Correlated stabilizing selection shapes the topology of gene regulatory networks

Apolline J. R. Petit[1,*], Jeremy Guez[2,3], and Arnaud Le Rouzic[1]

[1] *Université Paris-Saclay, CNRS, IRD, UMR EGCE, 91190, Gif-sur-Yvette, France*
[2] *UMR 7206 Eco-Anthropologie, CNRS, MNHN, Université Paris Cité, 75116 Paris, France*
[3] *Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France*
[*] *Corresponding author: Université Paris-Saclay, CNRS, IRD, UMR EGCE, 91190, Gif-sur-Yvette, France. Email: apolline.petit@universite-paris-saclay.fr*

Abstract

The evolution of gene expression is constrained by the topology of gene regulatory networks, as co-expressed genes are likely to be affected together by mutations. Conversely, co-expression can also be an advantage when genes are under joint selection. Here, we assessed theoretically whether correlated selection (selection for a combination of traits) was able to affect the pattern of correlated gene expressions and the underlying gene regulatory networks. We ran individual-based simulations, applying a stabilizing correlated fitness function to two genetic architectures: a quantitative genetics (multilinear) model featuring epistasis and pleiotropy, and a gene regulatory model, mimicking the mechanisms of gene expression regulation. Simulations showed that correlated mutational effects evolved in both genetic architecture as a response to correlated selection, but the response in gene networks was less precise due to the mechanistic constraints of gene regulation. The intensity of gene co-expression was mostly explained by the regulatory distance between genes (largest correlations being associated to genes directly interacting with each other), and the sign of co-expression was associated with the nature of the regulatory interaction (transcription activation or inhibition). These results concur to the idea that gene network topologies could partly reflects past correlated selection patterns on gene expression.

Key words Evolvability ; Mutation Covariance Matrix ; Evolution of Pleiotropy ; Evolutionary Systems Biology ; Quantitative Genetics.

Introduction

The development and physiology of living organisms are controlled by large and complex Gene Regulatory Networks (GRNs). The central role of GRNs is documented in all kinds of organisms, e.g. for the control of cell physiology in yeasts (Guelzim *et al.*, 2002), heart development in humans (Olson, 2006), skeleton development in sea urchins (Shashikant *et al.*, 2018), or flower development in angiosperms (Espinosa-Soto *et al.*, 2004). The organization of these networks have long been of central interest for systems biologists, and it is now widely acknowledged that GRNs tend to follow general structural rules: for instance, they tend to be sparse, modular (Wagner *et al.*, 2007; Espinosa-Soto, 2018), and scale-free (the number of connections per node follows a power law) (Babu *et al.*, 2004; Ouma *et al.*, 2018).

The reasons why real-life GRNs are organized in such a way are not completely clear (Espinosa-Soto, 2018; Taylor *et al.*, 2022). Because the expression of genes is expected to affect phenotypic traits, and thus to condition the fitness of organisms, the gene expression pattern is believed to respond to selection, at least for a subset of genes. For instance, specific sets of genes have been shown to change consistently in the wild (Philippe *et al.*, 2007; Verta and Jones, 2019; Huang *et al.*, 2021) or during experimental evolution (Philippe *et al.*, 2007; Ghalambor *et al.*, 2015; Jallet *et al.*, 2020). In contrast, the structure of the network itself is less directly submitted to natural selection. As multiple gene network topologies are capable of producing the same gene expression patterns, at least in theory

(Wagner and Wright, 2007), the main evolutionary mode of network structure may be related to non-adaptive processes, such as systems drift (Lynch, 2007), or mutation bias (Van Noort *et al.*, 2004). Nonetheless, it has been empirically observed that the gene network structure may be deeply rewired during species domestication (Swanson-Wagner *et al.*, 2012; Bellucci *et al.*, 2014), but whether these examples match theoretical predictions remains to be determined (Burban *et al.*, 2022). Similarly, whether or not gene network properties are influenced by indirect selection towards (or against) long-term evolvability remains theoretically unclear (Wagner, 2008; Mayer and Hansen, 2017). Overall, there is little theoretical predictions about how selection may affect the network topology, and about the possible role of adaptation in shaping the structure of GRNs.

May the evolution of GRN topology be predicted from quantitative genetics theory? After all, gene expressions can be assimilated to quantitative traits, and the complex result of regulatory interactions can be described as epistasis (non-additive gene-gene interactions) and pleiotropy (each gene affects several traits) (Fagny and Austerlitz, 2021). Statistical tools from quantitative genetics are routinely used to detect expression Quantitative Trait Loci (Gilad *et al.*, 2008; Michaelson *et al.*, 2009), and evolutionary quantitative genetics provide a wide corpus of evolutionary models (e.g. Walsh and Lynch, 2018) which could be relevant to understand and predict changes in gene expression. Particularly relevant to the evolution of genetic architectures are the models focusing on the evolution of pleiotropy and modularity of quantitative characters (Sgrò and Hoffmann, 2004; Pavličev and Cheverud, 2015). In particular, when the genetic architecture is epistatic, pleiotropy could theoretically evolve in response to stabilizing selection when the effect of traits on fitness are not independent (Jones *et al.*, 2014). This correlated selection, which corresponds to the selection of trait combinations (illustrated in Suppl. Fig. 1A), has been documented for various combinations of phenotypic characters (Sinervo and Svensson, 2002), but its consequences on the structure of genetic architectures is not well understood (Svensson *et al.*, 2021). The evolutionary mechanism involved in the evolution of pleiotropy relies on the fact that the genetic load of new mutations is minimized when the mutational structure matches the direction of the fitness function (Suppl. Fig. 1B). The results by Jones *et al.* (2014), translated to the gene expression level, thus formalises the hypothesis that correlated selection could favor gene network topologies promoting the co-expression of co-selected genes, even when the fitness function is constant. Yet, this important result from evolutionary quantitative genetics may not be straightforward to translate towards systems biology. For instance, the genetic architecture in Jones *et al.* (2014) was based on a bivariate multilinear model (Hansen and Wagner, 2001), featuring unconstrained and isotropic pleiotropic epistasis (i.e. any gene have the potential to modify the pleiotropy of any other gene). In contrast, the large epistatic and pleiotropic effects in GRNs are largely constrained and biased by gene expression regulation mechanisms, and by the topology of gene networks.

Here, we intend to understand the propensity of correlated stabilizing selection to shape the structure of GRNs by monitoring the evolution of mutational co-expressions in individual-based population simulations. We will use the theoretical framework proposed by Wagner (1994, 1996) to implement a simple gene regulatory network model as a genotype-phenotype map. This framework is well-suited for being coupled with individual-based simulations, as the genotype (the set of regulatory interactions between genes) and the phenotype (gene expressions) are explicit and clearly separated. We will address the evolution of gene co-expression at two levels: (i) at the gene expression level, can gene networks evolve to optimize mutational co-expression in regard to correlated selection ? (ii) at the network level, what is the effect of correlated selection on network structure and topology?

Material and Methods

Simulation model

Reproduction Simulations follow a traditional Wright-Fisher model. Populations consist in N sexually-reproducing hermaphrodite individuals. The genotype is encoded as n diploid, freely recombining loci, the (multivariate) phenotype being computed from the genotype according to one of the two genotype-phenotype models described below. Generations are non-overlapping; for each offspring, two parents are picked with a probability proportional to their fitness, and each give a haploid gamete (n genes). Populations evolve during 10,000 generations and are submitted to genetic drift, selection, and mutations.

Selection Individual fitness is determined by a multivariate stabilizing bell-shaped fitness function:

$$w(\mathbf{Z}_i) = \exp(-\frac{1}{2}(\mathbf{Z}_i - \mathbf{\Theta})^T \mathbf{S}^{-1}(\mathbf{Z}_i - \mathbf{\Theta})), \tag{1}$$

where $\mathbf{Z}_i$ is the vector of phenotypes for individual i, $\mathbf{\Theta}$ is the optimal phenotype (maximum fitness, $w(\mathbf{\Theta}) = 1$), and $\mathbf{S}$ is the covariance matrix of the fitness function. For simplicity, the number of phenotypic traits on which selection is applied is reduced to two traits whatever the genotype-phenotype model. As a consequence, the fitness function can be specified by five parameters: two localization parameters for the phenotypic optima in the 2-vector $\mathbf{\Theta}$, and three (co)variance parameters for the 2×2 matrix $\mathbf{S}$ (the strength of selection on traits 1 and 2, and the correlated selection).

Mutations Mutations affect the genotype during gamete formation, before the computation of the phenotype of individuals. They occur with a rate μ per gamete, affect a random locus, and a random element of this locus when the genotype has more than one dimension.

Mutational effects are cumulative, the new allelic value is drawn in a Gaussian distribution centered on the former value as $x' \sim \mathcal{N}(x, \sigma_m)$ where σ_m determines the mutation size.

Genotype-Phenotype maps

Multilinear model The multilinear model was developed by Hansen and Wagner (2001). Although provided as a multivariate model in its original description, it has been extensively used in its univariate form in the quantitative genetics literature (Hermisson *et al.*, 2003; Carter *et al.*, 2005; Jones *et al.*, 2007; Le Rouzic *et al.*, 2013), but more rarely in its multivariate implementation (Jones *et al.*, 2014).

The multilinear model is built as an extension of the traditional additive model, by adding epistatic terms. Restricting the model to second-order epistasis, the phenotypic value Z_m of a trait m (among K traits) is computed as:

$$Z_m = Z_{0m} + \sum_{i=1}^{n} y_m^i + \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i=1}^{n}\sum_{j>i}^{n} \varepsilon_{mkl}^{ij}\, y_k^i\, y_l^j, \tag{2}$$

in which y_m^i is the additive effect of the genotype at locus i on the phenotypic trait m. Z_{0m} is the phenotypic reference, i.e., the phenotypic value corresponding to an arbitrary reference genotype for which all $y_m^i = 0$. For every combination of traits, the epistatic coefficient ε^{ij} quantifies the directional epistasis between loci i and j. The coefficients ε_{mmm} describe "classical" epistasis, i.e., the interaction of allelic effects y_m^i and y_m^j on trait m. In contrast, ε_{mkl} with k and/or l different from m, correspond to interactions involving pleiotropy, i.e., how trait m is influenced by the interaction between the effects of alleles on traits k and l. When all $\varepsilon^{ij} = 0$, this model collapses towards an additive model. When $\varepsilon_{mkl} \neq 0$, pleiotropy can evolve (traits can become more or less dependent). In total, there are K^3 combinations of K traits, and for n loci, $n(n-1)/2$ independent epistatic coefficients (because $j > i$) for each combination of traits.

The multilinear model analysis is limited to $K = 2$ traits. In the simulation software, haploid genotypes are stored as two values (y_1^i, y_2^i) at each locus i, and summed up to obtain diploid genotypes. These individual effects y_m^i are evolvable, while epistatic coefficients are drawn in a Gaussian distribution $\varepsilon \sim \mathcal{N}(0, 1)$ at the beginning of each simulation run, as in (Jones *et al.*, 2014).

Gene network model We used a regulatory gene network model inspired from Wagner (Wagner 1994, 1996), which is a common abstraction of the transcription regulation process as it is a dynamic model with discrete time steps (Bergman and Siegal 2003; Azevedo *et al.* 2006; Leclerc 2008; Rhoné and Austerlitz 2011; Rünneburger and Le Rouzic 2016; Espinosa-Soto 2016). The structure of the regulation network among n genes is stored in a $n \times n$ matrix $\mathbf{W}$, corresponding to cis-regulatory interactions among transcription factors. Element W_{ij} corresponds to the effect of the product of gene j on the expression of gene i. Inhibiting interactions are negative values and activating interactions are positive values. Zero indicates the absence of direct regulation. Each gene of the network is susceptible to act as a transcription factor and to modify the expression of other genes; there is no self-regulation ($W_{ii} = 0$).

Gene expression is computed dynamically for 24 time steps. Initial gene expressions are set to their constitutive level 0.5 (expression in absence of regulation), and updated as a function of the concentration of the other genes of the network:

$$\mathbf{P}_{t+1} = F(\mathbf{W}\mathbf{P}_t), \tag{3}$$

where $\mathbf{P}$ is a vector of n quantitative gene expressions, scaled between 0 (no expression) to 1 (maximal expression), by a sigmoid function $F(x_1, \ldots, x_n) = (f(x_1), \ldots, f(x_n))$, where $f(x) = 1/(1+e^{-4x})$ (the same scaling function as in Rünneburger and Le Rouzic 2016 with a constitutive expression set to 1/2).

The phenotype $\mathbf{Z}$ corresponding to a genotype $\mathbf{W}$ is computed from the average expression of the two first genes of the network (hereafter called "a" and "b") for the 4 last time steps $\overline{\mathrm{P}} = (1/4)\sum_{t=21}^{24} \mathrm{P}_t$, as $\mathrm{Z}_i = \log[\overline{\mathrm{P}_i}/(1-\overline{\mathrm{P}_i})]$, rescaled over $(-\infty, +\infty)$ to be directly comparable with the multilinear model. The phenotypic value $Z_i = 0$ corresponds to the gene constitutive expression $P_i = 0.5$.

In addition to the selection on the phenotype, unstable networks are penalized (as in e.g. Siegal and Bergman 2002), by multiplying the individual fitness w by a factor $w_j = \exp(-s' \sum_i^n V_i)$, where s' quantify the selection against unstable network, and $V_i = (1/4)\sum_{t=21}^{24}(P_{it} - \overline{P}_i)^2$ is the variance in the expression of gene i during the last 4 steps of the network dynamics. We set $s' = 46,000$, as in Rünneburger and Le Rouzic 2016, which in practice means that unstable networks are virtually unviable.

This gene network model is coupled with the population genetic model described above. Individual genotypes correspond to the matrix $\mathbf{W}$, each line of the matrix (a vector of n cis-regulatory sites) stands for a locus. Gene networks are composed of 6 genes. Two genes are under correlated selection (with an optimal expression level set at their constitutive expression), two other genes are under stabilizing, non-correlated selection for slight over-expression, and the two remaining genes are not directly selected (they could participate to the network as transcription factors). The analyses are focused on the evolution of the two first genes.

Model output

The simulation software reports the means, variances and co-variances of the population phenotypes at regular time points. In addition, the population average mutation co-variance matrix $\mathbf{M}$ is estimated in the following way: 6 mutations are simulated for each of the $N = 5,000$ individuals i, leading to $5,000$ covariance matrices $\mathbf{C}_i$. These covariance matrices are then averaged out and multiplied by the mutation rate μ to estimate the mutational covariance matrix $\mathbf{M}$ in the population. The rest of the analysis are focused on 2×2 $\mathbf{M}$ matrices, corresponding to both selected genes.

We tracked the evolution of the direction $\alpha(\mathbf{M})$ of mutational matrices, expressed in the interval $(-\pi/2, \pi/2)$, as:

$$\begin{aligned} \alpha'(\mathbf{M}) &= \arccos(m_{1,1}) \mod \pi, \\ \alpha(\mathbf{M}) &= \begin{cases} \alpha'(\mathbf{M}) & \text{if } \alpha'(\mathbf{M}) < \pi/2 \\ \alpha'(\mathbf{M}) - \pi & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

where $m_{1,1}$ is the first element of the first eigenvector of the 2×2 matrix $\mathbf{M}$. The mean α over R replicates is obtained as the circular mean restricted to the $(0, \pi)$ interval:

$$\bar{\alpha} = \frac{1}{2}\left[\mathrm{atan2}\left(\frac{1}{R}\sum_{i=1}^{R}\sin 2\alpha_i, \frac{1}{R}\sum_{i=1}^{R}\cos 2\alpha_i\right)\right] \mod \pi, \tag{5}$$

where $\mathrm{atan2}(x, y) = 2\ \mathrm{atan}\ [y/(x + \sqrt{x^2 + y^2})]$ is the angle between the X-axis and the vector (x, y).

The eccentricity of the $\mathbf{M}$ matrix is computed as $e(\mathbf{M}) = \sqrt{1 - \lambda_2/\lambda_1}$, where λ_i stands for the ith eigenvalue of the matrix $\mathbf{M}$. Mutational correlation $r(\mathbf{M})$ between genes a and b were calculated from $\mathbf{M}$ matrices as $r(\mathbf{M}) = \mathrm{M}_{a,b}/(\sqrt{\mathrm{M}_{a,a}}\sqrt{\mathrm{M}_{b,b}})$.

Simulation parameters

Default simulation parameters were set as displayed in Table 1. For the multilinear model, the number of loci, number of selected traits, and epistasis parameters have been taken from Jones et al., 2014

Parameter	Multilinear	Gene network
Generations	10000	10000
Population size N	5000	5000
Loci n	20	6
Correlationally selected traits	2	2
Haplotype mutation rate μ	0.1	0.1
Mutation effect σ_m	0.1	0.1
Optimum phenotype	0	0
S matrix size	10	10
S matrix eccentricity	0.94	0.94

Table 1: Default parameters for both models.

(Jones *et al.*, 2014). Parameters for the gene network models were adapted to match those of the multilinear model, except for the number of loci ($n = 6$) due to computational constraints.

All simulations starts with genotypic values (y_i in the multilinear model, W_{ij} in the gene network model) set at 0, unless specified otherwise. In some simulations, the initial gene network topology is manipulated (positive or negative initial co-expression) by setting some initial regulatory interactions (W_{ab} and W_{ba}) with positive ($+0.5$) or negative (-0.5) values.

The bivariate stabilizing selection (**S** variance matrix) is parameterized in each simulation run by setting the angle of the major axis (between $-\pi/2$ and $\pi/2$); the selection matrix size (trace) and eccentricity remain constant (see Table 1 and orange ellipses in Figure 1).

Network optimization

The properties of the gene network genotype-phenotype map (and, in particular, its capacity to match the target fitness function) were further explored by optimizing networks with a quasi-Newton method (L-BFGHS-B from the optimx package in R, Nash and Varadhan, 2011). The algorithm minimizes the function:

$$o(\mathbf{W}) = \sum_i s_i(P_i - \Theta_i)^2 + s' \sum_i V_i + s_a(\alpha(\mathbf{M}) - \theta_\alpha)^2 + s_e(1 - e(\mathbf{M}))^2 + s_s(\mathrm{tr}(\mathbf{M}) - \theta_s)^2, \qquad (6)$$

where P_i, V_i, s_i and s' are the same as in the fitness function, index i representing each gene of the network. θ_α is the optimal direction of $\mathbf{M}$ (equivalent to $\alpha(\mathbf{S})$ in the simulations), $s_a = 1$, $s_e = 1$, and $s_s = 100$ are arbitrary coefficients weighting the strength of selection on the direction, the eccentricity, and the size of the $\mathbf{M}$ matrix, respectively. The target size (trace of the mutation covariance matrix) is fixed to $\theta_s = 0.026$, the average value from the control simulations. Optimization is performed on all mutable elements of the network $\mathbf{W}$ simultaneously. $\mathbf{M}$ matrices are computed deterministically from the (co)variances of gene expressions among all networks $\mathbf{W}'$ differing from $\mathbf{W}$ by a single $\pm\sigma_m = 0.1$ mutation.

Availability and model implementation

All data presented in this article have been generated by computer simulation of evolving populations with the C++ program Simevolv (Rünneburger and Le Rouzic 2016: `https://github.com/lerouzic/simevolv.git`), in wich both multilinear and GRN models are implemented. The scripts for these operations have been written in R and are available at `https://github.com/apetit8/Mmatrix_paper.git`.

Results

Mutational co-expressions can evolve in gene networks

We compared the evolution of simulated populations based on two genetic architectures: quantitative traits underlain by a multilinear genotype-phenotype map (from which the evolution of mutational covariance in response to correlated selection has already been assessed by Jones *et al.*, 2014), and a gene

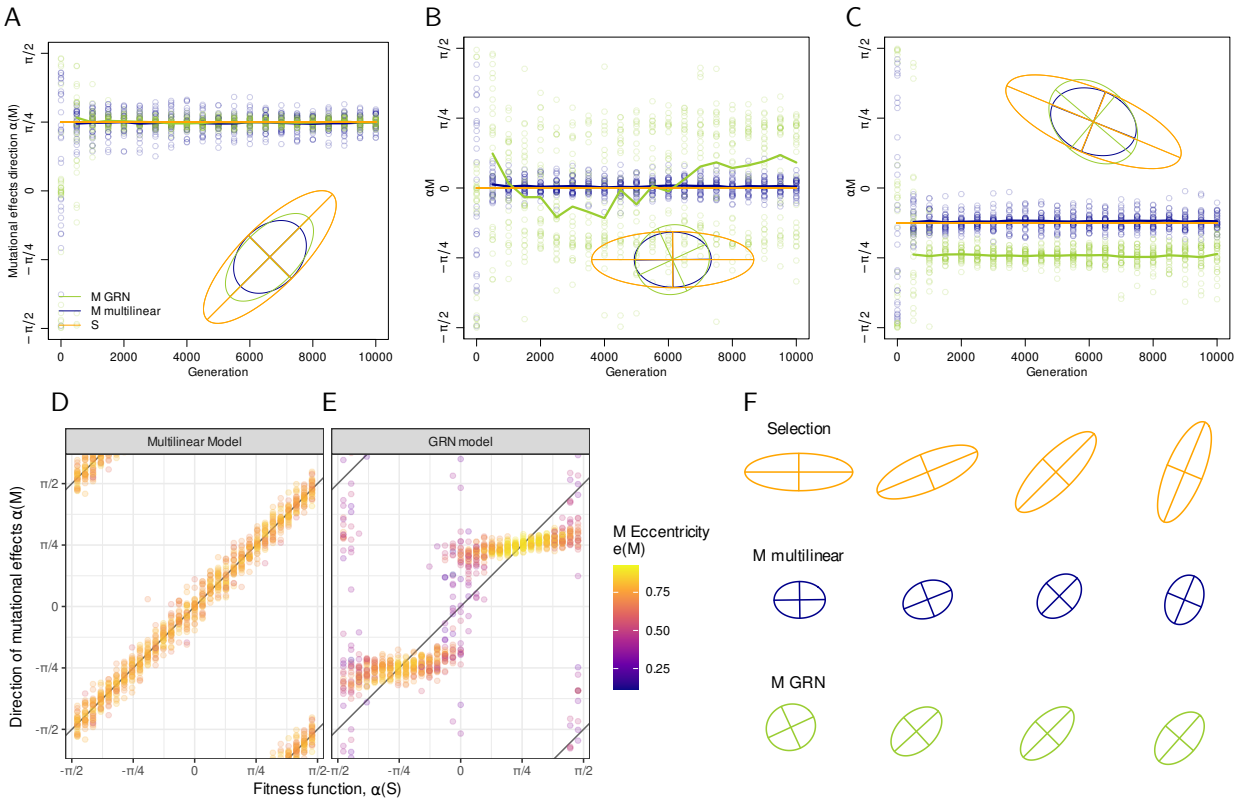


Figure 1: A, B, C: Evolution of the angle of the main axis of the mutation matrix ($\alpha(\mathbf{M})$) along generations. Blue: multilinear model, Green: GRN model. Orange ellipses represent the fitness function (scaled $\times 0.025$), which direction was $\alpha(\mathbf{S}) = +\pi/4$ (panel A), $\alpha(\mathbf{S}) = 0$ (panel B), and $\alpha(\mathbf{S}) = -\pi/8$ (panel C). Dots illustrate 30 simulation replicates, plain lines stand for circular means $\bar{\alpha}(\mathbf{M})$. Ellipses are the geometric representation of $\mathbf{M}$ matrices at the last generation (10,000); averaged over the 30 replicates. D and E: $\alpha(\mathbf{M})$ as a function of $\alpha(\mathbf{S})$ for both models. Dots represents the direction of the main axis of the $\mathbf{M}$ matrix at generation 10,000 in 30 simulation replicates for 31 values of $\alpha(\mathbf{S})$ regularly spaced between $-\pi/2$ and $\pi/2$; the color scale encodes the eccentricity of $\mathbf{M}$ (yellow for more pleiotropy). F: Average ellipses obtained at the last generations, showing the different nature of the response of $\mathbf{M}$ to correlated stabilizing selection with both genetic architecture models. Columns stand for $\alpha(\mathbf{S}) = 0, \pi/8, \pi/4$, and $3\pi/8$, respectively.

regulatory network architecture, considering gene expression levels as phenotypic traits. This approach made it possible to compare both models based on quantitative genetics concepts. Two traits (the expression of two genes in the GRN model) were submitted to correlated stabilizing selection, the fitness function being defined by the direction $\alpha(\mathbf{S})$ of the optimal trait combination. Our expectation was that pleiotropy (measured as the shape and direction of the mutational covariance matrix $\mathbf{M}$) should evolve in order to match the direction of the fitness function. In the multilinear model simulations, the alignment between the main axis of the mutational matrix $\mathbf{M}$ and the direction of the correlated fitness function was convincing after less than 500 generations (Figure 1A – C). This result reproduces the conclusions from Jones *et al.*, 2014, based on the same model. In contrast, with the gene network model, the pleiotropy (genetic correlation) was often stronger, but the alignment was less precise, even after 10,000 generations: the sign of the mutational correlation matched the sign of the fitness correlation, but they were substantial discrepancies when considering the direction of the main mutational axis.

The different nature of the response to correlated selection in both models is illustrated in Figure 1D and E. The accuracy of the response to the tilted fitness function was measured as the angular difference $\alpha(\mathbf{M})$ vs. $\alpha(\mathbf{S})$, while the intensity of pleiotropy (co-expression) was quantified by the eccentricity of the $\mathbf{M}$ matrix. The response of the multilinear model was isotropic, and the shape of the $\alpha(\mathbf{M})$ matrix did not depend on the angle of the correlated selection (Figure 1D). In contrast, when traits are gene expressions from a regulatory network, although both the direction and eccentricity of $\mathbf{M}$ evolve

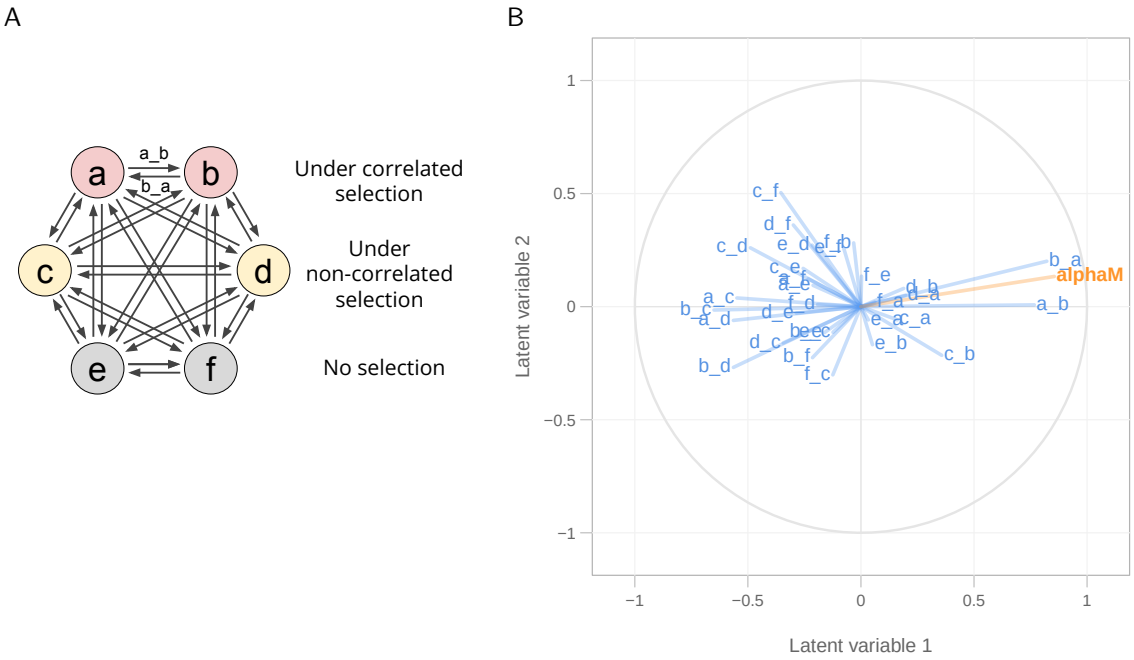


Figure 2: A: Diagram of our gene regulatory network design. a and b are the focal genes (the genes which expression is under correlated selection). c and d are genes selected to be slightly activated, independently from each other. e and f are free to evolve without affecting the fitness directly, and can thus act as transcription factors. B: Partial Least Square correlation circle. The networks included in the analysis (latent variables) were the final genotypes (generation 10,000) of the simulations displayed in Figure 1D (930 networks in total, corresponding to 31 regularly distributed angles for the correlated fitness function, with 30 simulation replicates for each angle). The only response variable was the direction $\alpha(\mathbf{M})$ of the mutation matrix between genes a and b.

(Figure 1E), there were preferential directions for the evolution of co-expression: they did match the sign, but not the precise direction of the fitness function. Intermediate angles ($\pi/4$: both gene expressions affected equally by mutations, and $-\pi/4$: opposite effects on both genes) were frequently observed, but mutational independence ($\pi/2$, $-\pi/2$, or 0 direction and round ellipses) was difficult to achieve.

Mutational co-expression is determined by local regulatory motifs

In the previous section, we used quantitative genetics tools to describe the structure of mutational correlations among traits and its evolution. Here we aim at deciphering the changes in the regulatory motifs that underlie the evolution of gene co-expression. We performed a Partial Least Square (PLS) regression to measure the influence of each of the 30 regulatory weights of the 6-gene network (the elements of the $\boldsymbol{W}$ matrix, Figure 2A) on the direction of the mutational covariances $\alpha(\mathbf{M})$. The PLS regression is dominated by the direct regulations between target genes (a and b) (Figure 2B): when a up-regulates b (and, respectively, when b up-regulates a), the co-expression between a and b increases. In contrast, down-regulation ($W_{ab} < 0$ or $W_{ba} < 0$) leads to negative $\alpha(\mathbf{M})$. More modest contributors to the direction of the $\mathbf{M}$ matrix are the feedback contributions of genes a and b to the regulation of genes c and d, as W_{ac}, W_{ad}, W_{bc}, and W_{bd} are all negatively associated with $\alpha(\mathbf{M})$. Opposite interactions (genes c and d regulating a and b) are positively (but slightly) associated with the direction of co-expression. There was little association between the other regulatory weights and the direction of the mutational matrix, suggesting that the co-expression between two genes is determined by the local regulatory structure. In particular, direct regulatory interactions appear as a major contributor to the evolution of mutational co-expression in response to correlated selection.

The influence of direct regulations was further assessed by fixing their value to positive (activation), negative (repression), or no regulation before running the simulations. The corresponding slots of the

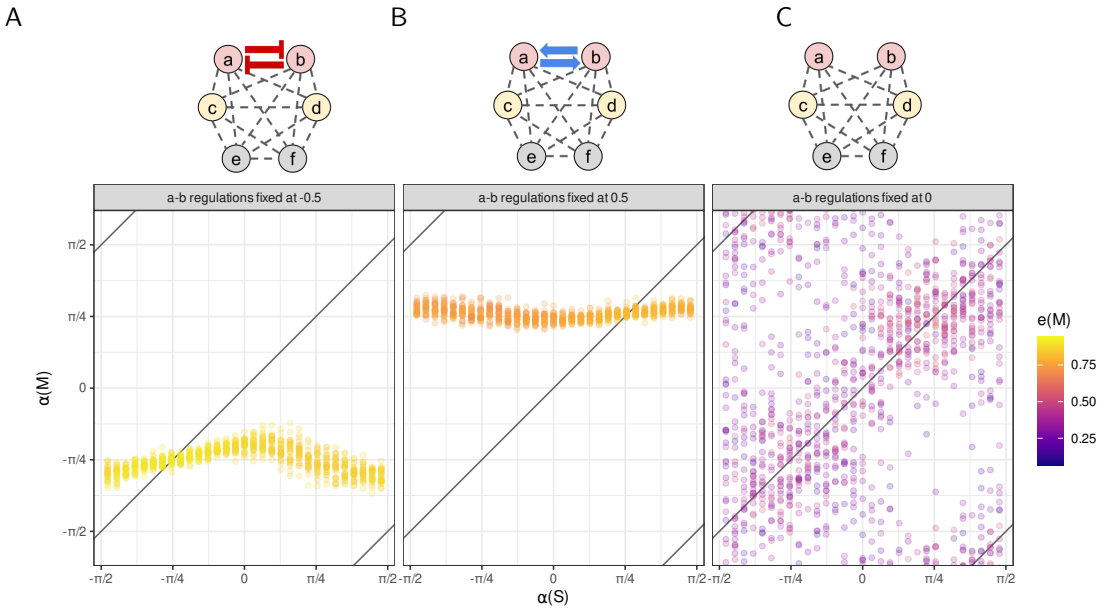


Figure 3: Direction of mutational matrices ($\alpha(\mathbf{M})$) after 10,000 generations of evolution under different correlated fitness functions (direction $\alpha(\mathbf{S})$). The representation is the same as in Figure 1D and E. A: fixed negative (inhibition) regulation between genes a and b, B: fixed positive (activation) regulation, C: no regulation. The corresponding networks are represented (red arrows: constant inhibition, blue arrows: constant activation, grey hyphenated connections: evolvable interactions, whether or not such regulatory interactions evolved in the simulations differed among replicates).

$\mathbf{W}$ matrix (W_{ab} and W_{ba}) were not evolvable and remained to their initial values, while the rest of the network was free to evolve. Other conditions were the same as in Figure 1E. Fixing regulations between the focal genes prevented the evolution of the mutational covariance structure (Figure 3A and B), and the mutational co-expression was constrained by the sign of the direct regulation interactions (mutually activating genes are always positively co-expressed, mutually inhibiting genes are always negatively co-expressed). Algorithmic optimization of the networks showed that these constraints were not absolute (Supp Fig 3), i.e., there exists theoretical networks that could overcome the sign of direct regulations, but those networks could not be found by a mutation-selection process with a correlated fitness function. Direct regulations thus highly constrain the evolution of the mutational structure in gene networks by a mutation-selection process, even if the structure of the genetic architecture makes workarounds theoretically possible.

The consequences of the absence of direct regulation interactions between gene a and b are more subtle (Figure 3C). The mutational direction $\alpha(\mathbf{M})$ is able to follow the direction of the fitness function, although the variance among simulation replicates is substantial. The eccentricity of the $\mathbf{M}$ matrix remains close to 0, indicating modest co-expression at best.

Figure 3C suggest that genes that are not directly connected in the network can still display some mutational co-expression, although weaker than when directly connected. We manipulated the network topology to increase further the network distance between focal genes and assess the effect of network distance on the possibility to mutational co-expression. All networks have evolved with the same fitness functions as in Figure 3, i.e., 31 correlated fitness functions oriented in various directions. Figure 4 summarizes the distribution of mutational correlations for various network topologies. While direct interactions between two genes allow for the evolution of mutational correlations ranging from -0.6 to $+0.7$, correlation intensity decreases very fast with the regulatory distance, as it barely spans ± 0.2 with one intermediate gene, and ± 0.05 with two intermediate genes. Correlations are slightly biased towards positive values, which is consistent with a trend to spontaneously evolve positive coexpressions in absence of correlated selection (i.e., Figure 1F). More than two intermediates discard any possibility of gene co-expression. Similar simulations with larger networks displayed the same trend (Suppl. Fig. 3).

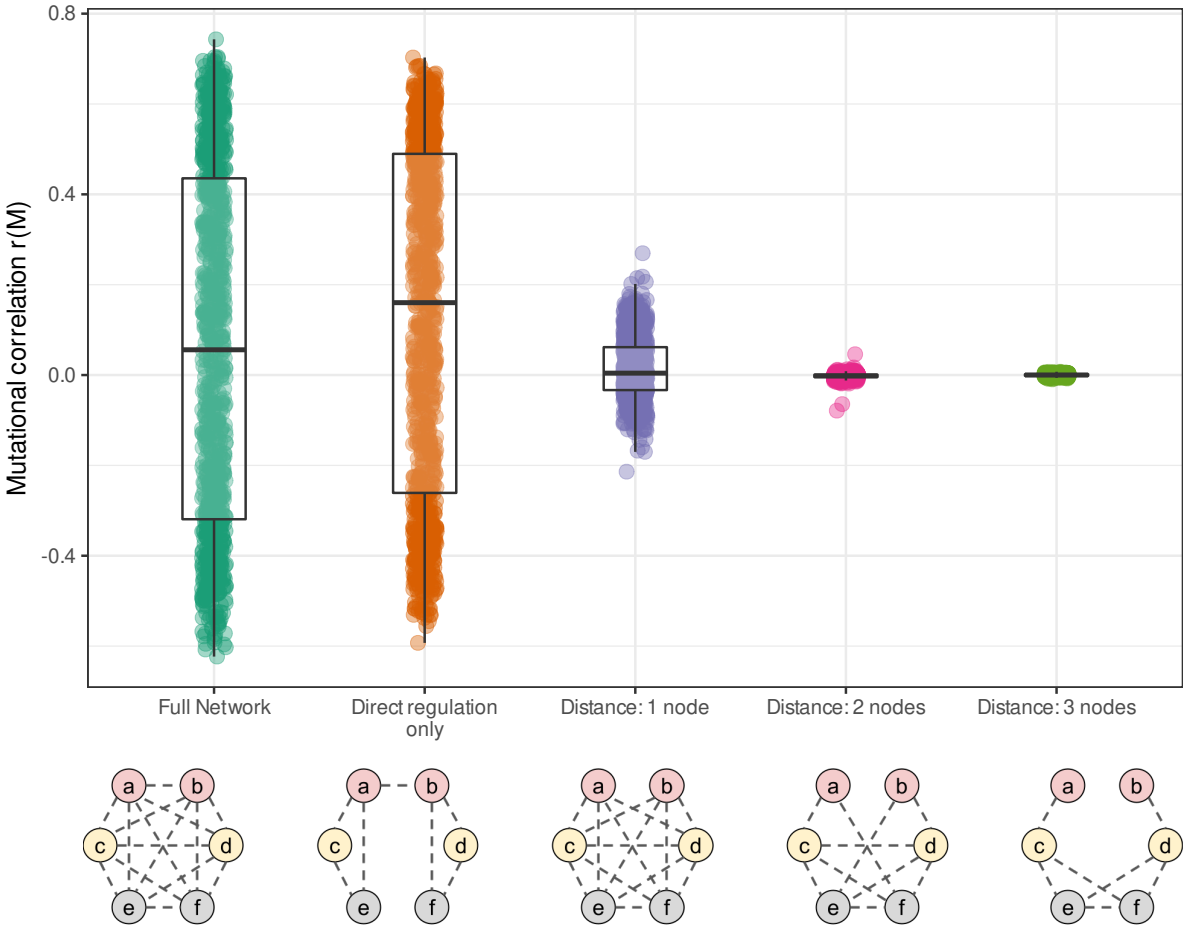


Figure 4: Distribution of mutational correlations $r(\mathbf{M})$ in networks of different topologies. Network topologies are illustrated under the x-axis. Genes colored in pink (a and b) are the genes under correlated selection, yellow genes (c and d) are under non-correlated selection, and grey genes are not selected (sans color code as in Figures 2 and 3). Regulatory connections indicate which regulations were possible; whether or not such regulatory interactions evolved in the simulations differed among replicates.

Discussion

Quantitative genetics theory predicts that mutational correlation between traits (pleiotropy) can evolve as the result of correlated stabilizing selection. We simulated the evolution of gene regulatory networks, for which correlated selection was applied on gene expression, in order to assess whether this same mechanism could drive the evolution of gene co-expression and affect the underlying gene regulatory network topology. We showed that co-expression could indeed evolve, at least qualitatively, as the sign of co-expression responded convincingly to the sign of the correlation of the fitness function. This co-expression was mostly driven by the regulatory distance and nature (activation or inhibition) of interactions between genes. Simulations confirmed that the correlation between gene expressions decreases with the network distance among genes, and that strongest co-expressions are associated with direct interactions. As correlated stabilizing selection promotes mutational co-expression, it thus promotes network topologies where selected genes are closely connected. Taken together, our results exemplify how the shape of a constant, stabilizing fitness function on gene expression can theoretically shape and constrain the topology of the underlying gene network.

Co-expression in gene regulatory network

The use of quantitative genetic tools and theory made possible to dissect the mutational co-expression structure. Theory predicts that the alignment of the direction of mutational co-variation to the (multivariate) direction of the fitness function minimizes genetic load and should be favored by natural selection (Cheverud, 1984; Jones *et al.*, 2014). As expected, simulations based on the multilinear model, parameterized in such a way that epistasis and pleiotropy are isotropic, agree quantitatively with this theoretical prediction. Whether or not stabilizing selection could have the same effect on gene expressions structure is not a trivial question. The mechanisms of gene regulation generate a lot of epistasis and pleiotropy at the gene expression level, but these are not expected to be equally distributed in the phenotypic space. In particular, co-expression among the genes belonging to the same regulatory module is unavoidable, suggesting that evolving expression independence might be more difficult that evolving correlated expression, as it requires to rewire the network and change its modularity.

Numerical optimization of networks confirmed that there exist gene networks which mutational structure match exactly the direction of any of the simulated correlated fitness functions: there are thus no absolute constraints preventing the network model to perform as well as the isotropic multilinear model. Yet, individual-based simulations confirmed the possibility of the evolution of the sign of co-expression, but the direction of the mutational covariance matrix was rarely close to that of the fitness function. Several non-exclusive mechanisms can be proposed to explain such a mismatch: (i) optimal networks might be too far away in the genotypic space to be reached within the time frame of the simulations, (ii) the complexity of the gene network model makes it likely to get trapped at a local optimum, and reaching a better network would require to cross fitness valleys; (iii) the selection pressure for optimized mutational co-expression was too weak and indirect to push networks to evolve pleiotropy (i.e., the effect of a single mutation on the correlation structure is smaller in the gene network model than in the multilinear model); (iv) mutations affecting gene co-expression have direct negative side effects (change in gene expression, decrease of the network stability), so that the fitness peak corresponds to a sub-optimal pleiotropic pattern; (v) mutations affecting gene co-expression have indirect negative side effects (e.g., decrease in mutational robustness) and actually do not decrease the genetic load.

Model approximations

The gene network model used for the simulation is based on the popular 'Wagner' model (Wagner, 1994, 1996). This model has already been explored in evolutionary biology to study the evolvability, the modularity, and the canalization of gene regulatory networks (Siegal and Bergman, 2002; Ciliberti *et al.*, 2007; Rünneburger and Le Rouzic, 2016, see Fierst and Phillips, 2015 for review). This model is computationally fast and requires few parameters in addition to the structure of the regulation network itself. Its lack of realism at the biochemical and cellular level (discrete time steps, no degradation kinetics, no compartments, arbitrary dose-response function) makes it unsuitable for physiological models of known regulation pathways (alternative models could be found in e.g. Karlebach and Shamir, 2008). The relevance of the Wagner model is thus probably limited to theoretical studies of network evolution on evolutionary time scales.

Computational constraints made it necessary to limit the population size ($N = 5,000$) and the number of network genes ($n = 6$) in the simulations (simulation time increases linearly with N and quadratically with n). Real regulatory networks can encompass dozens of transcription factors, and include hundreds of genes in co-regulated modules, which limits our study to small regulatory motifs isolated from a larger network. Yet, the fact that we did not observe any co-expression in genes separated by more than 2 nodes in the network made us confident that our results are valid for networks of larger diameter.

Theoretical approaches to the evolution of gene networks are often limited by the oversimplification of the selective constraints. Gene regulatory networks are expected to control gene expression in a complex temporal and spatial context, accounting to different developmental stages, different cell types, and different environmental conditions. In contrast, the Expression-to-Fitness map in our simulations was simple (multivariate bell-shaped) and gene optimal expression was constant and close to constitutive expression. Simulation results thus need to be interpreted as the evolution of networks in absence of structural and direct selective constraints on the network topology. Realistic patterns

and strengths of selection at the transcriptome level are not really known, but recent statistical and experimental progress makes it likely to obtain reliable empirically-based estimates in the near future (Whitehead and Crawford, 2006; Koch and Guillaume, 2020; Price *et al.*, 2022).

Effect of selection on gene network topology

It is not clear whether structural features of biological networks result from an adaptive process. For instance, modularity can emerge from different mechanisms, including direct selection for efficiency (Clune *et al.*, 2013), adaptation to modular environments (Kashtan and Alon, 2005), indirect selection for evolvability, or mutation bias (Wagner *et al.*, 2007). Large-scale properties, such as scale-freeness, may not have any impact on fitness, and the evolution of gene networks may be dominated by non-adaptive mechanisms, including genetic drift and mutation bias (Lynch, 2007).

Here, we showed theoretically how correlated stabilizing selection on gene expression could deterministically alter the topology of gene networks, by shortening the network distance between genes which expression levels interact at the fitness level. Simulations show that the evolution of regulatory connections as a result of correlated selection can realistically happen in non-restrictive conditions, even if it is difficult to estimate the extent by which real gene networks are affected by this phenomenon.

Nevertheless, correlated selection is not the only form of selection that may promote specific network topologies. Directional selection on a multivariate phenotype may for instance indirectly favor genetic backgrounds generating mutational variation towards the optimum. Fluctuating selection could similarly promote pleiotropy when the optimal phenotypes are correlated among traits (Crombach and Hogeweg, 2008). Independently, fluctuating selection also opens the possibility for the organisms to gather cues about the environment and evolve an adaptive plastic response (Via and Lande, 1985). If, as intuited by Waddington (1942), complex genetic architectures respond to mutational and environmental disturbances through shared molecular mechanisms, it is likely that mutational and environmentally-induced coexpressions will be similar – a property of complex genetic system that could fasten genetic adaptation (Brun-Usan *et al.*, 2021; Chevin *et al.*, 2021). Direct selection for correlated gene expression plasticity among sets of genes thus appears as a potentially powerful force that could drive the evolution of the topology of gene networks, perhaps strong enough to overcome the influence of correlated selection illustrated in our simulations.

Evolution of mutational robustness and pleiotropy

Constant stabilizing selection is often thought to promote stasis, and thus prevent the evolution of traits, but it is also suspected to modify the structure of genetic architectures, in particular through the minimization of the fitness load due to environmental disturbances and mutations (Wagner *et al.*, 1997). When considering phenotypic traits independently, stabilizing selection promotes genetic backgrounds that reduce mutational effects (Rice, 2002; Hermisson *et al.*, 2003). Simulations based in gene network models have reproduced this predictions, and have associated the decrease in the effects of mutations with systematic changes at the network level, including feedback loops, global network size and properties, and redundancy (Masel and Siegal, 2009; Payne and Wagner, 2015; Rünneburger and Le Rouzic, 2016).

From a multivariate quantitative genetics point of view, organizing the phenotypic space along measurable traits is a necessary consequence of how phenotypes are estimated empirically, but most models can easily be redefined for any linear combination of traits. For instance, when the bivariate multilinear model is parameterized in such a way that epistasis and pleiotropy are identical in all directions of the phenotypic space, mutational effects evolve indifferently towards robustness or pleiotropy depending on the direction of the fitness function (Jones *et al.*, 2014 and our simulations). This is where the quantitative genetics predictions break when applied to gene networks: in network models, genes both define observable phenotypes and structure the interaction patterns. As a consequence, correlated selection can also induce the evolution of pleiotropic gene expression, as predicted by theory, but the mechanisms involved into this response are different from those involved into the evolution of robustness. Our simulations pointed out the major effect of the network distance on co-expression, showing that evolving pleiotropic gene expression requires to rewire the network and reduce the distance (and possibly the sign of regulatory interactions) between co-selected genes.

The influence of correlated selection on the topology and on coexpression in real gene networks remains virtually unknown. Adaptive (e.g., plasticity) and non-adaptive (e.g., mutation bias) forces are

also prone to alter network topology, and condition the long-term evolvability of these complex genetic architectures. Since gene networks can be shaped by selection, but can also constrain the mutational availability of evolutionary path, the long-term influence of past environment on phenotypic variability and evolvability remains a challenging question in both quantitative genetics and evolutionary systems biology.

Acknowledgment

We thank Anne Genissel for valuable advice and discussion.

Funding

AP was supported by SDSV doctorate school. JG was supported by a French National Center for Scientific Research (CNRS) fellowship: ANR-80Prime TransIA.

Conflict of interest

The authors declare no conflict of interest.

References

Azevedo R, Lohaus R, Srinivasan S, Dang K, Burch C. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature. 440:87–90.

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. 2004. Structure and evolution of transcriptional regulatory networks. Current Opinion in Structural Biology. 14:283–291.

Bellucci E, Bitocchi E, Rau D, Rodriguez M, Biagetti E, Giardini A, Attene G, Nanni L, Papa R. 2014. Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. Genomics of Plant Genetic Resources. pp. 483–507.

Bergman A, Siegal M. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature. 424:549–552.

Brun-Usan M, Rago A, Thies C, Uller T, Watson RA. 2021. Development and selective grain make plasticity 'take the lead' in adaptive evolution. BMC Ecology and Evolution. 21:1–17.

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

Carter AJ, Hermisson J, Hansen TF. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. Theoretical Population Biology. 68:179–196.

Cheverud JM. 1984. Quantitative genetics and developmental constraints on evolution by selection. Journal of Theoretical Biology. 110:155–171.

Chevin LM, Leung C, Le Rouzic A, Uller T. 2021. Using phenotypic plasticity to understand the structure and evolution of the genotype – phenotype map. Genetica. .

Ciliberti S, Martin OC, Wagner A. 2007. Innovation and robustness in complex regulatory gene networks. Proceedings of the National Academy of Sciences of the United States of America. 104:13591–13596.

Clune J, Mouret JB, Lipson H. 2013. The evolutionary origins of modularity. Proceedings of the Royal Society b: Biological sciences. 280:20122863.

Crombach A, Hogeweg P. 2008. Evolution of evolvability in gene regulatory networks. PLoS Computational Biology. 4.

Espinosa-Soto C. 2016. Selection for distinct gene expression properties favours the evolution of mutational robustness in gene regulatory networks. Journal of Evolutionary Biology. 29:2321–2333.

Espinosa-Soto C. 2018. On the role of sparseness in the evolution of modularity in gene regulatory networks. PLoS Computational Biology. 14:1–24.

Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER. 2004. A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. The Plant Cell. 16:2923–2939.

Fagny M, Austerlitz F. 2021. Polygenic adaptation: Integrating population genetics and gene regulatory networks. Trends in Genetics. 37:631–638.

Fierst JL, Phillips PC. 2015. Modeling the evolution of complex genetic systems: The gene network family tree. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 324:1–12.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature. 525:372–375.

Gilad Y, Rifkin SA, Pritchard JK. 2008. Revealing the architecture of gene regulation: the promise of eqtl studies. Trends in genetics. 24:408–415.

Guelzim N, Bottani S, Bourgine P, Képès F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. Nature Genetics. 31:60–63.

Hansen TF, Wagner GP. 2001. Modeling genetic architecture: A multilinear theory of gene interaction. Theoretical Population Biology. 59:61–86.

Hermisson J, Hansen TF, Wagner GP. 2003. Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. American Naturalist. 161:708–734.

Huang Y, Lack JB, Hoppel GT, Pool JE. 2021. Parallel and population-specific gene regulatory evolution in cold-adapted fly populations. Genetics. 218.

Jallet A, Le Rouzic A, Genissel A. 2020. Evolution and Plasticity of the Transcriptome Under Temperature Fluctuations in the Fungal Plant Pathogen *Zymoseptoria tritici*. Frontiers in Microbiology. 11:1–18.

Jones AG, Arnold SJ, Bürger R. 2007. The mutation matrix and the evolution of evolvability. Evolution. 61:727–745.

Jones AG, Bürger R, Arnold SJ. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. Nature Communications. 5.

Karlebach G, Shamir R. 2008. Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Ciology. 9:770–780.

Kashtan N, Alon U. 2005. Spontaneous evolution of modularity and network motifs. Proceedings of the National Academy of Sciences. 102:13773–13778.

Koch EL, Guillaume F. 2020. Restoring ancestral phenotypes is a general pattern in gene expression evolution during adaptation to new environments in *Tribolium castaneum*. Molecular Ecology. 29:3938–3953.

Le Rouzic A, Álvarez-Castro JM, Hansen TF. 2013. The evolution of canalization and evolvability in stable and fluctuating environments. Evolutionary Biology. 40:317–340.

Leclerc R. 2008. Survival of the sparsest: robust gene networks are parsimonious. Molecular Systems Biology. 4:213.

Lynch M. 2007. The evolution of genetic networks by non-adaptive processes. Nature Reviews Genetics. 8:803–813.

Masel J, Siegal ML. 2009. Robustness: mechanisms and consequences. Trends in Genetics. 25:395–403.

Mayer C, Hansen TF. 2017. Evolvability and robustness: a paradox restored. Journal of Theoretical Biology. 430:78–85.

Michaelson JJ, Loguercio S, Beyer A. 2009. Detection and interpretation of expression quantitative trait loci (eqtl). Methods. 48:265–276.

Nash JC, Varadhan R. 2011. Unifying optimization algorithms to aid software system users: optimx for R. Journal of Statistical Software. 43:1–14.

Olson EN. 2006. Gene regulatory networks in the evolution and development of the heart. Science. 313:1922–1927.

Ouma WZ, Pogacar K, Grotewold E. 2018. Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. PLoS Computational Biology. 14:e1006098.

Pavličev M, Cheverud JM. 2015. Constraints evolve: context dependency of gene effects allows evolution of pleiotropy. Annual Review of Ecology, Evolution, and Systematics. 46:413–434.

Payne JL, Wagner A. 2015. Mechanisms of mutational robustness in transcriptional regulation. Frontiers in Genetics. 6.

Philippe N, Crozat E, Lenski RE, Schneider D. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. BioEssays. 29:846–860.

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022. Detecting signatures of selection on gene expression. Nature Ecology & Evolution. pp. 1–11.

Rhoné, B.and Brandenburg JT, Austerlitz F. 2011. Impact of selection on genes involved in regulatory network: A modelling study. Journal of Evolutionary Biology. 24:2087–2098.

Rice SH. 2002. A general population genetic theory for the evolution of developmental interactions. Proceedings of the National Academy of Sciences. 99:15518–15523.

Rünneburger E, Le Rouzic A. 2016. Why and how genetic canalization evolves in gene regulatory networks. BMC Evolutionary Biology. 16:1–11.

Sgrò CM, Hoffmann AA. 2004. Genetic correlations, tradeoffs and environmental variation. Heredity. 93:241–248.

Shashikant T, Khor JM, Ettensohn CA. 2018. From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. Genesis. 56:e23253.

Siegal ML, Bergman A. 2002. Waddington's canalization revisited: Developmental stability and evolution. Proceedings of the National Academy of Sciences of the United States of America. 99:10528–10532.

Sinervo B, Svensson E. 2002. Correlational selection and the evolution of genomic architecture. Heredity. 89:329–338.

Svensson EI, Arnold SJ, Bürger R, Csilléry K, Draghi J, Henshaw JM, Jones AG, De Lisle S, Marques DA, McGuigan K *et al*. 2021. Correlational selection in the age of genomics. Nature Ecology and Evolution. 5:562–573.

Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. 2012. Reshaping of the maize transcriptome by domestication. Proceedings of the National Academy of Sciences. 109:11878–11883.

Taylor TB, Shepherd MJ, Jackson RW, Silby MW. 2022. Natural selection on crosstalk between gene regulatory networks facilitates bacterial adaptation to novel environments. Current Opinion in Microbiology. 67:102140.

Van Noort V, Snel B, Huynen MA. 2004. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. EMBO reports. 5:280–284.

Verta JP, Jones FC. 2019. Predominance of *cis*-regulatory changes in parallel expression divergence of sticklebacks. eLife. 8:e43785.

Via S, Lande R. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution. 39:505–522.

Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. Nature. 150:563–565.

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 of the United States of America. 91:4387–4391.

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

Wagner A. 2008. Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society B: Biological Sciences. 275:91–100.

Wagner A, Wright J. 2007. Alternative routes and mutational robustness in complex regulatory networks. BioSystems. 88:163–172.

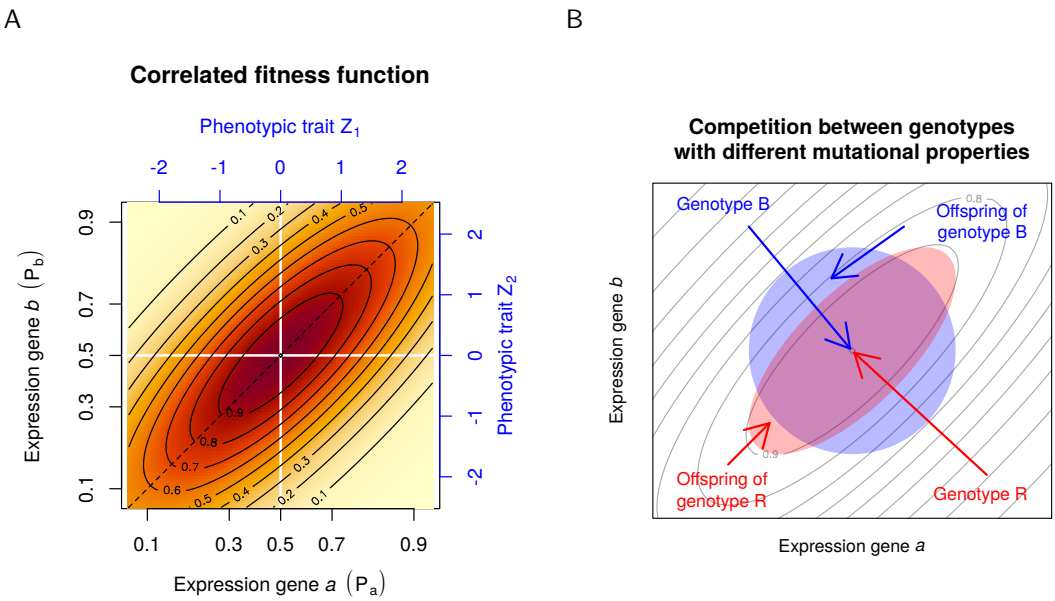
Wagner GP, Booth G, Bagheri-Chaichian H. 1997. A population genetic theory of canalization. Evolution. 51:329–347.

Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. Nature Reviews Genetics. 8:921–931.

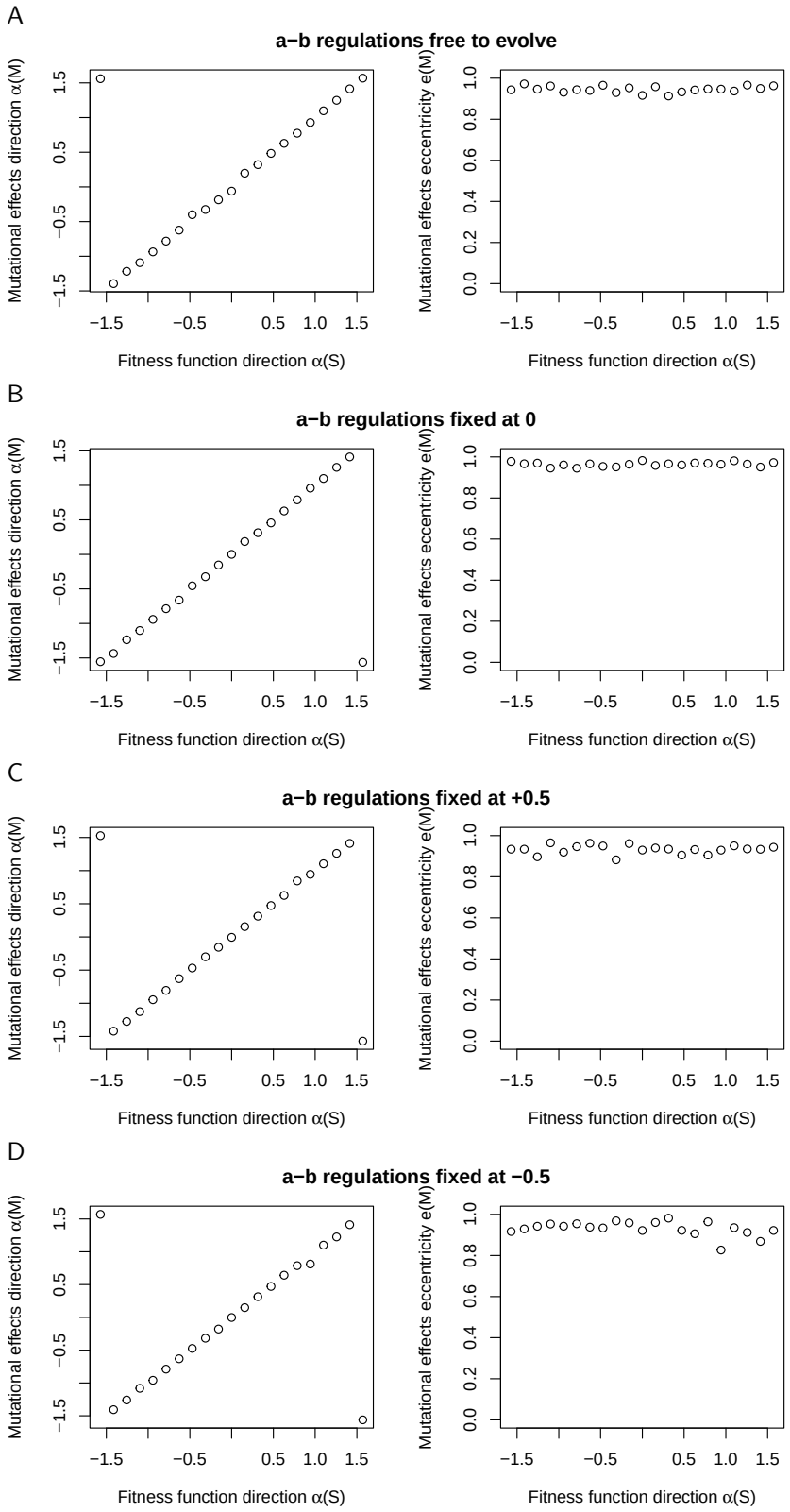
Walsh B, Lynch M. 2018. *Evolution and Selection of Quantitative Traits*. Oxford University Press. Oxford.

Whitehead A, Crawford DL. 2006. Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences. 103:5425–5430.

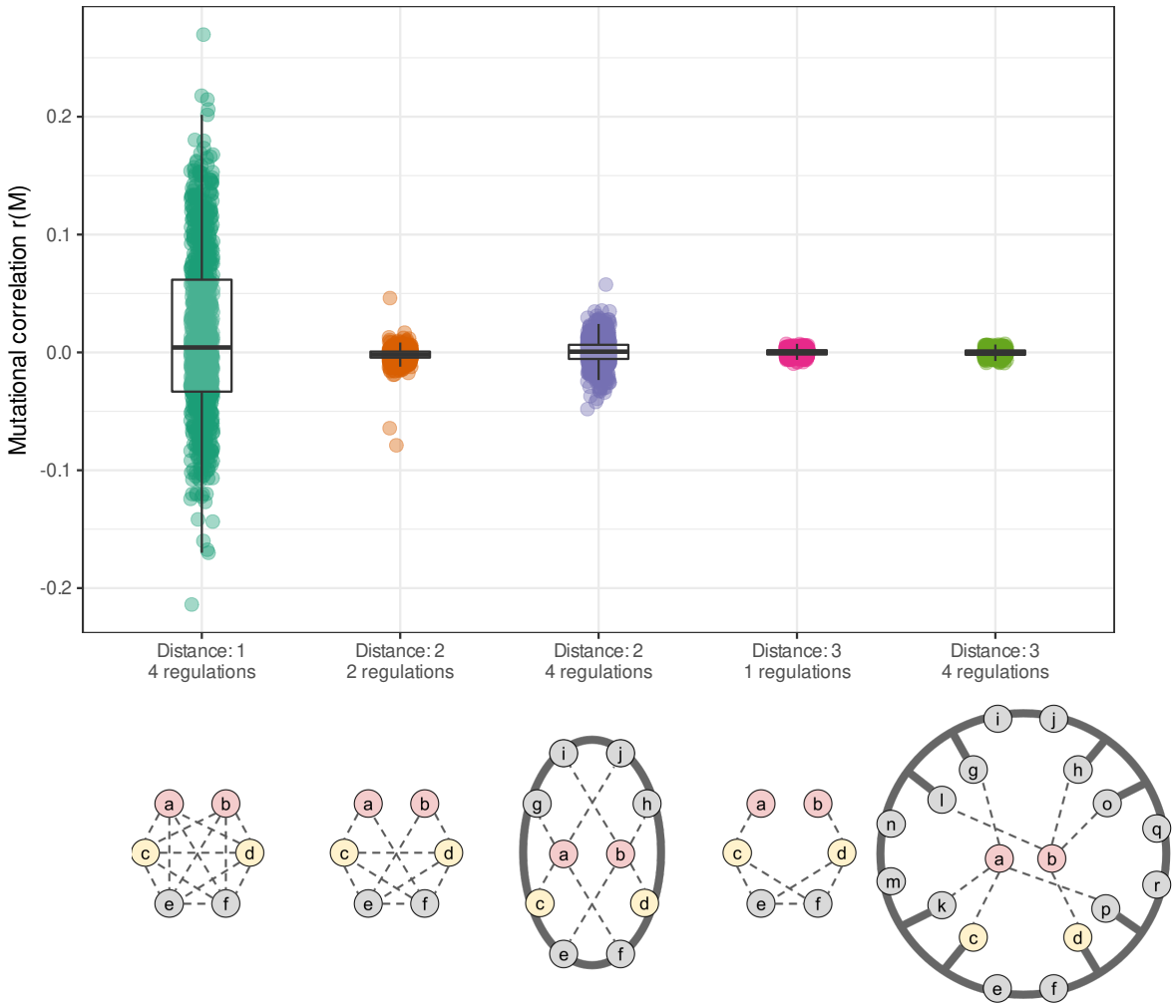
Supplementary material



Supplementary Figure 1: A: representation of the bivariate correlated fitness function used in the simulations (here, with an direction $\alpha(\mathbf{M}) = \pi/4$). Axes display both gene expression levels ($\mathbf{P}$) scaled between 0 (no expression) and 1 (full expression) in black, and the corresponding rescaled phenotype $\mathbf{Z}$ from which fitness is computed (in blue). White lines highlight the optimal expression level $\mathbf{\Theta}$ (for which the fitness is maximal). Any deviation from the optimal expression is penalized, but the penalty is weaker when both genes are over- or under-expressed together. B: Cartoon representation of the advantage of a genotype in which mutational effects are correlated (red, R) over a genotype in which mutational effects are uncorrelated (blue, B). Both genotypes display the optimal gene expression and thus have the same fitness; mutant offspring from both genotypes deviate from the optimum within the same range, but due to the genetic correlation, the average fitness of the mutant offspring (iso-fitness lines in gray) from genotype R is higher than the offspring from genotype B: the R lineage will progressively replace the B lineage in the population.



Supplementary Figure 2: Properties of $\mathbf{M}$ matrices obtained by mathematical optimization from different network topology. A: a and b can have any values. B: no regulations between a and b. C: positive regulations between a and b. D: negative regulations between a and b.



Supplementary Figure 3: Distribution of mutational correlations with different networks (same color code as in Figure 4). Two topologies were compared for the same network distance : one with the number of genes conserved (6 genes), and one with the number of edges connected to a and b conserved (4 possible regulatory interactions). Genes located on grey circle are all connected to each other; genes connected to the grey circle can interact with every gene on the circle but not with each other.