulated networks were not random, as a consequence of selection on gene expression levels. Yet, there were no plasticity-specific regulatory motifs, suggesting that environmental sensitivity is not encoded into the network topology.

Theoretical and empirical results strikingly similar Systems biology is rather convincing at describing general patterns in biological networks, based on measurement tools derived from graph theory [20, 21]. For instance, it is often admitted that biological networks tend to be scale-free [22],

relatively scarce [23], and modular [24]. Yet, there is a surprising lack of theory from which such patterns could be interpreted. Typically, the extent to which network evolution is driven by mutation bias, genetic drift, adaptation, or indirect selection remains widely unknown [25, 26, 27, 28]. Here, we show how a simplified gene network evolutionary model, generally exploited to address genetic architecture-related questions, can predict empirical patterns with a striking accuracy: our simulations could retrieve, both qualitatively and quantitatively, the larger number of regulatory genes for plastic genes, motif enrichment and depletion in selected *vs.* randomized networks, the most abundant motifs, and the relationship between the number of regulation and number of motifs. This validates the capacity of such simple models (in which gene regulations are encoded in an interaction matrix [18, 29]) to contribute to the theoretical corpus on gene network evolution, in complement to the more cumbersome physiologically-realistic models [8].

Nevertheless, we noticed a few differences between simulations and empirical patterns, that might be worthwhile to explore. The only obvious qualitative difference between both approaches was the ∼10-fold increase in the frequency of feedback loops in simulations compared to *E. coli* regulation network. Although feedback regulation exists in *E. coli* and is known to be involved into noise reduction or bistability [30], feedback loops happen to be outnumbered by feedforward loops in the Ecocyc transcriptional dataset and in pathway databases in general [31]. A possible explanation lies in the fact that most feedback mechanisms in prokaryotes rely on autoregulation (in *E. coli*, 50% of TFs self-regulate, [32]), while the minimum size for a motif in our analysis was set to 2 genes. A more subtle discrepancy between theoretical and empirical networks is that simulations predicted a small but systematic difference between motif frequencies in plastic vs. non-plastic genes, which was not statistically supported in *E. coli*. Insufficient statistical power could partly explain this observation, but more generally, the classification of *E. coli* genes into plastic and non-plastic categories remained coarse-grained, and may not match the same biological properties: simulations consider the slope of the reaction norm to a unique environmental factor, while the *E. coli* analysis relies on the sensitivity to several stressors, which might not reflect exactly the same mechanisms. The observed convergence between theoretical and empirical observations in spite of these different measurements of plasticity strengthen our conclusions.

Plastic genes have more regulators We identified regulatory differences between plastic genes and non-plastic genes in both real and simulated data. Plastic genes were regulated by more genes, and were the target of more regulatory loops (feedforward loops, diamond loops, and feedback loops). These results reinforces previous observations, which associated the number of regulators with the complexity of the gene expression patterns [33, 34]. Nevertheless, the distributions of the number of regulators and the number of regulatory loops largely overlap between plastic and non-plastic genes, both in simulated and in empirical networks. As a consequence, the number of regulators remains a poor predictor of gene expression plasticity.

A direct benefit of using computer simulations to complement the analysis of empirical network lies in the possibility to go beyond association statistics, and to investigate cause-consequence relationships. We showed that (i) plasticity was statistically associated with a higher number of regulators, even in the drift simulations, (ii) selection for plasticity increased the number of regulators, (iii) selection decreased the number of regulatory loops while keeping a high number of regulators. We propose the following scenario for the changes in the regulatory networks associated with the evolution of plasticity: selection for plastic gene expression directly promotes the recruitment of cis-regulators, which indirectly increases the number of regulatory loops.

If some regulatory motifs were outstandingly more present in plastic genes than non-plastic genes, it would have been possible to confidently associate a plastic function to those motifs. However, no differences between the motifs regulating plastic genes and the motifs regulating non-plastic genes were found in *E. coli*. Yet loop motifs were non-randomly distributed, and our simulations confirmed that this pattern was due to selection on gene expression. The selection pressure on motifs thus appears to be identical for plastic and non-plastic genes. Given our findings and the conflicting literature on this topic, we find it likely that there is no regulation motifs with a dedicated "plasticity" function.

Motifs evolve due to indirect selection, i.e., as a consequence of selection on gene expression. The *E. coli* motif distribution we found was coherent with previous work [35, 36], who also identified C1 and I1 as the most represented motifs. The high frequency of activating motifs like C1 for FF loops or PP for DMD loops could be a statistical consequence of the uneven balance between activation and

inhibition within the network. C1 and PP are indeed very rare in drift simulations, featured by an equal number of activation and inhibition interactions. Yet, arguments related to activation/inhibition ratios are fragile, because it is difficult to discard that the uneven activation/inhibition balance could be a consequence, rather than a cause, of an evolutionary bias towards specific motifs. In any case, both C1 and PP were over-represented in *E. coli* compared to randomly-permuted networks, suggesting some specific preference for such motifs. An attractive hypothesis can be derived from the fact that C1 and I1 have been predicted to be efficient at buffering expression noise [37, 11]. Selection on noise filtering and gene expression robustness could thus enrich gene networks in some specific motifs like C1 and I1. This hypothesis could not be confirmed by our simulations, in which the gene network dynamics was deterministic, but alternative theoretical settings [38, 39] could be considered.

Conclusion Our analysis revealed the limitations of regulatory network motifs in explaining environment-sensitive gene expression. We observed no topological feature associated with gene expression plasticity beyond a larger number of incoming regulations in average, so that the plasticity of a given gene cannot be deduced solely from the properties of its local regulation network.

If not the network motif, which regulatory features could reveal whether the expression of a specific gene has been selected for robustness, stability, cyclic expression, or sensitivity to an environmental or physiological input? To what extent is it possible to predict gene expression profiles from pathway databases [40, 41]? Three major hypotheses can be proposed: (i) the information about regulatory structures involved in plasticity is present in existing databases, but they are not caught by known descriptors of network structures and properties (loops, motifs, or graph metrics). For instance, it might depend on a more general regulatory context [42, 43], or on complex combinations of several intertwined network structures [44]. Alternatively, (ii) regulation databases remain the correct approach, but existing databases are not precise nor exhaustive enough. Typically, presence/absence data for regulatory interactions might be insufficient, and quantitative estimation of interaction strengths (weighted edges in the network graph [45]) could be necessary as suggested by other [46, 10, 47]. In contrast, hypothesis (iii) presumes that regulation might be too complex to be abstracted in regulatory networks, which omit important epigenetic and cellular mechanisms, including the spatial organisation of the genome, methylation, phosphorylation, protein conformation, or "missing regulation" from intergenic sequences [48]. The next challenge that the systems biology community will have to face might thus not to generate more data, but to identify which data is relevant to understand the evolutionary properties of biological systems, a task for which theoretical approaches might bring invaluable insights.

Material and Methods

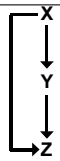
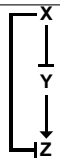
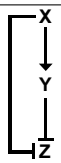
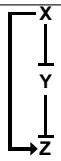
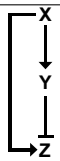
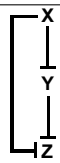
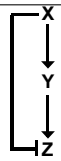
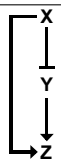
Coherent				
Nomenclature Abbreviation	Coherent 1 C1	Coherent 2 C2	Coherent 3 C3	Coherent 4 C4
Incoherent				
Nomenclature Abbreviation	Incoherent 1 I1	Incoherent 2 I2	Incoherent 3 I3	Incoherent 4 I4

Table 1: Three-genes feedforward loop nomenclature.

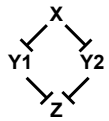
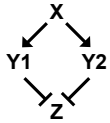
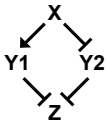
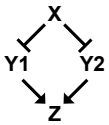
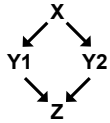
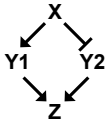
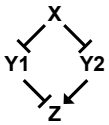
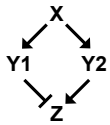
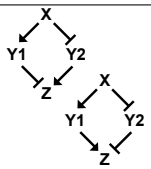
$X \rightarrow Y_n$ Sign	Inhibition	Activation	Mixed
Negative $Y_n \rightarrow Z$			
Nomenclature Abbreviation	Negative negative NN	Positive negative PN	Mixed negative MN
Positive $Y_n \rightarrow Z$			
Nomenclature Abbreviation	Negative positive NP	Positive positive PP	Mixed positive MP
Mixed $Y_n \rightarrow Z$			
Nomenclature Abbreviation	Negative mixed NM	Positive mixed PM	Mixed mixed 1 and 2 MM1 ; MM2

Table 2: Diamond motifs nomenclature.

Empirical data: *E. coli*

We used *E. coli* K-12 substr. MG1655 annotation (RefSeq: GCF_000005845.2). *E. coli* K-12 known regulations have been obtained from Ecocyc ([17], downloaded 29/01/2024). This database contained 9534 regulatory interactions.

To conduct our analyses, we established a list of "plastic" genes, to be contrasted to "non-plastic" genes. We selected 11 studies explicitly testing environmental conditions on gene expression in *E. coli*, and proposing lists of differentially-expressed genes. The environmental factors were pH [49, 50], growth medium (urine, gluconate, and lysogeny broth) [51], aerosolization and liquid suspension [52], exposure to apple juice [53], magnesium stress, sodium stress and carbon source [54], temperature [55, 56], metabolic stress (low pH, low oxygen, low nutrients) [57], oxydative stress [58], and amino acid starvation [59]. Genes were considered as plastic when reported as differentially expressed in at least two independent studies. Genes were considered as "non-plastic" if they were never reported as differentially expressed in these studies.

Gene regulatory network model

We used a quantitative gene regulatory network model, inspired from A. Wagner's seminar work on gene expression robustness and plasticity [18, 19]. This setting, coupling a traditional population genetics model with a gene network model as a genotype-to-phenotype map, was particularly popular to explore the evolution of complex genetic architectures [60, 61, 28].

Gene network The gene network model was dynamic with discrete time-steps. Genotype consisted in $n = 36$ genes encoding a gene network represented by a $n \times n$ gene regulatory matrix $\mathbf{W}$. Genes were expressed quantitatively between $P = 0$ (no expression) and $P = 1$ (maximum expression). Gene expression was computed dynamically: initial gene expressions were set to their basal level (constitutive expression) $\kappa = 0.15$, representing the gene expression in absence of regulation, and updated as a function of the concentration of the other genes in the network during 24 time steps:

$$\mathbf{P}_{t+1} = F(\mathbf{W}\mathbf{P}_t), \tag{1}$$

where $\mathbf{P}$ was a vector of n quantitative gene expressions scaled between 0 and 1 by a multivariate sigmoid function $F(x_1, \ldots, x_n) = (f(x_1), \ldots, f(x_n))$, with $f(x) = [1 + (1/\kappa - 1)\exp(-x/\kappa(1-\kappa))]^{-1}$ as in [62].

Individuals were characterized by the average expression of their n genes during the four last time-steps $\bar{\mathbf{P}} = [\bar{P}_1, \ldots, \bar{P}_i, \ldots, \bar{P}_n]$, with $\bar{P}_i = \sum_{t=21}^{24} P_{it}$.

Reproduction, mutation, and selection Populations consisted in 20,000 haploid, clonally-reproducing individuals, matching approximately *E. coli*'s life cycle. Simulations were carried on for 4,000 non-overlapping generations.

The genotype $\mathbf{W}$ was affected by mutations of the regulation strength W_{ij}, at a rate of $\mu = 0.15$ per individual. Each gene had the same probability μ/n to be affected by a mutation. A mutated gene i had a random regulation W_{ij} changed by a Gaussian random number of standard deviation 0.05. This setting (many mutations of small effect) ensured that the distribution of alleles at all loci was approximately Gaussian.

Successive generations were produced by drawing clones with a probability proportional to their relative fitness. The individual fitness w was the product of two components: $w = w_e \times w_s$. The first term w_e was determined by a stabilizing bell-shaped fitness function and decreased with the distance to the gene expression optimum $\boldsymbol{\theta}$ [63]: $w_e = \exp(-\sum_{i=1}^{n} s_i(\overline{\mathrm{P}_i} - \theta_i)^2)$. Unless specified otherwise, the selection coefficient was $s_i = 0$ for "unselected" genes, and $s_i = 10$ for "selected" genes. The second term w_s corresponds to the selection for network stability, calculated as $w_s = \exp(-s' \sum_i^n V_i)$, where s' quantifies the selection against unstable gene expression, and $V_i = (1/4)\sum_{t=21}^{24}(\mathrm{P}_{it} - \overline{\mathrm{P}_i})^2$ was the variance in the expression of gene i during the last 4 steps of the network dynamics. We set $s' = 46,000$, as in [62, 28], which was a large penalty, reflecting the common idea that unstable networks are not viable [64].

Fluctuating selection and plasticity In our simulations, the environment was modelled as a numerical value E_g changing randomly every generation g, drawn from an uniform distribution $U(0.15, 0.85)$. The state of the environment was transmitted to the network via a sensor gene, as in [65, 66]. The model implementation follows [67], the expression of the sensor gene was constant $P_s = E_g$ during all the network dynamics. The sensor gene was free to evolve regulations toward any genes of the network, but it was not regulated, as its expression depends solely of the environment.

The 35 remaining network genes were divided in 4 categories. 10 genes were "Transcription factors" not subjected to direct selection; 10 "Plastic Target Genes" were directly selected to have their expression correlated to the environment; 10 "Non-plastic Target Genes" were directly selected to have a constant expression; and 5 unselected "control genes" were let free to evolve (Table 3).

Selection on plastic genes was modelled by changing the gene expression optimum every generation g as a function of the environment E_g, as $\theta_i = a_i E_g + b_i$. The intercept b_i and the slope a_i were randomly drawn for each gene i at the beginning of each simulation under the constraint that gene expression had to remain between 0 and 1 for any $E_g \in (0.15, 0.85)$, and stayed constant thorough generations. "Non-plastic" genes had their constant optimum drawn from a uniform distribution $U(0.15, 0.85)$.

The reaction norm is the function describing the trait response to a quantitative environmental variation. We measured the expression plasticity of a gene i as the slope β_i of the linear regression between a set of environments and the corresponding gene expression $\bar{P}_i$. Simulated genes were classified into the category "plastic genes" and "non-plastic genes" according to their reaction norm slope. Genes with $\beta_i \geq 0.5$ were considered as plastic and genes with $\beta_i \leq 0.025$ as non-plastic. Genes outside of these categories were not kept during the analysis.

Analysis

We used *E. coli* Ecocyc regulation data to construct a discrete 4638-gene network as a 4638×4638 square regulatory matrix $\mathbf{Ec}$. A cell $\mathbf{Ec}_{ij} \in \{-1, 0, 1\}$ represents the regulation of gene j on i. Regulations that were both activating and inhibiting in the Ecocyc dataset were considered as null.

To study the regulatory motifs responsible for plasticity in the simulated networks with the same pipeline than for *E. coli*, we discretized the regulation matrix $\mathbf{W}$ into a matrix $\mathbf{Et}$, based on the concept of "essentiality" [2]. We tested one by one every element of the $\mathbf{W}$ matrix by changing their value to

	Sensor gene	Transcription factor (TF)	Plastic Target gene	Constant Target gene	Neutral gene
Selection s_i	0	0	10	10	0
Optimum θ_i			$a_i E_g + b_i$	b_i	
Evolution	Environment only	Free	Towards plasticity	Towards robustness	Drift
Regulating	All genes	All genes but sensor	TF	TF	None
Regulated by	None	All genes but Neutral	TF and Sensor	TF and Sensor	TF and Sensor

Table 3: Description of the five gene categories implemented in the gene regulatory network model.

0. If it changed the reaction norm slope or the intercept by more than 0.01, then the regulation was considered as "essential": $\mathbf{Et}_{ij} = \text{sign}(\mathbf{W}_{ij})$, and $\mathbf{Et}_{ij} = 0$ otherwise.

Motifs We investigated three loop categories: feedforward loops (corresponding to 8 unique motifs, Table 1), diamonds loops (10 unique motifs, Table 2), and feedbacks loops (of 4 different sizes). The nomenclature of the FFLs was based on [3], the genes under consideration hold the position Z (regulated gene) in the motifs. FFLs were divided into two groups, coherent and incoherent. The coherence of a motif was defined by the product of all regulations of the motif: coherent if the product was positive, incoherent otherwise (Table 1). The nomenclature of diamonds was defined by the the signs of the $X \rightarrow Yn$ regulations (positive, negative, or mixed) for the first part of the name and the sign of $Yn \rightarrow Z$ for the second part (Table 2). Feedback loops were divided into subcategories by size, from 2 to 5 genes.

Control datasets In drifting populations, populations of 200 individuals evolved for 20000 generations with a mutation rate of $\mu = 0.4$. "Shuffled" controls were obtained by randomly shuffling the values of non-null regulations, the number of regulations and loops were thus kept, but not the nature of loop motifs.

Availability and reproducibility

The simulation program was coded in C++, and is available at https://github.com/lerouzic/simevolv. The program was compiled and executed on the IFB-core cluster (https://www.france-bioinformatique.fr/en/ifb-core-cluster/). Simulation pipelines (in R and Bash) and analysis scripts (in R) are available at https://github/apetit/topoplast.

Aknowledgments

Simulations were performed on the Core Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). AP was supported by the doctorate school SDSV (ED 577, Université Paris-Saclay). The authors aknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-22-CE02-0026-01 (project EVOPLANET). The authors declare no conflict of interest.

References

[1] Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31(1):64–68.

[2] Burda Z, Krzywicki A, Martin OC, Zagorski M (2011) Motifs emerge from function in model gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* 108(42):17263–17268.

[3] Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* 100(21):11980–11985.

[4] Alon U (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8(6):450.

[5] Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences* 101(17):6355–6360.

[6] Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* 138(4):760–773.

[7] de Ronde WH, Tostevin F, ten Wolde PR (2012) Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Physical Review E* 86(2):021913.

[8] Xiong K, Lancaster AK, Siegal ML, Masel J (2019) Feed-forward regulation adaptively evolves via dynamics rather than topology when there is intrinsic noise. *Nature Communications* 10(1):2418.

[9] Hornung G, Barkai N (2008) Noise Propagation and Signaling Sensitivity in Biological Networks: A Role for Positive Feedback. *PLoS Computational Biology* 4(1):e8.

[10] Panovska-Griffiths J, Page KM, Briscoe J (2013) A gene regulatory motif that generates oscillatory or multiway switch outputs. *Journal of the Royal Society Interface* 10(79).

[11] Osella M, Bosia C, Corá D, Caselle M (2011) The Role of Incoherent MicroRNA-Mediated Feedforward Loops in Noise Buffering. *PLoS Computational Biology* 7(3):e1001101.

[12] Kadelka C, Murrugarra D, Laubenbacher R (2013) Stabilizing gene regulatory networks through feedforward loops. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23(2):1–19.

[13] Roy S, Raj M, Ghosh P, Das SK (2017) Role of motifs in topological robustness of gene regulatory networks. *IEEE International Conference on Communications* 1.

[14] Kwon YK, Cho KH (2007) Analysis of feedback loops and robustness in network evolution based on Boolean models. *BMC Bioinformatics* 8(1):430.

[15] Sen S, Kim J, Murray RM (2014) Designing robustness to temperature in a feedforward loop circuit in *53rd IEEE Conference on Decision and Control*. (IEEE), Vol. 2015-Febru, pp. 4629–4634.

[16] Weldemichael T, Asemoloye MD, Marchisio MA (2022) Feedforward Loops: Evolutionary Conserved Network Motifs Redesigned for Synthetic Biology Applications. *Applied Sciences* 12(16):8292.

[17] Keseler IM, et al. (2021) The ecocyc database in 2021. *Frontiers in Microbiology* 12.

[18] Wagner A (1994) Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proceedings of the National Academy of Sciences* 91(10):4387–4391.

[19] Wagner A (1996) Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biological cybernetics* 74(6):557–567.

[20] Mason O, Verwoerd M (2007) Graph theory and networks in biology. *IET systems biology* 1(2):89–119.

[21] Pavlopoulos GA, et al. (2011) Using graph theory to analyze biological networks. *BioData mining* 4:1–27.

[22] Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. *Current opinion in structural biology* 14(3):283–291.

[23] Espinosa-Soto C (2018) On the role of sparseness in the evolution of modularity in gene regulatory networks. *PLoS computational biology* 14(5):e1006172.

[24] Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. *Nature Reviews Genetics* 8(12):921–931.

[25] Conant GC, Wagner A (2003) Convergent evolution of gene circuits. *Nature genetics* 34(3):264–266.

[26] Lynch M (2007) The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics* 8(10):803.

[27] Tsuda ME, Kawata M (2010) Evolution of gene regulatory networks by fluctuating selection and intrinsic constraints. *PLoS computational biology* 6(8):e1000873.

[28] Petit AJR, Guez J, Le Rouzic A (2023) Correlated stabilizing selection shapes the topology of gene regulatory networks. *Genetics* 224(2). iyad065.

[29] Fierst JL (2011) A history of phenotypic plasticity accelerates adaptation to a new environment. *Journal of evolutionary biology* 24(9):1992–2001.

[30] Smits WK, Kuipers OP, Veening JW (2006) Phenotypic variation in bacteria: the role of feedback regulation. *Nature Reviews Microbiology* 4(4):259–271.

[31] Karp PD (2001) Pathway databases: a case study in computational symbolic theories. *Science* 293(5537):2040–2044.

[32] Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *Journal of molecular biology* 323(5):785–793.

[33] Promislow D (2005) A regulatory network analysis of phenotypic plasticity in yeast. *The American Naturalist* 165(5):515–523.

[34] Mochizuki A (2008) Structure of regulatory networks and diversity of gene expression patterns. *Journal of Theoretical Biology* 250(2):307–321.

[35] Mangan S, Itzkovitz S, Zaslaver A, Alon U (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *Journal of Molecular Biology* 356(5):1073–1081.

[36] Cooper MB, Loose M, Brookfield JF (2008) Evolutionary modelling of feed forward loops in gene regulatory networks. *BioSystems* 91(1):231–244.

[37] Ghosh B, Karmakar R, Bose I (2005) Noise characteristics of feed forward loops. *Physical Biology* 2(1):36–45.

[38] Draghi J, Whitlock M (2015) Robustness to noise in gene expression evolves despite epistatic constraints in a model of gene networks. *Evolution* 69(9):2345–2358.

[39] Schmutzer M, Wagner A (2020) Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. *PLoS computational biology* 16(10):e1007727.

[40] Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM (2020) Gene regulatory network inference resources: A practical overview. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1863(6):194430.

[41] Escorcia-Rodríguez JM, et al. (2023) Improving gene regulatory network inference and assessment: The importance of using network structure. *Frontiers in Genetics* 14(February):1–14.

[42] Lipshtat A, Purushothaman SP, Iyengar R, Ma'ayan A (2008) Functions of bifans in context of multiple regulatory motifs in signaling networks. *Biophysical Journal* 94(7):2566–2579.

[43] Jiang Y, Hao N (2021) Memorizing environmental signals through feedback and feedforward loops. *Current Opinion in Cell Biology* 69:96–102.

[44] Knabe JF, Nehaniv CL, Schilstra MJ (2008) Do motifs reflect evolved function?—No convergent evolution of genetic regulatory network subgraph topologies. *Biosystems* 94(1-2):68–74.

[45] Freyre-González JA, et al. (2022) System Principles Governing the Organization, Architecture, Dynamics, and Evolution of Gene Regulatory Networks. *Frontiers in Bioengineering and Biotechnology* 10.

[46] Ingram PJ, Stumpf MP, Stark J (2006) Network motifs: structure does not determine function. *BMC Genomics* 7(1):108.

[47] Kim S, Wysocka J (2023) Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular Cell* 83(3):373–392.

[48] Connally NJ, et al. (2022) The missing link between genetic association and regulatory function. *eLife* 11:1–35.

[49] Tucker DL, Tucker N, Conway T (2002) Gene expression profiling of the pH response in *Escherichia coli*. *Journal of Bacteriology* 184(23):6551–6558.

[50] Maurer LM, Yohannes E, Bondurant SS, Radmacher M, Slonczewski JL (2005) pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *Journal of Bacteriology* 187(1):304–319.

[51] Feugeas JP, et al. (2016) Links between Transcription, Environmental Adaptation and Gene Variability in *Escherichia coli*: Correlations between Gene Expression and Gene Variability Reflect Growth Efficiencies. *Molecular Biology and Evolution* 33(10):2515–2529.

[52] Ng TW, et al. (2018) Differential gene expression in *Escherichia coli* during aerosolization from liquid suspension. *Applied Microbiology and Biotechnology* 102(14):6257–6267.

[53] Bergholz TM, Vanaja SK, Whittam TS (2009) Gene expression induced in *Escherichia coli* O157:H7 upon exposure to model apple juice. *Applied and Environmental Microbiology* 75(11):3542–3553.

[54] Caglar MU, et al. (2017) The *E. coli* molecular phenotype under different growth conditions. *Scientific Reports* 7(October 2016):1–15.

[55] White-Ziegler CA, Malhowski AJ, Young S (2007) Human body temperature (37°C) increases the expression of iron, carbohydrate, and amino acid utilization genes in *Escherichia coli* K-12. *Journal of Bacteriology* 189(15):5429–5440.

[56] Kim S, et al. (2020) Heat-responsive and time-resolved transcriptome and metabolome analyses of *Escherichia coli* uncover thermo-tolerant mechanisms. *Scientific Reports* 10(1):1–13.

[57] Bhatia RP, Kirit HA, Predeus AV, Bollback JP (2022) Transcriptomic profiling of *Escherichia coli* K-12 in response to a compendium of stressors. *Scientific Reports* 12(1):1–14.

[58] Wang S, et al. (2009) Transcriptomic response of *Escherichia coli* O157:H7 to oxidative stress. *Applied and Environmental Microbiology* 75(19):6110–6123.

[59] Durfee T, Hansen AM, Zhi H, Blattner FR, Ding JJ (2008) Transcription profiling of the stringent response in *Escherichia coli*. *Journal of Bacteriology* 190(3):1084–1096.

[60] Bergman A, Siegal M (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424:549–552.

[61] Leclerc R (2008) Survival of the sparsest: robust gene networks are parsimonious. *Molecular Systems Biology* 4:213.

[62] Rünneburger E, Le Rouzic A (2016) Why and how genetic canalization evolves in gene regulatory networks. *BMC evolutionary biology* 16(1):239.

[63] Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33(1Part2):402–416.

[64] Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. *Proceedings of the National Academy of Sciences* 99(16):10528–10532.

[65] Draghi JA, Whitlock MC (2012) Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. *Evolution: International Journal of Organic Evolution* 66(9):2891–2902.

[66] Vadée-Le-Brun Y, Rouzaud-Cornabas J, Beslon G (2015) Epigenetic inheritance speeds up evolution of artificial organisms in *Proceedings of the European Conference on Artificial Life 13*. (MIT Press), pp. 439–446.

[67] Burban E, Tenaillon MI, Le Rouzic A (2022) Gene network simulations provide testable predictions for the molecular domestication syndrome. *Genetics* 220(2):iyab214.