Living with noise:
The evolution of gene expression noise in gene regulatory networks

Dissertation
in fulfillment of the requirements for the degree
Doctor rerum naturalium
of the Faculty of Mathematics and Natural Sciences
at the Christian Albrechts University of Kiel

Submitted by
Nataša Puzović

Max Planck Institute for Evolutionary Biology
Plön, Germany
June, 2023.

First examiner:	Prof. Dr. Julien Yann Dutheil
Second examiner:	Prof. Dr. rer. nat. Tal Dagan
External examiner:	Prof. Dr. Jan Baumbach
Chairperson:	Prof. Dr. Matthias Leippe
Date of oral examination:	19.07.2023.
Place of oral examination:	Christian-Albrechts-Universität zu Kiel, Germany

Zusammenfassung

Einer der Grundpfeiler der Evolutionsbiologie ist die Untersuchung, wie sich Merkmale von Organismen im Laufe der Zeit verändern. Technologische Fortschritte in den letzten zwanzig Jahren haben es uns ermöglicht, die Variation eines wichtigen Merkmals, der Genexpressionsebene, auf Einzelzellauflösung zu untersuchen. Eine der Ursachen für die Variation des Genexpressionsniveaus ist Genexpressionsrauschen, ein Ergebnis der angeborenen Stochastizität des Genexpressionsprozesses. Genexpressionsrauschen ist genspezifisch und kann durch Selektion eingestellt werden, aber was die Entwicklung des genspezifischen Expressionsrauschens antreibt, bleibt eine offene Frage.

In dieser Dissertation untersuche ich den selektiven Druck und die Evolvierbarkeit von genspezifischem Expressionsrauschen in Genregulationsnetzwerken. Ich verwende Evolutionssimulationen, indem ich Mutations-, Rekombinations- und Reproduktionsrunden auf Populationen von Modellnetzwerken zur Genregulation in verschiedenen Selektionsszenarien anwende.

Im ersten Kapitel untersuche ich die Reaktion von genspezifischem Expressionsrauschen in Genregulationsnetzwerken in konstanten Umgebungen, die eine stabilisierende Selektion auf der Genexpressionsebene erfordern. In diesen Simulationen konnte sich das Ausdrucksrauschen über Tausende von Generationen hinweg weiterentwickeln. Die Wahrscheinlichkeit, auf die Auswahl zu reagieren, und die Stärke der selektiven Reaktion wurden durch lokale Netzwerkzentralitätsmetriken beeinflusst. Gene mit höheren Zentralitätsmetriken hatten nämlich eine höhere Wahrscheinlichkeit, auf die Selektion zu reagieren, und eine stärkere Reduzierung des genspezifischen Expressionsrauschens als Reaktion auf die stabilisierende Selektion. Darüber hinaus beeinflussten globale Netzwerkmerkmale wie Netzwerkdurchmesser, Zentralisierung und durchschnittlicher Grad die durchschnittliche Expressionsvarianz und den durchschnittlichen Selektionsdruck, der auf konstituierende Gene wirkt.

Im zweiten Kapitel untersuche ich die Reaktion des mittleren Genexpressionsniveaus und des genspezifischen Expressionsrauschens in isolierten Genen und Genen in Genregulationsnetzwerken in sich verändernden Umgebungen. In diesen Simulationen konnten sich sowohl das Genexpressionsniveau als auch das genspezifische Expressionsrauschen über Tausende von Generationen unter gerichteter oder schwankender Selektion entwickeln. Das genspezifische Expressionsrauschen von Genen nahm bei schwankender Selektion zu, was auf die Entwicklung einer Bet-Hedging-Strategie hinweist. Unter direktionaler Selektion nahm das genspezifische Expressionsrauschen vorübergehend zu, was zeigt, dass Expressionsrauschen eine Rolle im Anpassungsprozess hin zu einem neuen mittleren Expressionsoptimum spielt. In beiden selektiven Szenarien reagierten Zielgene, also Gene, die von anderen Genen reguliert werden, eher als Regulatorgene.

Diese Ergebnisse zeigen, dass die Selektion auf Netzwerkebene zu einem unterschiedlichen Selektionsdruck auf Genebene führt und dass lokale und globale Netzwerkeigenschaften von Genregulationsnetzwerken ein wesentlicher Bestandteil der genspezifischen Expressionsrauschentwicklung sind. Sie zeigen weiterhin, dass erhöhtes Expressionsrauschen als adaptive Strategie genutzt werden kann und dass der Hintergrund des Gennetzwerks evolutionäre Einschränkungen für die Entwicklung des mittleren Expressionsniveaus und des genspezifischen Expressionsrauschens mit sich bringt. Diese Ergebnisse stellen einen Fortschritt beim Verständnis der Entwicklung des Genexpressionsrauschens in Gennetzwerken dar.

Summary

One of the keystones of evolutionary biology is the study of how organismal traits change in time. Technological advancements in the past twenty years have enabled us to study the variation of an important trait, gene expression level, at single cell resolution. One of the sources of gene expression level variation is gene expression noise, a result of the innate stochasticity of the gene expression process. Gene expression noise is gene-specific and can be tuned by selection, but what drives the evolution of gene-specific expression noise remains an open question.

In this thesis, I explore the selective pressure and evolvability of gene-specific expression noise in gene regulatory networks. I use evolutionary simulations by applying rounds of mutation, recombination and reproduction to populations of model gene regulatory networks in different selection scenarios.

In the first chapter, I investigate the response of gene-specific expression noise in gene regulatory networks in constant environments, which imposes stabilizing selection on gene expression level. In these simulations, the expression noise was allowed to evolve over thousands of generations. The probability of responding to selection and the strength of the selective response was affected by local network centrality metrics. Namely, genes with higher centrality metrics had higher probability of responding to selection and a higher reduction in gene-specific expression noise in response to stabilizing selection. Furthermore, global network features, such as network diameter, centralization and average degree affected the average expression variance and average selective pressure acting on constituent genes.

In the second chapter, I investigate the response of mean gene expression level and gene-specific expression noise in isolated genes and genes in gene regulatory networks in changing environments. In these simulations, both gene expression level and gene-specific expression noise were allowed to evolve over thousands of generations under directional or fluctuating selection. Gene-specific expression noise of genes increased under fluctuating selection, indicating the evolution of a bet-hedging strategy. Under directional selection gene-specific expression noise transiently increased, showing that expression noise plays a role in the adaptation process towards a new mean expression optimum. In both selective scenarios, target genes, genes regulated by other genes, were more likely to respond than regulator genes.

These results show that selection at the network level leads to differential selective pressure at the gene level, and local and global network characteristics of gene regulatory networks are an essential component of gene-specific expression noise evolution. They further demonstrate that increased expression noise can be utilized as an adaptive strategy and that the gene network background imposes evolutionary constraints on the evolution of mean expression level and gene-specific expression noise. These findings represent a step forward in understanding the evolution of gene expression noise in gene networks.

Sky Blue, Wassily Kandinsky (1940)

“I would like to see anyone, prophet, king or God, convince a thousand cats to do the same thing at the same time.”

- Neil Gaiman, *A Dream of a Thousand Cats* (1990)

Foreword

The richness of macroscopic life on Earth is apparent to anyone who has spent any time out in nature. Take a walk in a forest and you will see a palette of living creatures: dozens of species of trees and shrubs, bees pollinating flowers, differently colored beetles roaming around, squirrels scattering up the tree trunks, birds perched atop the branches, paw prints of weasels and martens, and perhaps a well-hidden badger sett if you have a keen eye. Invisibly, below the ground, there will be hundreds of fungi and bacteria eating away at the soil. In all likelihood, in your neighbourhood there will be hundreds of species of plants and animals and even more prokaryotes and fungi, all coexisting together. If we consider the neighbourhood at it's largest scale, the planet, it is estimated that there are around eight million species on Earth, of which only around one million are described. It is easy to be awestruck by the sheer amount of visible diversity of species on Earth.

However, the limitations of our naked eyes make us overlook an equally impressive assortment - the same amount of richness and diversity we can see at the macroscopic level can be found at the microscopic level within living creatures. Living beings consist of cells, which harbor a fascinating abundance of diversity of their own kind. Namely, in each cell there are thousands of different molecules that orchestrate its life, molecules big and small, which shuffle and bump into each other and create products useful to the cell. These molecules, just like living creatures in the forest, come in all shapes and sizes. If we could shrink ourselves into a homunculus and take a walk through a cell, we would be struck by diversity just like in the forest. Structural proteins will be the beams holding the cell components together, thousands of enzymes will be catalyzing metabolic reactions such as digestion and respiration, tube-like ion channels will be shuttling ions between the cell and its environment, motor proteins will be transporting their cargo up and down the microtubule filaments to distribute it to different areas of the cell. Proteasomes, protein complexes that look and function like garbage cans, will be floating around suspended in cytoplasm, waiting to recycle any incorrectly made proteins. Inside the center of the cell, there will be a two meter long chain of nucleotides, holding the genetic information of the cell and tightly bound around barrel-like proteins called histones. The way in and out of the center of the cell, the nucleus, will be gated by nuclear pores, intricate molecular structures consisting of over thirty different kinds of proteins, which have three rings and the outer ring has filaments jutting outside, while the inner ring resembles an actual basket. Altogether, in the landscape of a cell there are dozens of thousands of distinct proteins, over a thousand distinct lipids, an unknown number of distinct carbohydrates and nucleic acids. Micromolecules included, the total number of molecular species in a cell can easily yield over a hundred thousand distinct species. Most of these species are undescribed, much like in the macroscopic world. Surely, the inner world of the cell is a rich and mysterious tapestry, which we have begun to unravel only a couple of decades ago.

What is perhaps most interesting in this picture of the inner functioning of the cell is the lack of a central command unit. Even though so many processes need to be tightly controlled for the cell to survive, there is no central mainframe controlling the behaviour of each molecule in the cell. Instead, the cell is a self-organized structure, in which molecules diffuse freely inside the cell body and within the cell compartments. Naturally, this means that there is an inevitable randomness within the cell, a randomness that is a feature of all life. Proteins, and molecules in general, are constantly synthesized with small differences, or in too low or too high amounts. Other molecules might take long to find their targets, or bind by mistake

to wrong targets. Keeping in mind the set of tens of thousands of molecular species working together to keep the cell alive, it is tempting to think of a cell as a perfectly organized machine, in which every molecular behaviour is intentional and controlled. In fact, it is the opposite. The orchestra of tens of thousands of molecular species has no conductor, and more closely resembles a chaotic jungle in which, on average, the necessary is carried out, and many mistakes are made along the way. One might think then - how on Earth do cells, and living beings in general, survive like this? One might even stumble into a PhD thinking about it.

How do we go about answering this question? There are multiple ways to tackle a scientific question. A good start is to observe what is present in nature and make inferences based on our observations. Sometimes, it is difficult or impossible to manipulate a biological sample to direct an experimental study. In that case, scientific modelling comes in. Scientific models are simplified representations of natural phenomena that capture some key aspect of their behaviour, which we can use to explore the behaviour in different conditions. As simplified representations, all models are wrong (following George Box's aphorism), in the sense they do not perfectly recreate every behavior of the real phenomena. They are, nonetheless, useful. Models help us save time and resources that would otherwise be spent on experiments, or enable us to study what is impossible by experimental means. For example, an aircraft engineer will come up with different wing prototypes and test their aerodynamics using computational models by simulating the movement of air across it, instead of physically building the prototypes and putting it in an air chamber. Pharmacists use molecular and pharmacokinetic models to come up with new drugs and test their efficiency, before actually synthesizing them in the lab. The behavior of the model gives us an implication of the real system would have behaved, or it might suggest that there is something in nature we have not noticed so far. The existence of Neptune was predicted based on the mathematical model of Sir Isaac Newton's law of gravity, and the existence of the Higgs boson was predicted based on a theoretical model about fifty years before it was experimentally observed. The modelling and simulation approach is used in this thesis. A simple model of interacting molecules in the cell is used to study how randomness in cells evolves and affects the evolution of the molecular components.

Contents

Chapter 1

Background: The inherent randomness of living beings

"Each cell is, within certain limits, an Individual, an independent Whole."
– Theodor Schwann, *Microscopical Researches (1839)*

In this thesis, I study the evolution of an aspect of biological randomness, gene expression noise, in the context of gene networks. I will first introduce the concept of gene expression, then the concept of gene networks, and lastly, list reasons to investigate noise in biological systems.

1.1 Gene expression

1.1.1 History of studying genes and gene expression

The fact that offspring resemble their parents has been known since before empirical science, but what exactly determines the traits of an organism and how it is propagated into the next generation remained a mystery. The basic concepts of inheritance were set in 1886 by Gregor Mendel using pea experiments, which showed that offspring inherit single units of inheritance which determine some phenotypic traits from their parents (Mendel, 1865). Mendel's principles of inheritance were largely unnoticed until they were rediscovered in 1900 by Hugo de Vries, Carl Correns, and Erich von Tschermak, who have reported to had reached similar conclusions. Around the same time, the term *gene* was introduced by Wilhelm Johannsen to denote the basic physical unit of inheritance. He also introduced the terms *genotype* and *phenotype*, albeit in a slightly different meaning than they are used today, to distinguish between the hereditary genetic material and the observable traits of an organism. However, the molecular nature of genes was unknown and for a long time it was hypothesized that the genetic material is chemically a protein. The molecular nature of genes was not discovered until the work of Oswald Avery, Colin MacLeod and Maclyn McCarty in 1944 (Avery et al., 1944) and Alfred Hershey and Martha Chase in 1952 (Hershey and Chase, 1952), which showed that the genetic material is a deoxyribonucleic acid, or DNA. A year later, in 1953, the double helix chain-like structure of DNA is published, which solved the problem of encoding a large amount of genetic material in a molecule using only four nucleotides (Watson and Crick,

1953). Finally, it was accepted that organisms carry genes, encoded in their DNA, which fully, or to some extent, determine their phenotypic traits.

How the genetic information stored in genes is realized into phenotypic traits has been explained by the central dogma of molecular biology, first formulated by Francis Crick in 1958, which posits that genetic information flows from DNA through RNA to proteins (Crick, 1958). The multi-step process by which information in a gene is used to synthesize a gene product is called *gene expression*. Gene products can be proteins or non-coding RNA, and usually, the measure of gene expression is the amount of gene product in the population or individual cell. The process of gene expression is comprised of many reactions in which specific molecules must bind to each other to facilitate a reaction and intracellular molecules mostly move by diffusion, making the binding event, more or less, a random process. Also, these molecules are present in low amount, increasing the susceptibility to random fluctuations. In other words, the biochemical reactions that constitute the gene expression process are inherently noisy, yielding a nondeterministic outcome in terms of expression level. The variation in the expression level resulting from the randomness of gene expression, which persists even among genetically identical cells living in an identical environment, has been termed *expression noise* (Elowitz, 2002). When looking at single cells, the genome can be considered the cell's genotype and the expression levels of all genes the cell's phenotype.

In classical biology, genetically identically cells living in an identical environment are assumed to have identical phenotypes, disregarding inherent randomness. On the other hand, intracellular randomness has been predicted from basic physical principles since the 1940s (Schrödinger, 1944) and phenotypic heterogeneity reported (Delbrück, 1945; Novick and Weiner, 1957; Maloney and Rotman, 1973; Spudich and Koshland, 1976). However, even though the dogma of a completely deterministic phenotype played a role in disregarding phenotypic randomness, there were obstacles in studying noise due to the scarcity of experimental tools available to study cell individuality. Gene expression was usually measured qualitatively (whether a gene is expressed or silenced) or quantitatively (as average protein level).

A decisive pivot happened in 2002, with the study of the variation of *E. coli* cells growing in the same environment by Michael Elowitz and coauthors (Elowitz, 2002). They presented experimental evidence of inherent stochasticity of gene expression creating phenotypic heterogeneity, as well as a way to disentangle the intrinsic and intrinsic sources of gene expression noise. This study highlighted the non-negligible effect expression noise has on phenotypic variation and brought the topic of expression noise to the limelight. Still, experimental methodology at the time allowed the measurement of noise on only few genes at the time, and even though some experimental datasets were reported (Newman et al., 2006; Bar-Even et al., 2006), genome-wide noise measurements were largely unavailable.

Recent technological advancements in the form of single-cell sequencing have opened the possibility of deeply exploring cells at the individual level. First single-cell RNA sequencing was reported in 2009 (Tang et al., 2009), opening the door to genome-wide measurements of expression levels at the resolution of single cells. Single-cell sequencing (single-cell transcriptomics and genomics) was named "Method of the Year for 2013" by Nature Methods for its potential to study heterogeneity of cells at the individual level (Nature, 2014). Since then, numerous studies working on unravelling the factors shaping the expression level mean and expression noise in single cells have been published.

1.1.2 What shapes expression noise levels (molecular and evolutionary causes)?

There are several reported molecular factors that influence expression noise levels, at different stages of the gene expression process. At the epigenetic level, certain chromatin features, such as presence of chromatin remodelling complexes (Newman et al., 2006) and gene proximity resulting in common chromatin dynamics (Sun and Zhang, 2019) have been reported to influence expression noise. At the transcriptional level, presence of a TATA box (Newman et al., 2006), the promoter shape (Sigalova et al., 2020), presence and number of transcription factor binding sites (Sharon et al., 2014), transcription factor binding dynamics (Azpeitia and Wagner, 2020), presence of transcription factor decoy binding sites (Dey et al., 2020), and transcription rate have been reported to affect noise. At the translational level, miRNA targetting (Schmiedel et al., 2015), mRNA lifetime, and translation rate. At the posttranslational level, the protein degradation rate and compartmentalization of proteins by phase separation have also been shown to influence noise (Klosin et al., 2020).

At evolutionary time scales, some factors shaping expression noise have been identified. Evidence of selection on expression noise was first seen in the fact that dosage-sensitive genes (Lehner, 2008) and essential genes exhibit lower levels of expression noise (Fraser et al., 2004; Wang and Zhang, 2011). Intrinsic noise was also reported to correlate with the strength of selection acting on the encoded protein. Namely, proteins with a lower ratio of non-synonymous over synonymous substitution rate (Ka/Ks) have a lower level of expression noise (Barroso et al., 2018). Sequence conservation is one proxy for evolutionary constraints, and it is one of the predictive features of low expression variation. Conserved genes between *Drosophila* and human are significantly less variable, and are highly enriched for broad promoters (Sigalova et al., 2020). Changes in the expression noise of a single gene may be either beneficial or deleterious, depending on how far its mean expression is from the optimal expression level (Duveau et al., 2018). Expression noise is deleterious if the mean expression level is close to the optimal, as higher variation, in this case, generates a larger number of less fit individuals, reducing the population fitness. Conversely, expression noise can be beneficial if the mean expression level is far from the optimum, as noisy genes are more likely to generate cells with an expression level closer to the optimum. Noisy gene expression can thus be part of a bet-hedging strategy and was observed in genes involved in immune and environmental response (Beaumont et al., 2009; Bódi et al., 2017; Farquhar et al., 2019; Nevozhay et al., 2012). On the other hand, low-variability genes are enriched for housekeeping and developmental functions in *Drosophila* (Sigalova et al., 2020). The fitness cost of changes in the level of expression noise in the fitness landscapes of ≈ 30 yeast genes have been shown to be on the same order as fitness costs of changes in mean expression level (Schmiedel et al., 2019). Since the fitness effect of different levels of expression noise can be as detrimental as different mean expression levels, which are thought to be extensively under selection (Gilad et al., 2006), it can be assumed that expression noise is extensively under selection genome-wide. Prevalent selection on expression noise has been demonstrated in naturally segregating promoter variants of *E. coli* (Vlková and Silander, 2022).

1.1.3 From genotypes to phenotypes through networks

How do genes constituting the genotype translate into the actual organism, the phenotype? Inspired by Mendelian genetics, early population genetics assumed genes to be independent and equivalent entities shaping the phenotype, and that evolution happens by changing the frequency of gene variants. This framing has been jokingly termed 'beanbag genetics' by Ernst Mayr (Mayr, 1963), where the population of organisms is described as a bag full of beans and mutations represent an exchange of one bean for another, and famously led to a public dispute with J. B. S. Haldane (Haldane, 1964). This view also assumes a deterministic relationship between genotype and phenotype, in which the genetic material of the individual fully determines the phenotypic traits, and by extension, the gene pool determines the phenotypic traits of the population. Mayr criticized the assumption for its disregard of the complexities of organismal development, underlying that selection acts on phenotypes, whose development cannot be disregarded by assuming a simple additive effect of each gene. The interactions between genes have since started to be taken into consideration, and a commonly used concept when studying the relationship between the genotype and phenotype is the 'genotype-phenotype map', introduced by Pere Alberch in 1991 (Alberch, 1991). The relationship between the genotype and phenotype in this case is understood not as a one-to-one mapping, but through a mapping function which maps parameter space to a phenotypic space. Investigating the relationship between the genotype and phenotype requires a more nuanced approach in which the complexities of development have been more accurately modelled by biological networks, *i.e.* a shift from focusing on genes to focusing on gene networks (Pigliucci, 2010).

1.2 Biological networks

Biological networks are representations of relationships between biological entities, such as proteins, genes, species in the food web, and others. Most commonly studied networks are protein-protein interaction networks, DNA-protein interaction networks (gene regulatory networks), food webs and neuronal networks. Computational models allow us to investigate the behaviour of biological systems which may be difficult or impossible to do by experimental means. The models used to study each network type differ, so here I present a brief overview of the models used to study gene regulatory networks (GRNs), which are the focus of this thesis. Gene regulatory networks are representations of the regulatory interactions between genes, *i.e.* a directed graph that defines which gene influences the expression level of the genes it is connected to.

1.2.1 Models of gene regulatory networks

Numerous models and methodologies for reconstructing GRNs from single-cell RNA-seq data have been reported (*e.g.* Oubounyt et al. (2023); Mao et al. (2022); Shu et al. (2021); Jackson et al. (2020); Huang et al. (2017)), even some that take phylogenetic information into account for increased accuracy (Mignone et al., 2020; Koch et al., 2017). Nonetheless, the reconstructed GRNs are far from being complete (Röttger et al., 2012) and being consistent with gene expression data (Larsen et al., 2019). The models used for simulating the behaviour of GRNs are, however, different from the models used in reconstructing them from real data, apart from one recent study reports a model that can be used simultaneously for reconstruction and sim-

ulation of GRNs (Ventre et al., 2023). Common models used to simulate GRNs can be roughly classified into qualitative and quantitative models (Barbuti et al., 2020).

Qualitative models used to simulate GRNs include logical models, such as Boolean networks or Petri nets, in which the expression levels are represented as binary states (on or off). For example, Boolean networks have been used to study the cell-cycle regulatory network in *Saccharomyces cerevisiae* (Li et al., 2004), and Petri nets have been used to study the sporulation regulatory network of *Bacillus subtilis* (Steggles et al., 2007). These models are computationally and parametrically undemanding, but simplistic, and have been used less often in the past decade.

Quantitative models used to simulate GRNs include differential equations and the stochastic simulation algorithm. These models represent expression levels as continuous values, such as the concentrations of molecular species or their absolute number. Most commonly used differential equation models are ordinary differential equation (ODE) models. For example, the regulatory network response of yeast to hyperosmotic shock was implemented as an ODE model (Klipp et al., 2005). The values for ODE parameters are rarely known, so due to the number of parameters that need to be estimated differential equation models are suitable for modeling only small-scale networks. They also assume a deterministic gene expression process, making them unable to represent stochasticity in gene expression. Gillespie's stochastic simulation algorithm (SSA) (Gillespie, 1977) models the randomness of reactions in molecular networks by implementing molecular reactions as probabilistic events. It is computationally demanding, so approximations to SSA which exchange accuracy for speed have been developed, such as tau-leaping (Gillespie, 2001) and the chemical Langevin equation (Gillespie, 2000).

Both differential equation and Gillespie algorithm models require a large amount of parametric information, which is often unavailable, and have low-scalability to larger network sizes. To study the evolution of gene expression noise using evolutionary simulations, in which millions of individual networks are to be realized over thousands of generations, a gene regulatory network model that is computationally fast and represents the stochastic behaviour of real networks is needed.

1.2.2 Wagner's gene regulatory network model

One of the commonly used gene regulatory network models is the Wagner gene network model, first published in 1996 (Wagner, 1996). It was used to study evolutionary plasticity, and many alterations to the model have been made since, but the original model is outlined here.

The gene network of N genes is represented by a dynamical system whose state variables $\vec{S}(t)$ denote the expression states of the genes in the network. The expression levels of each gene are updated at each time step during the developmental process and can have only two values, +1 and -1. The relationships of all genes in the network are defined by a $N \times N$ size regulatory matrix $W = \left(w_{ij}\right)_{1 \le i \le n,\ 1 \le j \le n}$. The values of the entries of the regulatory matrix w_{ij} define the strength of the regulatory interaction between gene j and gene i. Diagonal elements can have non-zero values, defining autoregulation of the gene by its own product.

The network development process is modelled as a succession of gene expression states,

and given by a set of differential equations:

$$S_i(t+\tau) = \sigma[\sum_{j-1}^{N} w_{ij} \cdot S_j(t)] = \sigma[\, h_i(t)\,] \tag{1.1}$$

The expression level state of gene i at time $t+\tau$, $S_i(t+\tau)$ is given by the function $h_i(t)$, which is a weighted sum of the expression states of all genes in the network at time t and represents the sum of the effects all regulatory genes have on gene i.

In the original model from 1996, a filter step function was applied on gene expression states to discretize them:

$$\sigma(x) = \begin{cases} -1, & \text{if } x < 0 \\ 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \end{cases} \tag{1.2}$$

In later modifications to the model, a sigmoidal function was used to make expression level states a continuous distribution.

Evolutionary simulations were performed as cycles of reproduction, mutation and selection steps. A genotype is considered viable if its gene expression level pattern can reach a stable pattern. Fitness was determined based on expression level stability: stable expression patterns had a fitness of 1, unstable expression patterns had a fitness of 0.

Wagner's gene regulatory network model has been used for studying the evolution of gene networks, epistasis, robustness, canalization of development (Bergman and Siegal, 2003; Azevedo et al., 2006; Huerta-Sanchez and Durrett, 2007; Espinosa-Soto and Wagner, 2010; Rhoné et al., 2011; Odorico et al., 2018; Espinosa-Soto, 2018), neutral networks (Ciliberti et al., 2007), as well as evolvability of gene networks with noisy gene expression level states (Pinho et al., 2015).

1.3 Why should a biologist care about expression noise?

Understanding phenotypic variation is one of the central questions in biology. Traditionally, studies of phenotypic variation are framed as a question of the contribution of genetic and environmental factors to the focal phenotype, colloquially known as the nature vs. nurture debate. This framing imposes a deterministic view of the genotype-phenotype map, *i.e.* it assumes that the phenotype is fully determined by the genetic information and environmental conditions. In many cases, a large portion of the phenotypic variation cannot be ascribed to either genetic variation or environmental effects. For example, 35% to 40% of the behavioral individuality in orientation towards a visual object in fruit flies can be ascribed to stochastic variation in the brain wiring, which originates from non-heritable noise during brain development (Linneweber et al., 2020).

Expression noise was shown to be a selectable trait that affects fitness and the numerous evidence was outlined in section 1.1.2. Furthermore, the nongenetic variation created in a population can favor the fixation of beneficial mutations (Schmutzer and Wagner, 2020), increasing the speed of evolution. Expression noise is also a mechanism of creating phenotypic noise which leads to division of labor among cooperating individuals. In some cases, individual cells in the population independently and randomly specialize into distinct roles by increasing stochastic fluctuations in biochemical reactions of each cell (Liu et al., 2021). For example,

it was shown that the production and secretion of extracellular proteases in a population of *Bacillus subtilis* are determined randomly (Veening et al., 2008). The extracellular proteases released by the altruistic cells degrade complex proteins in the medium into freely dispersed smaller peptides, which can be taken up by any cell in the population, indicating cooperative or altruistic behaviour. Furthermore, the phenotypic noise responsible for the random activation of the production pathway was demonstrated to be a result of increased expression noise of the transcriptional regulator controlling the pathway.

1.4 Thesis scope

The objectives of this thesis are:

i) Introduce a model of gene regulatory networks which includes stochastic gene expression and allows distinction of extrinsic and intrinsic gene expression noise. The model must also represent noise propagation between connected genes in the network and be computationally feasible to implement within a forward-in-time simulation procedure.

ii) Investigate how gene expression noise responds to stabilizing selection on gene expression levels in gene regulatory networks through in silico evolutionary experiments.

iii) Investigate how gene expression mean and noise responds to directional and fluctuating selection on gene expression levels in gene regulatory networks through in silico evolutionary experiments.

Chapter 2

Evolution of gene-specific expression noise under stabilizing selection

> *No man is an island, entire of itself; Each is a piece of the continent, a part of the main; if a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thine own, or of thine friend's were; Each man's death diminishes me, for I am involved in mankind; Therefore, send not to know for whom the bell tolls, it tolls for thee.*
>
> – John Donne, *Devotions Upon Emergent Occasions* (1624)

2.1 Abstract

Expression noise, the variability of the amount of gene product among isogenic cells grown in identical conditions, originates from the inherent stochasticity of diffusion and binding of the molecular players involved in transcription and translation. It has been shown that expression noise is an evolvable trait and that central genes exhibit less noise than peripheral genes in gene networks. A possible explanation for this pattern is increased selective pressure on central genes since they propagate their noise to downstream targets, leading to noise amplification. To test this hypothesis, we developed a new gene regulatory network model with inheritable stochastic gene expression and simulated the evolution of gene-specific expression noise under constraint at the network level. Stabilizing selection was imposed on the expression level of all genes in the network and rounds of mutation, selection, replication and recombination were performed. We observed that local network features affect both the probability to respond to selection, and the strength of the selective pressure acting on individual genes. In particular, the reduction of gene-specific expression noise as a response to stabilizing selection on the gene expression level is higher in genes with higher centrality metrics. Furthermore, global topological structures such as network diameter, centralization and average degree affect the average expression variance and average selective pressure acting on constituent genes. Our results demonstrate that selection at the network level leads to differential selective pressure at the gene level, and local and global network characteristics are an essential component of gene-specific expression noise evolution.

2.2 Introduction

Living beings are complex systems constituted of many genes that interact with each other and the environment to create an organism. From prokaryotes with a few hundred essential genes, to eukaryotes with possibly several thousands, cells require many gene products to work together to perform housekeeping functions and to replicate. Fine-tuned molecular processes, generally referred to as *gene expression*, ensure how, where and when these products are generated. However, gene expression is an inherently noisy process (Elowitz, 2002; Raser and O'Shea, 2005), which involves many steps where molecules participating in the expression machinery diffuse and bind to target molecules. Additionally, these molecules are often present in small copy numbers, increasing the susceptibility of gene expression to stochastic events. Consequently, there is a variation in gene expression levels among cells, even if they are isogenic and grown in a homogeneous environment, and this inevitable variation has been termed *gene expression noise*. Organisms have to express hundreds of genes, each one of which is noisy – raising the question of how they evolved to cope with this inevitable noise.

The expression noise level of a particular gene may be decomposed into two components, called *extrinsic* and *intrinsic*. Extrinsic noise affects all genes equally and results from the sharing of key molecules, such as RNA polymerases and ribosomes, by all genes in the expression process, as well as, for instance, differences in cell size and phase in the cell cycle. Intrinsic noise is gene-specific and results from different chromatin states, cis-regulatory elements and kinetic parameters of transcription and translation of each gene (Chalancon et al., 2012). Minor sequence mutations can have a significant effect on the level of expression noise. For example, a small number of single-nucleotide changes in a transcription factor binding site were reported to have a large effect on the expression noise level (Sharon et al., 2014). Since (i) there is variation in the level of intrinsic noise of genes, and (ii) intrinsic noise is genetically determined – and, therefore, heritable – gene expression noise can be shaped by natural selection.

Evidence of selection on expression noise was first seen in the fact that dosage-sensitive genes (Lehner, 2008) and essential genes exhibit lower levels of expression noise (Fraser et al., 2004; Wang and Zhang, 2011). Intrinsic noise was also reported to correlate with the strength of selection acting on the encoded protein. Namely, proteins with a lower ratio of non-synonymous over synonymous substitution rate (Ka/Ks) have a lower level of expression noise (Barroso et al., 2018). Changes in the expression noise of a single gene may be either beneficial or deleterious, depending on how far its mean expression is from the optimal expression level (Duveau et al., 2018). Expression noise is deleterious if the mean expression level is close to the optimal, as higher variation, in this case, generates a larger number of less fit individuals, reducing the population fitness. Conversely, expression noise can be beneficial if the mean expression level is far from the optimum, as noisy genes are more likely to generate cells with an expression level closer to the optimum. Noisy gene expression can thus be part of a bet-hedging strategy and was observed in genes involved in immune and environmental response (Beaumont et al., 2009; Bódi et al., 2017; Farquhar et al., 2019; Nevozhay et al., 2012). The fitness cost of changes in the level of expression noise in the fitness landscapes of ≈ 30 yeast genes have been shown to be on the same order as fitness costs of changes in mean expression level (Schmiedel et al., 2019). Since the fitness effect of different levels of expression noise can be as detrimental as different mean expression levels, which are thought to be extensively under selection (Gilad et al., 2006), it can be assumed that expression noise is extensively un-

der selection genome-wide. Prevalent selection on expression noise has been demonstrated in naturally segregating promoter variants of *E. coli* (Vlková and Silander, 2022).

The phenotype (and, therefore, the fitness) of an organism depends on the interaction of many genes. As a result, genes do not evolve independently, and the selective pressure acting on a gene's intrinsic noise depends on its interactions with other genes. Understanding the evolution of gene expression noise requires accounting for such gene-to-gene interactions, commonly depicted by a gene network. The propagation of noise from gene to gene in the network was established both theoretically and experimentally (Pedraza, 2005; Blake et al., 2003). Genes with many connections propagate their noise to a more substantial extent than genes with fewer connections and, therefore, contribute more to the global noise levels of the network. Gene networks are robust to variation in the expression level of their system components to some degree, but at a critical point the global noise of the network becomes too high and leads to network collapse. Selection against noise at the network level was, therefore, hypothesized to result in stronger constraints on the intrinsic noise of highly connected genes (Barroso et al., 2018). Moreover, the topological structure of the network has been shown to affect the pattern of noise propagation (Hens et al., 2019), suggesting that the topology of the network might impose additional selective constraints on the constituent genes.

Here, we test the hypothesis that expression noise of highly connected genes in gene networks is under stronger selective pressure than expression noise in peripheral genes using an *in silico* evolutionary experiment. We introduce a new gene regulatory network evolution model, which includes an evolvable component of stochastic gene expression, and use it to evolve thousands of network topology samples over 10,000 generations. These simulations showed that highly connected genes have a more constrained intrinsic expression noise. They further revealed that not all genes might evolve in response to network-level selection, and the probability that they do so depends on local network properties. Lastly, the average selective pressure acting on genes in a network is affected by topological features such as network diameter, centralization and average degree.

2.3 Materials and methods

We introduce a new gene regulatory network model that incorporates intrinsic expression noise. We then use this model within a forward simulation framework to simulate the evolution of populations of networks with mutable levels of intrinsic expression noise. These simulations allow us to study how the selective pressure acting on expression noise varies within the regulatory network.

2.3.1 A gene regulatory network model with stochastic gene expression

To investigate the evolution of stochastic gene expression in gene regulatory networks, we first extend Wagner's gene network model (Wagner, 1996) to integrate gene-specific expression noise.

We model a network of n genes ($n = 40$ in this study) defined by a regulatory matrix $W = \left(w_{ij}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$, and a vector of intrinsic, gene-specific noise $\{\eta_i^{\text{int}}\}_{1\leq i\leq n}$. Each element w_{ij} of the regulatory matrix W defines the regulatory effect of gene j on gene i. The value of w_{ij}

is a real number and is referred to as regulatory strength of gene j on gene i. In case $w_{ij} > 0$, gene j is an activator of gene i and increases its expression level. Conversely, when $w_{ij} < 0$, gene j is a repressor of gene i and decreases its expression level. Lastly, if $w_{ij} = 0$, gene i is not regulated by gene j and gene j has no effect on expression level of gene i. Two genes i and j are connected by an edge in the network if at least one of w_{ij} and w_{ji} is non-null. The intrinsic noise vector $\{\eta_i^{\text{int}}\}_{1\leq i\leq n}$ defines the gene-specific expression noise of each gene in the network. The regulatory matrix and the intrinsic noise vector together constitute a unique genotype in this modeling framework (Fig 2.1A).

The phenotype (the expression level of each gene) in the model is represented by a state vector $\{S_i\}_{1\leq i\leq n} = \{s_1, s_2, \ldots, s_n\}$, which describes the expression level of each gene. The state vector at t_0 is set to an arbitrary basal expression level value ($\{S_i^0\}_{1\leq i\leq n} = \{S_i^{basal}\}_{1\leq i\leq n} = \{20, ..., 20\}$ in this study). In every time step t ($1 \leq t \leq T_r$, with $T_r = 50$ in this study), the expression level of each gene is recomputed. The cumulative effect of all transcription factors in the expression level of each gene is for simplicity considered to be additive, *i.e.* we assume there is no cooperative or competitive binding of transcription factors to transcription factor binding sites. This assumption removes the small degree of non-linearity in the response of the regulated gene to transcription factor concentrations, which is present in real transcription factor regulation dynamics. The activation rate $a_i(t)$ is defined as the sum of all effects the regulators of gene i have on its expression level at time step t:

$$a_i(t) = \sum_{j=1}^{n} w_{ij} \cdot s_j(t), \tag{2.1}$$

in which case the dynamic equation for the expression level of each gene in the following time step is:

$$s_i(t+1) \sim \mathcal{N}\left(s_i^{basal} + a_i(t), \eta_i^{int}\right). \tag{2.2}$$

In every time step the expression level of a gene is drawn from a random distribution. We implemented a simple Gaussian noise, where the mean of the normal distribution equals the sum of basal expression level (s_i^{basal}) and activation rate ($a_i(t)$), and the variance equals the gene noise genotype (η_i^{int}). If the expression level value drawn from the normal distribution is below the minimal ($s_{min} = 0$) or above the maximal expression level ($s_{max} = 100$), it is set to the minimal or maximal expression level, respectively. We note that the shape and variance of the distribution is constant in realization time in our model, but that the expression levels of each individual is the product of the trajectory of the expression levels during the realization process, during which expression levels can exhibit phenotypic switching between stable states. Consequently, there can be a non-normal expression level distribution of a certain gene in the clonal population, even though the expression levels in each time step are drawn from a normal distribution.

The expression levels of all genes are synchronously updated in each time step. The steady state expression levels are invariant to whether the expression levels of each gene are updated synchronously or asynchronously (Appendix A.1.2). Similarly, mean expression level, expression variance, CV, noise and Fano factor are invariant to the updating mode (Appendix A.1.2). The model may be realized as stochastic or deterministic, depending on the noise parameter values (Fig 2.1B). The deterministic realization has been used to benchmark the model and to set up the mean expression levels for the starting populations, and the stochastic realization has been used in the main bulk of the simulations, in which intrinsic noise is evolved.

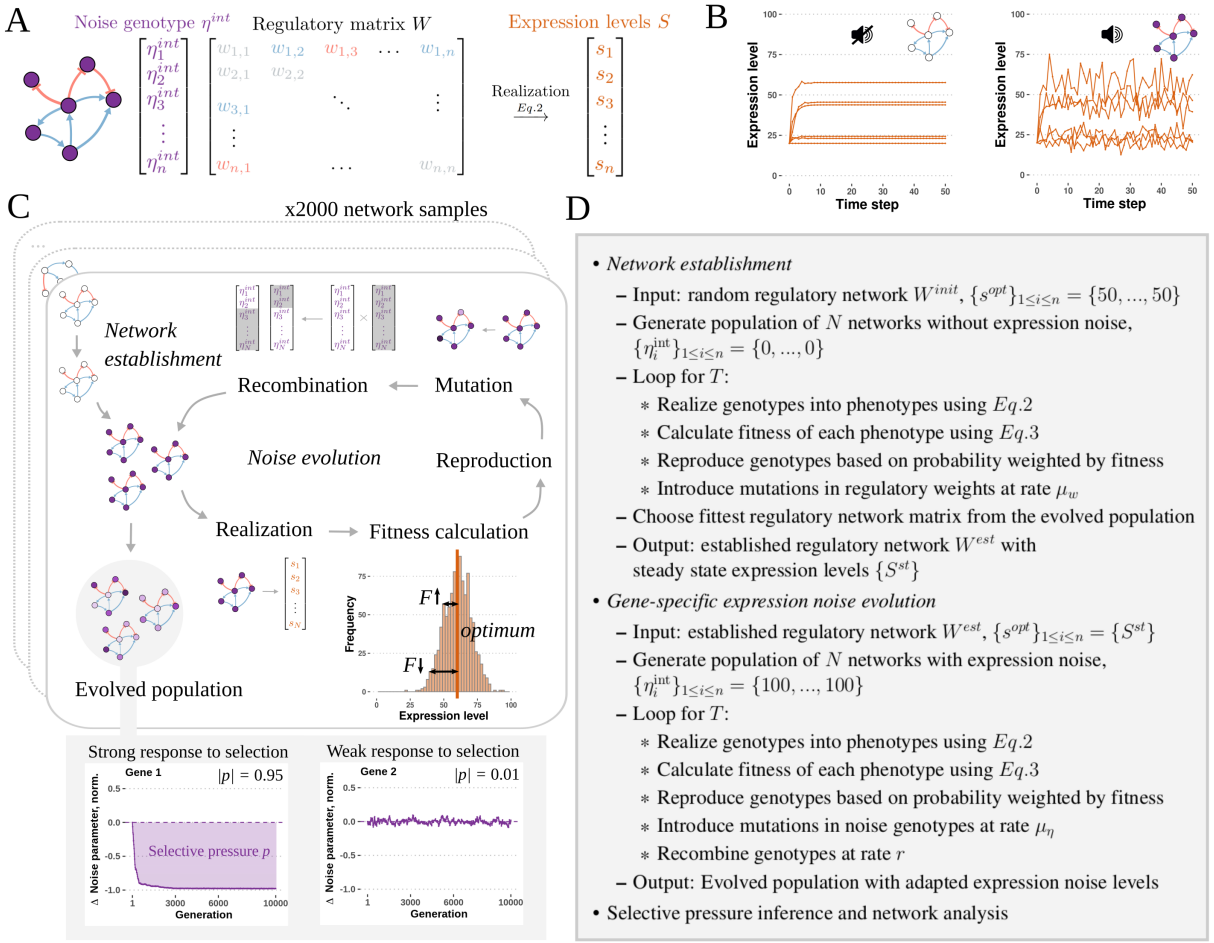

Figure 2.1: **The evolution of gene-specific expression noise was simulated using populations of model gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions. A** - Gene regulatory network model. The genotype consists of the intrinsic noise vector η^{int} and regulatory matrix W. The intrinsic noise vector defines the gene-specific expression variance of each gene in the network. The regulatory matrix defines the regulatory interactions in the network. The genotype is realized into the phenotype using the dynamical equation described in the main text. The phenotype is given by the state vector S, which represents the expression level of each gene in the network. **B** - Deterministic (left) and stochastic (right) realizations of the model. **C** - Steps of the evolutionary simulation process. Each established network configuration was used as a founding network for the network populations used in the noise evolution simulation. In every generation, genotypes are realized and phenotypes (expression levels) are sampled from the last time step. Fitness is calculated from the expression levels. If the populations are evolved under selection, fitness is calculated as the distance of the expression level of each gene from the optimal expression level. Genotypes are reproduced based on their relative fitness and mutations in the intrinsic noise vectors are introduced. Noise genotype vectors are recombined by randomly choosing individuals for recombination and shuffling their noise vectors. The process is repeated for 10,000 generations. **D** - Algorithm overview.

2.3.2 Forward-in-time simulation of expression noise evolution

To investigate how gene-specific expression noise of constituent genes responds to stabilizing selection at the network level, we used the newly introduced model to perform forward-in-time evolutionary simulations in which we allow the gene-specific noise levels to mutate. An *in silico* evolutionary process consisting of rounds of mutation, selection, recombination and replication events of a population of N ($N = 1,000$ in this study) individuals was performed for T ($T = 10,000$) generations (Fig 2.1C).

We first generated network topologies that would serve as the founding network for the populations in our simulations. We generated 2,000 random (Erdős–Rényi model) network topologies of 40 nodes with regulatory strength values drawn from a uniform distribution $\mathcal{U}(-3,3)$. The network density was $d = 0.05$. Only connected network graphs were used, meaning there is only one component and there are no disconnected subgraphs. Autoregulation is not present, because it affects gene-specific noise levels and would be a confounding factor in the analysis. In order to assess the effect of the topology structure on the evolution of expression noise, we also generated an additional 1,000 scale-free (Barabási–Albert model) and 1,000 small-world (Watts–Strogatz model) network topologies with the same size and density. Both random and small-world networks are characterized by a Poisson degree distribution and short mean shortest path length, but random networks have a low clustering coefficient, while small-world networks have a high clustering coefficient. Scale-free networks are characterized by a degree distribution that follows a power law. Real-world networks exhibit degree distributions similar to power-law distributions, high clustering and short path lengths. As such, real-world networks have features of both scale-free and small-world networks (Newman, 2010).

In the simulation of expression noise evolution the regulatory interactions were immutable and the values of the noise genotype vectors were allowed to mutate. Stabilizing selection, the selection scenario in which individuals with extreme phenotypic values have a lower fitness, was imposed on all constituent genes by setting the value of optimal expression level as the mean equilibrium expression level of each gene. The fitness $F(s)$ of a phenotype s was calculated as in Laarits et al. (Laarits et al., 2016), where fitness is defined as the distance from the optimal expression state vector $\{s_i^{opt}\}_{1\leq i\leq n}$, weighted by the fitness contribution given by $\{\rho_i\}_{1\leq i\leq n}$:

$$F(s) = e^{-\sum_{i=1}^{n} |s_i^{opt} - s_i|/(n\rho_i)} \tag{2.3}$$

The fitness contribution parameters $\{\rho_i\}_{1\leq i\leq n}$ define the contribution of each gene to the fitness of the phenotype, *i.e.* it is a scaling factor of the decrease of fitness as a function of the distance of the expression level from the optimal expression level for each gene. In this study, the strength of the imposed selective pressure is set to be identical for all constituent genes ($\forall i\, \rho_i = 1$). The assumption of all genes having identical fitness contribution is biologically unrealistic, so we have also performed simulations in which we impose unequal fitness contributions among genes in the same network. We found consistent conclusions (Appendix A.5), and, for simplicity, we report the results with equal fitness contributions here. Since the fitness contribution of all genes is identical, any differences in the evolutionary outcome we observe after removing the effect of drift will be due to gene differences in their network interactions. Individuals were reproduced into the next generation with a probability equal to their relative phenotype fitness. The fitness of all phenotypes in populations evolved un-

der non-selective conditions was set to an equal constant value, regardless of gene expression levels. Mutations were introduced at a rate μ_η ($\mu_\eta = 0.01$) per gene per replication event. The values for noise genotype mutations were drawn from a normal distribution $\mathcal{N}(100, 40)$. There is no experimental evidence for the shape of the distribution of the expression noise and regulatory strength mutations. We chose a normal distribution because: 1) it defines equally frequent beneficial and deleterious mutations and 2) most mutations would have a small effect, which reflects the characteristic of many studied distributions of fitness effects in model organisms. Recombination was implemented by choosing a random offspring individual at a rate r ($r = 0.05$) and introducing a random break point in the linear genome. The genotype values in the genome segment defined by the break point were then exchanged with another randomly chosen individual from the offspring population. A constant population size N ($N = 1,000$) was maintained. To account for the effect of genetic drift, the noise evolution simulations of each founding network population were replicated 10 times under selection and 10 times under neutrality.

We found that the expression level of most genes in networks with random configurations converge to either s_{min} or s_{max} under a deterministic realization. The measurement of variance of genes that are either not expressed at all or expressed at the maximal level would be impaired since their expression range is constrained by the lower and upper expression level boundary. Since the study of expression variance is our main focus, we added a network establishment step before the noise evolution simulations, in which we subject the network regulatory matrix to mutation and selection for intermediate expression levels. During the network establishment step networks are realized deterministically, *i.e.* the intrinsic noise genotype of all genes is 0. Networks with intermediate steady state expression levels were established through the evolutionary process by imposing a target expression level $\{s_i^{opt}\}_{1\le i\le n}$ ($\{s_i^{opt}\}_{1\le i\le n} = \{\frac{s_{max}}{2}, ..., \frac{s_{max}}{2}\}$) for all genes and allowing the strength of regulatory interactions to mutate. Mutations were introduced at a rate μ_w ($\mu_w = 0.1$) in non-zero entries in the regulatory matrix, preserving the network topology structure (Erdős–Rényi, Barabási–Albert, or Watts–Strogatz model). The values for regulatory strength mutations were drawn from a normal distribution $\mathcal{N}(0, 2)$. Recombination was not implemented at this stage. Fitness of each individual was computed as the distance of the phenotype to the optimal expression state vector using *Eq.*1. Individuals were reproduced with a probability equal to the relative fitness and the population size kept constant. Network regulatory configurations in which the expression level of all genes would not converge to a fixed point and would oscillate were discarded, as in previous studies (Laarits et al., 2016). Oscillating gene expression level patterns create population-level heterogeneity generated by the system oscillations and not by stochastic gene expression. Since we are studying the evolution of gene-specific expression noise, expression noise generated by oscillations would be a confounding factor in our analysis. We note, however, that oscillatory networks can be frequent in simulations (Pinho et al., 2012) and biological systems (Zhang et al., 2014), and the role of expression noise in their behavior is an interesting perspective for follow-up studies. Expression level dynamics were termed oscillating if the sum of the differences between expression level in the last time step and previous τ time steps ($\tau = 10$) was higher than ϵ ($\epsilon = 10^{-6}$). A stable, *i.e.* non-oscillating, expression level dynamics satisfied the following criterion (Laarits et al., 2016):

$$\Phi(S(t)) = \frac{1}{\tau}\sum_{\theta=t-\tau}^{t} D(S(\theta), S(t)) < \epsilon \tag{2.4}$$

where D is the distance between two vectors $D(S^1, S^2) = \sum_{i=1}^{n} |S_i^1 - S_i^2|/n$.

The network establishment process consisting of rounds of mutation, selection and reproduction of a population of N ($N = 1,000$) individuals was performed for T ($T = 10,000$) generations, for each network topology. At the end of the network establishment process, 68% (54333/80000) of genes had intermediate expression levels (Appendix A.1.3). The reason why a minority of the genes do not reach close to optimum expression levels could be potential network configuration constraints or a non-extensive optimization/fitting algorithm. Genes that had an expression level of 0 or s_{max} were filtered out from the dataset used in the final analysis. The network regulatory configuration with the highest fitness was chosen from the evolved population and this network configuration was used to generate the starting population for the noise evolution simulations.

The gene network model and evolutionary simulations were implemented in C++ and the source code is available at https://gitlab.gwdg.de/molsysevol/supplementarydata_expressionnoise/cpp.

2.3.3 Analysis of simulation results: expression noise and network centrality measures

The evolutionary outcomes (*i.e.* the change of phenotypes and genotypes) were measured as change of expression noise and selective pressure for each network, respectively. Expression noise in the first and last generation in each evolved population was measured as the variance of the population expression level states for each gene. The change of expression noise (phenotypic evolution) between the first and last generation was measured as the relative change of expression noise, calculated as the difference of expression variance between the first and last generation divided by their sum $(\sigma^2_{gen1} - \sigma^2_{gen10k})/(\sigma^2_{gen1} + \sigma^2_{gen10k})$.

The selective pressure (genotypic evolution) acting on each gene was measured as the average change of noise genotype in every second generation relative to the starting level (Fig 2.1C). To compare the effect of node centrality on the selective pressure acting on constituent genes, we computed node-level network centrality measures for each node in the networks. We focused our analysis on two local network centrality measures, node instrength and outstrength, but over 30 network centrality measures were analyzed (Appendix A.2). In-strength of node i is measure of the strength and number of in-going links, *i.e.* how strongly a gene is being regulated:

$$\text{Instrength}(i) = \sum_{j}^{n} |w_{ij}|. \tag{2.5}$$

Conversely, the outstrength of node j is a measure of the strength and number of outgoing links, *i.e.* how strongly a gene regulates other genes downstream:

$$\text{Outstrength}(j) = \sum_{i}^{n} |w_{ij}|. \tag{2.6}$$

Further, we computed global graph-level metrics, such as mean graph distance and performed a principal component analysis to reduce the dimensionality (Appendix A.2). The results were analysed in `R 3.6.3` (Team, 2021). Network analyses were performed using the `igraph 1.2.4.2` Csardi and Nepusz (2006) and `statnet 2019.6` (Hunter et al., 2008) packages.

Principal component analysis was performed using the `ade4 1.7.15` (Dray and Dufour, 2007) package.

2.3.4 Analysis of simulation results: linear modeling

We fitted linear mixed-effects models using network centrality measures as fixed effect variables and the network topology sample as a random effect variable, allowing for control of intra-network correlation in the response variable. We tested different transformations of the response and explanatory variables in order to improve linearity, and variance structures to account for heteroskedasticity of the residuals. A model where the residual variance was an exponential function of the node absolute instrength was shown to provide the best fit according to the minimal Akaike's Information criterion and was used for all subsequent models (Appendix A.3). Two types of models were fitted: a logistic regression where the response variable was set to whether a gene answered to selection or not, and standard regressions that used expression variance, relative change of expression variance or selective pressure as response variables. Linear mixed-effect modelling was performed using the `nlme 3.1.144` (Pinheiro et al., 2022) and `lme4 1.1.27.1` (Bates et al., 2015) packages. Marginal and conditional R^2 values were computed using the `MuMIn 1.43.17` (Bartoń, 2020) package. Network centrality measures used as explanatory variables in our linear models were correlated (Pearson's $r = -0.17$, p-value $< 2.2 \times 10^{-16}$, Appendix A.2), so we computed the variance inflation factor (VIF) using the `car 3.0.11` (Fox and Weisberg, 2019) package. The VIF of all linear models was less than 3; therefore, colinearity was considered to have negligible impact on the inferred statistical significance (James et al., 2013). To improve homoskedasticity of the residuals in the linear models, we also performed each model fit on two filtered datasets: one in which genes with zero values of instrength or outstrength were removed, and one in which only genes with zero values of instrength or outstrength were kept. The same pattern of effects and significance is observed in the filtered as in the main dataset, so we included the results of the complete dataset in the main text and reported the results of the reduced dataset in Appendix A.6.

Finally, since in some cases variable transformation, heterogeneous variance modeling and data filtering did not ensure normality and independence of the residuals, we assessed the amount of resulting bias in the estimation of p-values using a randomization test, in which we fitted a selected model on 10,000 permuted datasets. We chose the model of relative noise change (Appendix A.3), as the corresponding residuals were significantly departing normality (Shapiro-Wilk test, p-value $< 2.2 \times 10^{-16}$) and independence (Box-Ljung test, p-value $= 8.9 \times 10^{-7}$). For each permutation, we shuffled the values of the response variable (relative change of variance) within each network topology, which removes the effect of network metrics on the change of noise, but preserves the distributions of each metric per network, as well as putative colinearity between explanatory variables. Using $\alpha = 0.05$ as a significance cutoff value, we found a false discovery rate (FDR) of 6.0% for the effect of instrength and and 6.7% for the effect of outstrength. While these values are above the expected 5%, the FDR inflation was found to be relatively low and we concluded that the non-normality of residuals did not affect our conclusions.

2.3.5 Analysis of simulation results: information-based metrics

Generalized linear mixed-effects models make several assumptions that might be violated by the data in some cases. Namely, they assume a normal distribution and homoskedasticity of Pearson's residuals, and a normal distribution of random effects. To further validate our conclusions, we computed the mutual information (MI) between variables, which does not have any prior assumptions. We calculated mutual information between the expression noise and centrality metrics using the `infotheo 1.2.0` (Meyer, 2014) package. Monte Carlo permutation tests with 10,000 permutations were used to compute p-values for the significance of the mutual information between each pair of tested variables.

2.4 Results

We investigate how selection at the gene network level may lead to the evolution of differential gene-specific expression noise, as observed in biological systems. To do so, we introduce a new gene regulatory model with stochastic gene expression, which extends Wagner's model (Wagner, 1996) by adding node-specific intrinsic noise parameters (Fig 2.1A-B). In this framework, the phenotype is represented by the expression level of each gene, and is the realization of a random distribution determined by the genotype. The fitness of an individual is further determined by its distance to an optimal phenotype, therefore, stabilizing selection is implemented as acting on the expression level. We used this model to simulate the evolution of populations of gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions (Fig 2.1C-D), and assessed how node properties affect the evolution of intrinsic noise.

2.4.1 Expression noise propagates along the regulatory network

We first investigated how noise propagated in the model gene regulatory networks. It was shown that noise is additive in biological networks and, therefore, propagates from regulators to regulated genes (Pedraza, 2005; Blake et al., 2003). To assess whether our model successfully captured this property, we generated a dataset of 2,000 realized random network topologies, and tested whether gene expression variance increased with the number of ingoing regulatory links. As expected, we found that the absolute instrength of a gene had a significant positive effect on gene expression variance (linear mixed-effects model with coefficient β = 0.28, p-value < 2.2×10^{-16}) (Fig 2.2A), indicating that noise propagation was captured in our model. Furthermore, the mutual information between gene expression variance and absolute instrength was significant (MI = 0.67, p-value ≤ 10^{-4}, permutation test). High node instrength increases expression noise, in line with the experimental evidence that the noisiness of promoters increases with the number of regulatory inputs (Urchueguía et al., 2021).

We then looked at fitness costs associated with high expression noise in regulators and regulated genes. In a dataset of 1,000 random network topologies, we assessed the mean fitness of the clonal populations of 1,000 individuals under stabilizing selection on the expression level. Each gene was imposed 5 different levels of intrinsic noise, while the intrinsic noise of the rest of the network was kept at 0. We found that increasing the level of expression noise of a single gene decreased the mean fitness of the network (linear mixed-effects model with coefficient β = -0.002, p-value < 2.2×10^{-16}), as expected. However, the strength of this

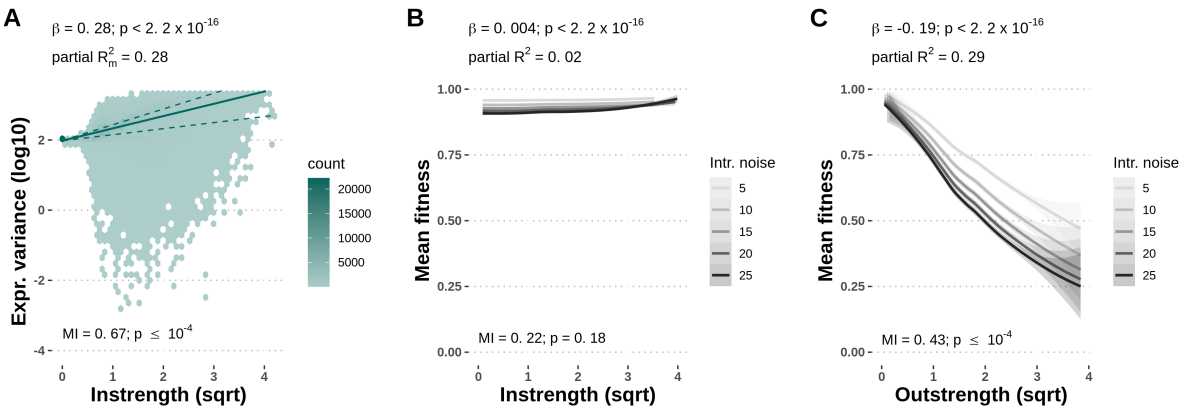

Figure 2.2: **Noise propagation is captured by the gene regulatory network model. A** - Gene-specific expression variance increases with the absolute instrength of the node, indicating noise propagation is reflected in the gene regulatory network model. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **B, C** - Gene-specific expression variance decreases fitness in gene networks under stabilizing selection on gene expression level. Increasing the level of gene-specific expression noise reduces the mean fitness of the clonal population. The mean fitness of the population is significantly, but marginally, increased by noise in genes with higher node instrength (B), and significantly decreased by noise in genes with higher node outstrength (C). Lines represent the smoothed conditional means and grey bands represent the 95% confidence interval bands. Coefficients, p-values and partial marginal R^2 measures are estimated using linear mixed-effects models with expression variance or mean fitness as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed with a permutation test with 10,000 permutations.

effect depended on the gene centrality. The reduction of fitness due to gene-specific expression noise was significantly, but marginally, affected by instrength (linear model with coefficient β = 0.004, p-value < 2.2×10^{-16}, Fig 2.2B). The mutual information between mean fitness of the population and absolute instrength was not significant (MI = 0.22, p-value = 0.18, permutation test). However, the mean fitness significantly decreased with node outstrength (linear model with coefficient β = -0.19, p-value < 2.2×10^{-16}, Fig 2.2C). The mutual information between mean fitness of the population and absolute outstrength was significant (MI = 0.43, p-value ≤ 10^{-4}, permutation test). Higher fitness cost of expression noise in gene with high outstrength suggests there is a differential selective pressure acting on genes based on their centrality in the gene regulatory network, which we explore in the next section using an *in silico* evolutionary experiment.

2.4.2 Gene expression noise is reduced under a stabilizing selection regime

To investigate how gene-specific expression noise responds to stabilizing selection at the network-level, we simulated the evolution of 2,000 random network topologies with and without selection on the gene expression level. We observed that gene expression variance de-

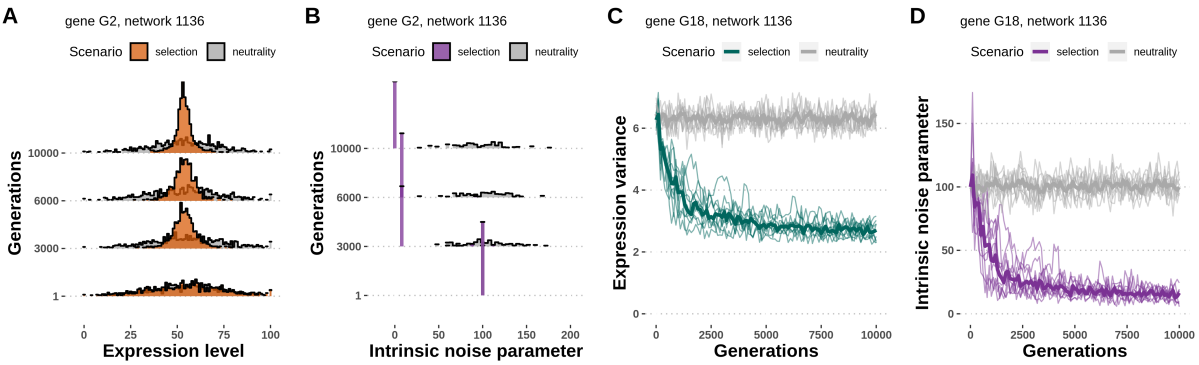

Figure 2.3: **Gene-specific expression noise evolves in a model with selection. A** - The distribution of expression levels of an example gene throughout evolution in populations evolved under stabilizing selection on gene expression level and under neutrality. The variance of gene expression level is reduced under selection, but not under neutrality. **B** - The distribution of intrinsic noise parameters of an example gene throughout evolution in populations evolved under selection and under neutrality. The median intrinsic noise parameter skews to lower values under stabilizing selection, but not under neutrality. **C, D** - Replicates of the simulations with the same input network and parameters. Replicates have different dynamics, but reach similar outcomes in terms of expression variance (C) and median intrinsic noise parameter (D) in the evolved populations. The evolution of each network topology sample was replicated 10 times under selection and 10 times under neutrality.

creased throughout evolution under selective conditions (Fig 2.3A), and the distribution of intrinsic noise parameters in the population shifted towards lower noise genotype values (Fig 2.3B), indicating that low-noise alleles conferred a fitness increase to the network. Conversely, gene expression variance remained constant throughout evolution under neutral conditions, and the distribution of noise genotypes reflected only the distribution of random mutations. Replicating the simulations for each network topology sample yielded similar reduction of gene expression variance (Fig 2.3C) and median noise parameter in the population (Fig 2.3D). As the initial networks were at their optimal expression level, the mean expression level did not change during evolution and was highly correlated between the first and last generations (Pearson's r = 0.99, p-value < 2.2×10^{-16}, Appendix A.1.5), confirming that selection acted only on the gene expression variance. Population size had a positive effect on the selective pressure acting on genes, as expected, selection being more efficient in large populations (Appendix A.1.4). A population size of 1,000 individuals was chosen for the main simulations as the optimal population size in the trade-off between selecting mutations with small effects and reducing computational speed.

Next, we investigated how individual nodes within a network respond to selection, based on their centrality properties.

2.4.3 Evolutionary change in phenotypes: regulators reduce their expression noise to a higher degree

We first analysed the phenotype change, *i.e.* the relative change in gene-specific expression variance after evolution. The variance of gene expression depends both on the intrinsic noise

of the genes (that is, its genotype in our model) and the number and noise of the genes it is connected with.

We fitted linear models to assess the impact of the absolute instrength and outstrength measures on the relative change in expression variance for each node in each network. Under selection, both absolute instrength and absolute outstrength had a significant negative effect (linear mixed-effects model with coefficients $\beta_{\text{instrength}}$ = -0.003, p-value = 2.9×10^{-10}, Fig 2.4A; $\beta_{\text{outstrength}}$ = -0.046, p-value < 2.2×10^{-16}, Fig 2.4B), meaning that genes with more and stronger connections reduced their expression variance to a larger extent than less connected genes. The effect was notably stronger for outstrength (marginal R^2 = 0.15) than for instrength (marginal R^2 = 5.2×10^{-4}). Similarly, the mutual information was significant between the relative change in gene expression variance under selection and absolute instrength (MI = 0.09, p-value ≤ 10^{-4}, permutation test) and absolute outstrength (MI = 0.14, p-value ≤ 10^{-4}, permutation test). Genes with high outstrength are strong regulators and their reduction of expression variance to a larger extent indicates that high expression noise is more detrimental in regulators than in regulated genes. Under neutrality, absolute instrength had a significantly positive effect (linear mixed-effects model with coefficient $\beta = 8.3 \times 10^{-4}$, p-value < 2.2×10^{-16}, Fig 2.4C) and absolute outstrength did not have a significant effect on the relative change in gene expression variance (linear mixed-effects model with coefficient $\beta = 7.1 \times 10^{5}$, p-value = 0.26, Fig 2.4D). The mutual information was significant between the relative change in gene expression variance under neutrality and absolute instrength (MI = 0.03, p-value ≤ 10^{-4}, permutation test) and absolute outstrength (MI = 0.01, p-value ≤ 10^{-4}, permutation test). These effects are much smaller and of opposite direction than the ones measured in selective conditions, indicating that genetic drift did not cause the effect of centrality measures on expression variance observed in selected populations.

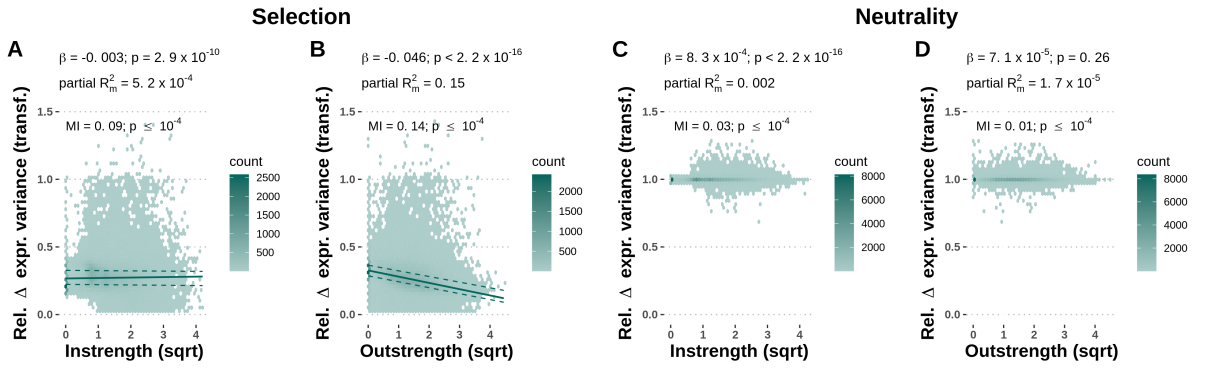

Figure 2.4: **Node-level network centrality measures affect the relative change of gene-specific expression variance under network-level selection.** For each gene, the relative change of expression variance before and after evolution (Rel. Δ expr. variance) was averaged over all replicates. **A, B** - Absolute instrength (A) and absolute outstrength (B) have a significant negative effect on the relative change in gene expression variance in populations evolved under selection. A lower value of relative change of expression variance indicates a bigger reduction in expression variance between the first and last generation and a stronger response to selection. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **C, D** - Absolute instrength (C) and absolute outstrength (D) have a significant, but negligible, negative effect on the relative change in gene expression variance in the populations evolved under neutrality. The dataset consists of 74,443 genes from 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. Coefficients, p-values and partial marginal R^2 measures were estimated using linear mixed-effects models with relative change of gene-specific variance as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed using 10,000 permutations.

2.4.4 Evolutionary change in genotypes: regulators are more likely to respond – and display a stronger response – to selection

To investigate differential selective pressure acting on gene-specific expression noise, we analysed the change of intrinsic noise parameters in populations of gene regulatory networks evolved with or without stabilizing selection on the expression level. We measured the selective pressure acting on individual genes as the average reduction in the intrinsic noise parameter relative to the beginning of the evolutionary simulation (see Methods). The selective pressure on genes was found to be close to 0 in neutrally evolving populations, as expected (Fig 2.5B). In the presence of selection, however, the distribution of selective pressures was found to be bimodal (Fig 2.5A). Therefore, we binned genes in two categories according to whether they responded to selection (selective pressure > 0.5) or not (selective pressure ≤ 0.5). We then separately analysed the probability to respond to selection and the strength of the response.

Absolute instrength had a significant and strongly negative effect (logistic regression with coefficient β = -1.87, p-value < 2.2×10^{-16}, Fig 2.5C) on the probability of a gene to respond to selection, that is, genes with more and stronger incoming links are less likely to respond to selection. Absolute outstrength also had a significant effect on the probability of a gene to respond to selection (logistic regression with coefficient β = -0.08, p-value = 6.7×10^{-7}, Fig 2.5D). However, this effect was small and was lost when the interaction terms between instrength and outstrength were included in the model (SI).

For a qualitative analysis of the effect of network centrality on the selective pressure acting on individual genes, we fitted linear-mixed effects models on the set of genes that responded to selection, with selective pressure as the response variable. In the genes that responded to selection from the selected populations, absolute instrength had a significant negative effect (linear mixed-effects model with coefficient β = -0.04, p-value < 2.2×10^{-16}, Fig 2.5E). Conversely, absolute outstrength had a significant positive effect (linear mixed-effects model with coefficient β = 0.03, p-value < 2.2×10^{-16}, Fig 2.5F) on the selective pressure. In the selected populations, the mutual information was significant between the selective pressure and absolute instrength (MI = 0.19, p-value ≤ 10^{-4}, permutation test) and absolute outstrength (MI = 0.31, p-value ≤ 10^{-4}, permutation test). In the neutral populations, neither absolute instrength nor absolute outstrength had a significant effect (linear mixed-effects model with coefficient $\beta_{\text{instrength}} = 2.4 \times 10^{-8}$, p-value = 0.99, Fig 2.5G; $\beta_{\text{outstrength}} = -1.2 \times 10^{-5}$, p-value = 0.49, Fig 2.5H) on the selective pressure. Similarly, the mutual information was not significant between the selective pressure and absolute instrength (MI = 0.005, p-value = 0.34, permutation test), nor absolute outstrength (MI = 0.005, p-value = 0.45, permutation test).

The increased selective pressure in genes with high outstrength (strong regulators) can be explained by noise propagation to downstream elements. Namely, expression noise in regulators propagates to the genes they regulate, increasing the overall expression noise in the gene regulatory network. If gene expression levels in the network are under stabilizing selection, expression noise is deleterious. Therefore, regulator genes experience a comparatively higher selective pressure to reduce expression noise than regulated genes. In a genome-wide expression noise screen in *Drosophila melanogaster*, transcription factors were found to have lower expression variation (Sigalova et al., 2020). Suppression of expression noise can be attained through negative autoregulation (Becskei and Serrano, 2000; Dublanche et al., 2006; Grönlund et al., 2013), whereby a regulator acts as its own repressor. Incidentally, 40% of transcription

factors in *E. coli* (Rosenfeld et al., 2002) and many eukaryotic transcription factors (Alon, 2007) have negative autoregulation, indicating a wide-spread control of expression noise in natural regulatory networks.

In contrast to regulator genes, we found that regulated genes, *i.e.* genes with high node instrength, are less likely to respond to selection and the selective pressure decreases with node instrength. Since the expression noise of genes is a sum of their intrinsic noise and noise propagated from upstream elements, the contribution of intrinsic noise to the total noise of the gene will be comparatively smaller in strongly regulated genes. The network can thus respond to selection either by reducing the intrinsic noise of the focal gene, or by reducing the intrinsic noise of any of the upstream elements, which would reduce propagated noise. As a result, there is a relaxation of selective pressure in regulated genes, which is distributed on upstream genes. On the other hand, the same mechanism increases the selective pressure on upstream genes, *i.e.* regulators.

To check the robustness of our results, we performed the node-level network centrality analysis on two additional datasets with different topology structures: scale-free (BarabásiAlbert) and small-world (Watts–Strogatz) topology models. We find consistent effects (direction and significance) of local network centrality metrics on the selective pressure acting on gene-specific noise across topology models, showing that our findings are robust to the topology model used (Appendix A.4). However, the effect size of network centrality metrics differed between the topology models, pointing at an effect of the topology model on noise propagation and the evolution of gene-specific expression noise, which we investigate in the next section.

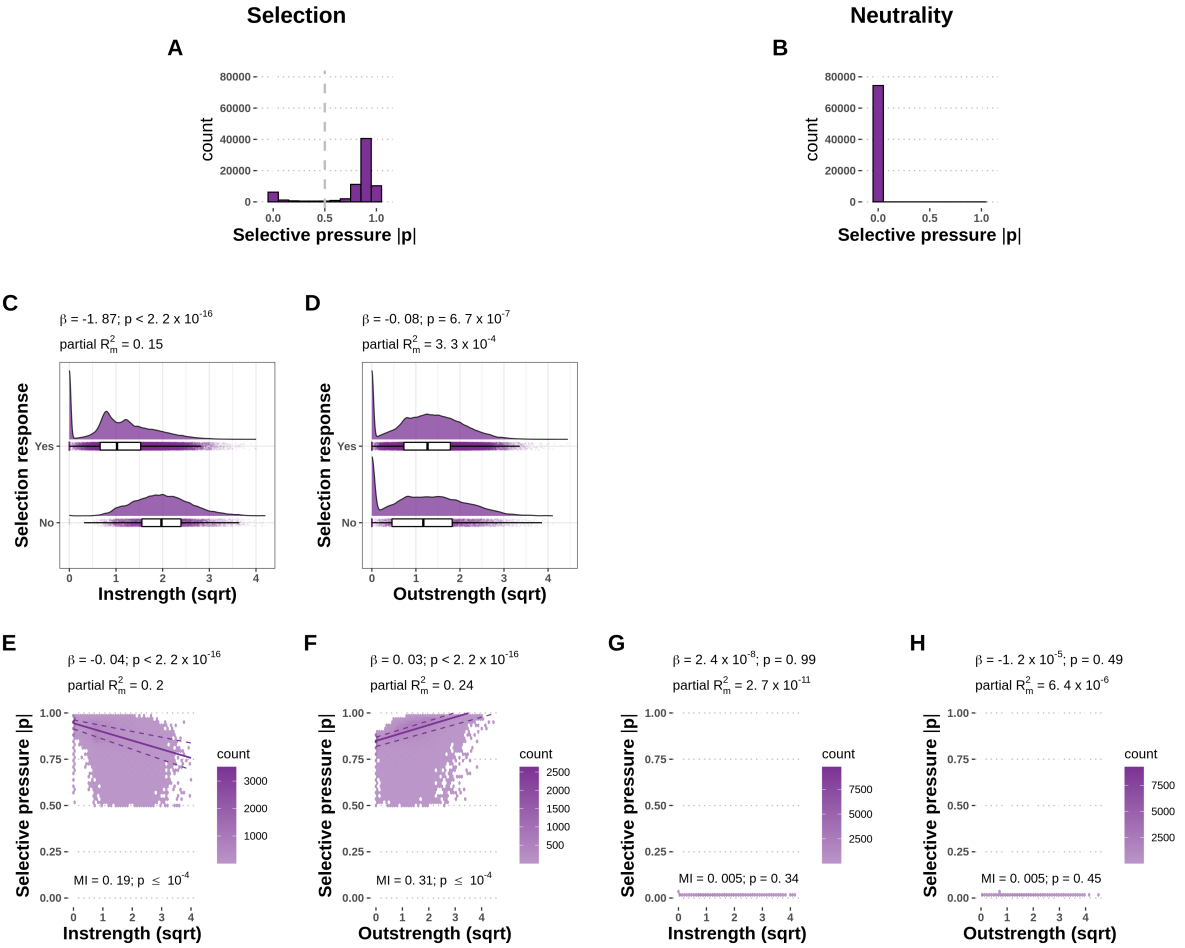

Figure 2.5: **Differential selective pressure is acting on genes based on their centrality. A, B** - Distributions of the measured selective pressure in selected (A) and neutral (B) populations. Genes with a selective pressure above 0.5 were categorized as responsive to selection. **C, D** - High instrength genes are less likely to respond to selection. Absolute instrength (C) has a strong significant negative effect on the probability of selection response. Absolute outstrength (D) has a weak significant negative effect on the probability of selection response. **E, F** - In the subset of genes that responded to selection, high instrength (E) decreases the selective pressure, while high outstrength (F) increases the selective pressure acting on individual genes. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **G, H** - Absolute instrength (G) and outstrength (H) have no significant effect on the selective pressure in the non-selected populations. The dataset consists of 74,443 genes from 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Coefficients, p-values and partial marginal R^2 measures are estimated using logistic regression and linear mixed-effects models with selection responsiveness or selective pressure as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were using 10,000 permutations.

2.4.5 Global network properties affect the evolvability of expression noise and selective pressure on constituent genes

Lastly, we analysed how topological structures and graph-level network properties affect the expression noise response of constituent genes to selection on a joint dataset of random (Erdős-Rényi), scale-free (Barabási–Albert) and small-world (Watts–Strogatz) network topologies. Jointly analysing genes from all three topology types with linear models, we observed statistically significant interactions between instrength and outstrength and network topology types on both the probability to respond to selection and the selective pressure acting on gene-specific expression noise (Table 2.1). We found that genes in scale-free networks have a significantly higher probability of responding to selection than genes in random networks. These results are in agreement with previous studies reporting a higher evolvability of scale-free networks (Oikonomou and Cluzel, 2006; Greenbury et al., 2010). Conversely, genes in small-world networks have a significantly lower probability of responding to selection than genes in random networks. Furthermore, there are significant effects of interactions between instrength and outstrength with the topology type on the selective pressure on constituent genes.

To investigate which global topological features of the three network models affect expression noise evolution, we performed a principal component analysis (PCA) on 12 graph-level measures. The first two dimensions of the PCA expressed 85.4% of the total dataset inertia (Appendix A.2), so we used the first two principal components (PCs) as synthetic explanatory variables in linear mixed-effects models. The loading of the first synthetic variable (PC1) is dominated by negative loadings of diameter and mean path distance, and the centralization measures, namely positive loadings of outdegree and closeness centralization and negative loadings of indegree and betweenness centralization. The diameter of a network is defined as the longest shortest path between any two nodes. Centralization is a measure of the extent to which a network is centered around a single node and can be computed from different centrality metrics. The loading of the second synthetic variable (PC2) is dominated by the negative loading of the average degree, average indegree and average outdegree measures (Appendix A.2). For a more intuitive interpretation, the signs of both PCs have been switched in the statistical analysis. Therefore, PC1 shown in the results is dominated by positive loadings of diameter, mean path distance, indegree centralization and negative loadings of outdegree centralization, and PC2 is dominated by positive loadings of average degree. We refer to PC1 and PC2 as synthetic network diameter and centralization and synthetic average degree, respectively.

The average expression variance per network is significantly negatively affected by synthetic network diameter and centralization (linear model with synthetic network diameter and centralization coefficient β = -6.19, p-value < 2.2×10^{-16}) and significantly positively affected by the synthetic average degree (linear model with synthetic average degree coefficient β = 13.26, p-value < 2.2×10^{-16}). The mutual information was significant between the average expression variance per network and synthetic network diameter and centralization (MI = 0.21, p-value ≤ 10^{-4}, permutation test) and synthetic average degree (MI = 0.21, p-value ≤ 10^{-4}, permutation test). This finding means that global network properties affect the amplification of noise through noise propagation between the genes. Specifically, networks with a lower diameter, mean path distance, indegree centralization, and higher outdegree centralization and average degree, had higher average gene expression variance. In the selected populations, the average

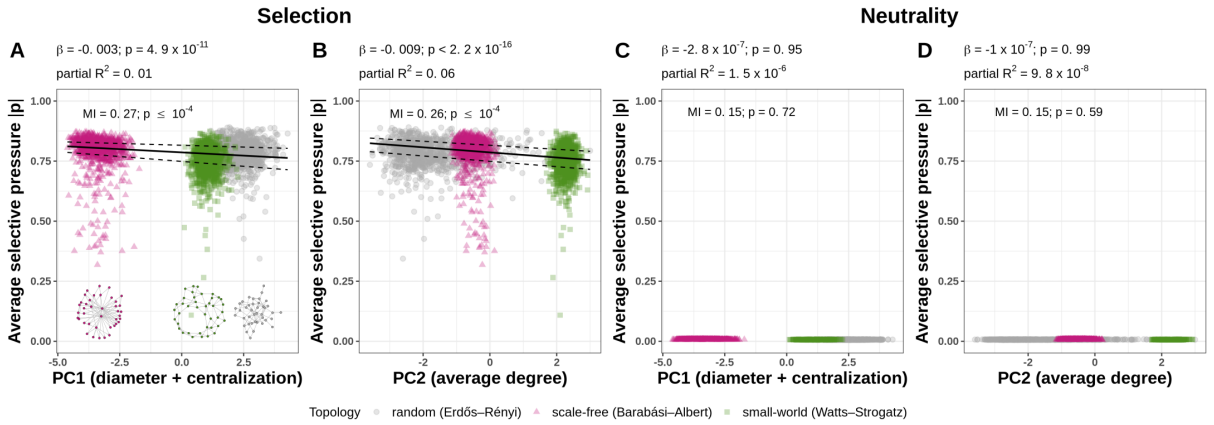

Figure 2.6: **Global network properties affect the average selective pressure acting on gene expression noise under stabilizing selection on gene expression level. A, B** - Principal component variables consisting of the diameter and network centralization (A) and average degree (B) have a significant negative effect on the average selective pressure per network. The two synthetic variables were constructed by performing a principal component analysis on 12 graph-level network metrics. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized over all replicates in each scenario. Coefficients and p-values are estimated using a linear model with average selective pressure as the response variable, and PC1 and PC2 as explanatory variables. Mutual information (MI) p-values were computed with permutation test with 10,000 permutations.

selective pressure per network was significantly negatively affected by both synthetic network diameter and centralization and the synthetic average degree (linear model with synthetic network diameter and centralization coefficient β = -0.003, p-value = 4.9×10^{-11}, Fig 2.6A; synthetic average degree coefficient β = -0.009, p-value < 2.2×10^{-16}, Fig 2.6B). The mutual information was significant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.27, p-value ≤ 10^{-4}, permutation test) and synthetic average degree (MI = 0.26, p-value ≤ 10^{-4}, permutation test). This result shows that the average selective pressure acting on gene-specific expression noise in networks decreases with an increase of network diameter, mean path distance, indegree centralization and average degree per network. Conversely, the average selective pressure increases with an increase of outdegree centralization (Fig 2.6A-B). In the populations evolved under neutrality, neither synthetic network diameter and centralization, nor synthetic average degree, had a significant effect on the average selective pressure per network (linear model with synthetic network diameter and centralization coefficient $\beta = -2.8\times10^{-7}$, p-value = 0.95; synthetic average degree coefficient $\beta = -1\times10^{-7}$, p-value = 0.99, Fig 2.6C-D). Similarly, the mutual information was insignificant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.15, p-value = 0.72, permutation test) and synthetic average degree (MI = 0.15, p-value = 0.59, permutation test).

Table 2.1: **Network topology type affects the probability of responding to selection and selective pressure on gene-specific expression noise under stabilizing selection on gene expression level.**

Response	Explanatory variable	Beta	SE	p-value[1]	
Probability of responding to selection	Instrength	-1.9270	0.0284	$< 2.2 \times 10^{-16}$	****
	Outstrength	-0.0829	0.0226	$< 2.6 \times 10^{-4}$	***
	Scale-free (BA) topology[2]	0.9209	0.1075	$< 2.2 \times 10^{-16}$	****
	Small-world (WS) topology[3]	-0.2684	0.0945	0.0045	**
	Instrength:BA[4]	0.0120	0.0516	0.8159	n.s.
	Instrength:WS	0.0006	0.0401	0.9873	n.s.
	Outstrength:BA	-0.2947	0.0252	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0728	0.0333	0.0287	*
Gene-specific selective pressure	Instrength	-0.0377	0.0004	$< 2.2 \times 10^{-16}$	****
	Outstrength	0.0347	0.0003	$< 2.2 \times 10^{-16}$	****
	Scale-free (BA) topology	0.0019	0.0012	0.1404	n.s.
	Small-world (WS) topology	0.0222	0.0013	$< 2.2 \times 10^{-16}$	****
	Instrength:BA	0.0143	0.0007	$< 2.2 \times 10^{-16}$	****
	Instrength:WS	-0.0055	0.0006	$< 2.2 \times 10^{-16}$	****
	Outstrength:BA	-0.0151	0.0003	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0075	0.0005	$< 2.2 \times 10^{-16}$	****

[1] Coefficients and their significance were computed using linear mixed-effects models (see Methods). The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene was calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Genes were termed responsive to selection if their selective pressure was above 0.5. Asterisks indicate statistical significance: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.
[2] Barabási–Albert network model.
[3] Watts–Strogatz network model.
[4] Colons (':') indicate variable interactions.

2.5 Discussion

In this work, we aimed at understanding how natural selection shaped the distribution of expression noise levels between genes in the genome. We hypothesized that selection for low noise at the network level translates into differential selective pressures at the gene level. To test this hypothesis, we developed a new gene regulatory network evolution model that incorporates stochastic gene expression, where the gene expression mean and variance are both heritable and, therefore, potentially subject to natural selection. We simulated the evolution of gene-specific expression noise in populations of model gene regulatory networks under selective and non-selective conditions. In agreement with our hypothesis, we observed that individual genes respond differently to the global selective pressure and that this response depends on the local and global network properties. In particular, we found that genes of high centrality exhibit a stronger selective pressure to reduce gene-specific expression noise under stabilizing selection on the expression level and that the genetic network structure affects the propagation and evolvability of gene-specific expression noise. In the following, we further discuss the implications of differential selective pressure acting on constituent genes in gene networks.

2.5.1 Mechanisms of intrinsic noise reduction

In this study we abstracted and summarized the many determinants of intrinsic expression noise into a single parameter, which can be viewed as a modifier locus that can directly change the intrinsic noise of a given gene. This simplification permitted us to investigate the evolution of expression noise in gene networks with computationally feasible evolutionary simulations. In reality, multiple factors that affect gene expression variance in biological systems have been reported. These include epigenetic factors, such as chromatic dynamics (Sun and Zhang, 2019) and presence of chromatin remodelling complexes (Newman et al., 2006). Other factors affect transcription directly and can, therefore, control expression noise: the promoter shape (Sigalova et al., 2020), presence of a TATA box (Newman et al., 2006), presence and number (Sharon et al., 2014) of TF binding sites, TF binding dynamics (Azpeitia and Wagner, 2020), presence of TF decoy binding sites (Dey et al., 2020), and transcription rate. Factors affecting translation have also been shown to play a role in controlling noise: miRNA targetting (Schmiedel et al., 2015), mRNA lifetime, translation rate, and post-translational modifications such as the protein degradation rate. Compartmentalization of proteins by phase separation has also been shown to reduce noise (Klosin et al., 2020). Lastly, gene expression costs can also affect the gene expression level distributions, and thereby expression level noise (Charlebois, 2015). We have demonstrated the existence of a general selective pressure acting on gene expression noise. Biological organisms may differ in the mechanisms used to respond to this selective pressure, calling for further, data-driven, investigations.

2.5.2 Global network structure impacts noise propagation and evolution

By simulating thousands of networks with distinct structures, we were further able to assess the impact of global network characteristics on gene-specific selective pressure. Given that there is a trade-off between the fitness advantage of reducing gene-specific expression noise

at the gene level and its mechanistic cost (for instance, in terms of mRNA processing (Hausser et al., 2019)), evolving the global network structure may offer an alternative way to reduce network-level noise. Several motifs recurrently found in regulatory networks have an impact on expression noise, such as negative (Becskei and Serrano, 2000; Dublanche et al., 2006; Grönlund et al., 2013) and positive autoregulation (Alon, 2007), feed-forward loops (Alon, 2007; Charlebois et al., 2014; Camellato et al., 2019) and interlinked feed-forward loops (Chepyala et al., 2016).

It is important, however, to distinguish two aspects when considering the effect of the network structure on the expression dynamics of constituent genes: the network structure, *i.e.* the topology of the graph, and the strength of each of the regulatory interactions, both of which impact expression noise. The same network topology, but with different regulatory interactions strengths, can give rise to markedly different network behaviours. In the *gap* gene system, for example, it was shown that multiple subcircuits share the same regulatory structure, but yield different expression patterns because of their differences in active components and strength of regulatory interactions (Verd et al., 2019). It results that network models of gene expression noise must incorporate both graph topology and interaction strength between all constituent genes. The Wagner model constitutes a simple framework that fulfills these two conditions. However, it has its limitations. Namely, it is not fine-grained enough to capture the complex dynamics of real regulatory networks. Models that incorporate higher molecular detail, such as large systems of differential equations, are necessary to precisely capture in fine detail the expression dynamics of a real biological network, but they come with a cost in terms of high computation time (preventing their use in evolutionary simulations), low tractability and, often, the inability to model noise.

2.5.3 Implications of selection on expression noise on the evolution of genomes and gene regulatory networks

One mechanism by which networks and genomes evolve is gene duplication. Gene duplications are a major source of new genes and thought to be a primary source of evolutionary novelties. It has been long proposed that new functionality arises from duplicated genes by allowing the other gene copy to acquire new functions (neofunctionalization) or improve existing functions (subfunctionalization) by relaxing the selective pressure acting on a single gene through an additional redundant copy (Ohno, 1970). However, most of the time the redundant copy is lost before new functionality can arise (Lynch and Conery, 2000), either by genetic drift alone or because having the extra copy is deleterious. The redundant copy has a chance to evolve a new function or improve an existing one while it is evolving neutrally or reaches fixation in the population, or alternatively, if there is some fitness benefit of the additional copy that increases its frequency in the population. Some benefits of having additional gene copies have been shown, such as increased expression level for genes whose pre-duplication expression level was far from the optimum (Riehle et al., 2001). Moreover, duplicating a gene reduces its expression noise (Rodrigo and Fares, 2018; Chapal et al., 2019), averaging the stochastic events over the two gene copies. The reduction of expression noise may, therefore, constitute another benefit of a gene duplication, increasing its chance of fixation in the population. As the gene number increases in bacterial genomes, the number of regulatory genes increases 4-fold (Molina and van Nimwegen, 2008), indicating a gene duplication is more likely to stay if the gene is a regulatory gene. We hypothesize that selection on expression noise, particu-

larly on regulatory genes, could, therefore, be one of the forces driving the maintenance of duplicated genes.

2.5.4 Applications of the model framework to study complex systems

In this study, we developed a new regulatory and evolutionary model to study expression noise in gene regulatory networks. The model represents key features of evolving gene regulatory networks, namely the non-independence of gene expression levels and fitness determined by the expression level of many or all genes in the network. Our results revealed that differential selective pressure acts on intrinsic expression noise of constituent genes and that network-level topological properties affect noise propagation within the network.

Although our study focused on gene regulatory networks, our conclusions potentially apply to a broader range of systems. In particular, we posit that any system that fulfills two essential properties will exhibit a similar behavior: (i) the amount of product of each system component (here called "expression level") is not independent and (ii) the performance (here termed "fitness") is determined by the product level of one or several of the components of the system. There are many other complex systems that fulfill these criteria, such as biological metabolic networks, ecological food webs, neural networks, economies, transportation and other infrastructure networks, and social networks. We expect that the same constraints act on noise in elements of these systems, and that some of the conclusions from gene regulatory networks could be carefully applied to other complex systems.

2.6 Conclusion

Our results show that selection for low expression noise acting on a system (the gene network) resulted in differential selective pressures on its individual components (the genes). We demonstrated that the position of the gene in the network and the global network structure act as important drivers of the evolution of intrinsic expression noise. Investigating how gene networks evolve to cope with expression noise will reveal mechanisms of how complex biological systems adapt to function with an inevitable molecular noise in their components. A better comprehension of these mechanisms is a prerequisite to understand the evolution of complexity in biological systems, from the first self-replicating RNA systems to modern eukaryotic cells expressing tens of thousands of genes.

Chapter 3

The evolution of gene expression level and noise in changing environments

"A simple change of scenery can bring about powerful shifts in the flow of time and emotions."

– Haruki Murakami, *South of the Border, West of the Sun* (1992)

3.1 Abstract

The variability of gene expression levels, also known as gene expression noise, is an evolvable phenotypic trait subject to selection. Gene expression noise is detrimental under stabilizing selection on gene expression level. On the other hand, in changing environments, where genes are subject to directional or fluctuating selection, expression noise may be beneficial. However, expression noise propagates along the gene network, making the evolution of connected genes interdependent. Here, we explore how their position in the gene network constrains the evolution of genes under selection using an in silico evolution experiment. We simulate the evolution of populations of model gene regulatory networks under directional and fluctuating selection on the gene expression level while allowing the basal expression level and expression noise level to mutate. We find that target genes, regulated by other genes, were more likely to respond to directional selection by changing the mean expression level than regulator genes. Moreover, the intrinsic expression noise of genes under directional selection transiently increased, showing that expression noise may play a role in the adaptation process towards a new mean expression optimum. Similarly, target genes under fluctuating selection were more likely to increase their gene-specific expression noise than regulator genes. These findings suggest that both the mean and variance of gene expression levels respond to selection due to changing environments – and do so in a network-dependent manner.

3.2 Introduction

Gene expression levels directly affect the viability and fitness of the organism. An important component of the gene expression profile, in addition to the mean gene expression level, is the variability of this level around the mean, *i.e.* gene expression noise. Gene expression noise is the inevitable variability of gene expression levels due to the stochasticity of the process of gene expression itself. It has been shown to be an evolvable trait independent of expression mean. For example, stabilizing selection on gene expression level reduces gene expression noise (Lehner, 2008). However, the response of expression noise to other selection scenarios, such as directional and fluctuating selection, has not yet been sufficiently explored.

Expression noise has been suggested to be beneficial during the adaptation of the mean expression level to a new expression level optimum in genes under directional selection (Duveau et al., 2018). Increasing expression noise as a bet-hedging strategy of adapting to fluctuating environments was experimentally demonstrated in previous studies. Expression noise can create phenotypic heterogeneity in clonal population and improve its survivability in changing environments. For instance, the increased cell-to-cell variability of a signal transduction system in *E. coli* permits growth in case of rapid oxygen availability fluctuations (Carey et al., 2018). It was shown that expression noise of the transcription factor *comK* drives cell fate determination in *Bacillus subtilis* and enables competency in a proportion of the cell population (Maamar et al., 2007), and increasing the noise increases the response range of the competency circuit (Mugler et al., 2016). Furthermore, several observed properties of gene regulatory networks have been attributed to fluctuating selection in gene networks (Tsuda and Kawata, 2010).

The response of expression noise to different selection scenarios has been studied in singular genes, but not in context of genetic networks. Since expression noise has been demonstrated to propagate in gene networks (Pedraza, 2005), it is important to take into account the network background of a gene under selection when investigating the evolution of expression noise. Changing the expression noise of a gene changes the expression noise of downstream elements, because expression noise propagates between genes in the network. It has been argued that most genes are under stabilizing selection on expression level and, consequently, it can be assumed that low expression noise will be maintained. Increasing expression noise in some part of the network as a response to fluctuating or directional selection would increase noise in neighbouring parts of the network, which might conflict with stabilizing selection acting on other genes and decrease the efficiency of selection. Therefore, the adaptation to directional or fluctuating selection on gene expression level of individual genes might be constrained by their position in the gene network.

In this study, we used a computational gene regulatory network model to simulate the evolution of populations of model gene regulatory networks in two different selection scenarios, directional and fluctuating selection acting on gene expression levels, in order to investigate the adaptability constraints imposed on individual genes by their gene network background. We find that regulator genes are less likely to adapt to directional selection that non-regulator genes, because the change to their mean expression level changes the mean expression level of downstream genes. Regulators are also less likely to adapt to fluctuating selection by increasing their expression noise as part of a bet-hedging strategy.

3.3 Material and Methods

To study the evolution of gene expression noise and mean expression level in gene regulatory networks in changing environments, we simulated the evolution of populations of model gene regulatory networks under directional and fluctuating selection.

3.3.1 Gene regulatory network model with evolvable gene expression mean and noise level

The noisy gene regulatory network model introduced in Chapter 1 is modified and summarized here. An individual is represented by a genotype, which is realized into the phenotype through a set of difference equations. The genotype consists of the regulatory network matrix $W = \left(w_{ij}\right)_{1\le i\le n,\ 1\le j\le n}$, intrinsic noise vector $\{\eta_i^{\text{int}}\}_{1\le i\le n}$, and the basal expression level vector $\{s_i^{\text{basal}}\}_{1\le i\le n}$. The regulatory network matrix W determines the presence, strength and sign of the regulatory interactions between each pair of genes. The sign of the elements of the regulatory network matrix w_{ij} determines whether the interaction is activation or repression of downstream expression level, and the value determines the strength of the interaction. The intrinsic noise vector $\{\eta_i^{\text{int}}\}_{1\le i\le n}$ determines the intrinsic expression variance of each gene in the network. The basal expression level vector $\{s_i^{\text{basal}}\}_{1\le i\le n}$ determines the constitutive expression level of each gene, which is present regardless of the input from regulatory genes. The genotype is realized by updating the expression level of each gene in every time step using the following rule:

$$s_i(t+1) \sim \mathcal{N}\left(s_i^{basal} + a_i(t), \eta_i^{int}\right). \tag{3.1}$$

The expression level of each gene in each time step is drawn from a normal distribution, which has a mean of the sum of the activation rate in the previous time step and the basal expression level, and a variance of the intrinsic noise genotype. The activation rate is defined as the sum of the effects of all regulators:

$$a_i(t) = \sum_{j=1}^{n} w_{ij} \cdot s_j(t), \tag{3.2}$$

The expression levels of each gene are synchronously updated in every time step t for T_r ($T_r = 50$) timesteps. The expression level vector at the final time step t_{50} is taken as the phenotype of each genotype.

3.3.2 Forward-in-time simulation of expression mean and noise evolution

Populations of model gene regulatory networks were evolved for 10,000 generations using an evolutionary algorithm consisting of repeated cycles of phenotype realization, reproduction, mutation and recombination. Firstly, all individuals in the population had their genotype realized into a phenotype. Next, fitness of each individual was calculated as a function of the distance of gene expression levels from an optimal gene expression level vector, weighted by fitness contribution given by ρ:

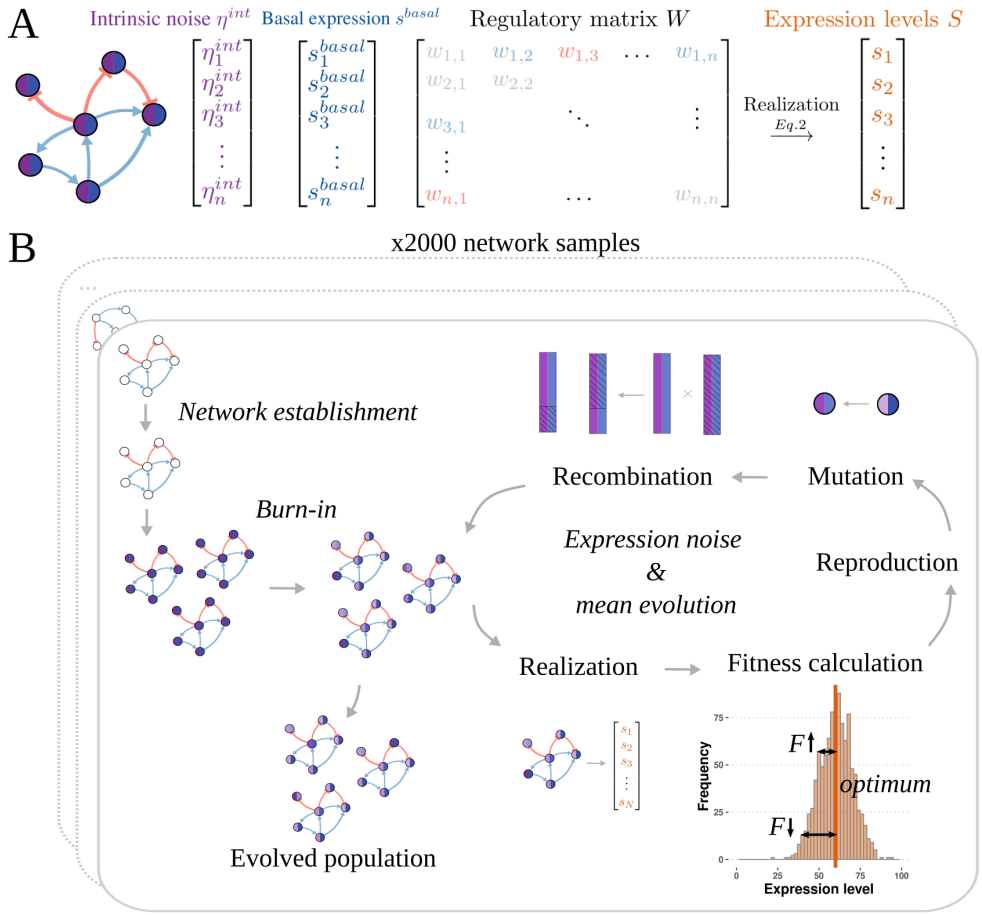

Figure 3.1: **The evolution of gene-specific expression noise was simulated using populations of model gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions. A** - Gene regulatory network model. The genotype consists of the intrinsic noise vector η^{int}, basal expression level vector s^{basal} and regulatory matrix W. The intrinsic noise vector defines the gene-specific expression variance of each gene in the network. The basal expression level vector determines the constitutive expression level of each gene, which is present regardless of the input from regulatory genes. The regulatory matrix defines the regulatory interactions in the network. The genotype is realized into the phenotype using the dynamical equation described in the main text. The phenotype is given by the state vector S, which represents the expression level of each gene in the network. **B** - Steps of the evolutionary simulation process. Each established network configuration was used as a founding network for the network populations used in the noise evolution simulation. In every generation, genotypes are realized and phenotypes (expression levels) are sampled from the last time step. Fitness is calculated from the expression levels. If the populations are evolved under selection, fitness is calculated as the distance of the expression level of each gene from the optimal expression level. Genotypes are reproduced based on their relative fitness and mutations in the intrinsic noise vectors are introduced. Noise genotype vectors are recombined by randomly choosing individuals for recombination and shuffling their noise vectors. The process is repeated for 10,000 generations.

$$F(s) = e^{-\sum_{i=1}^{n} |s_i^{opt} - s_i|/(n\rho_i)}. \tag{3.3}$$

The fitness contribution ρ defines the magnitude of the fitness cost resulting from the deviance from the optimum expression level and was set to 1 for all genes, defining an equally strong selective pressure on all genes. Individuals were reproduced into the next generation by taking the fitness value as the probability of reproducing in this generation and drawing the genotypes with replacement until the population size is reached. A constant population size was maintained. Mutations were introduced in the intrinsic noise vector and the basal expression level vectors with a probability of μ_η ($\mu_\eta = 0.005$) and μ_{sb} ($\mu_{sb} = 0.005$) per gene, respectively. The mutation values for intrinsic noise mutations were drawn from an uniform distribution $\mathcal{U}(0, 200)$ and the mutation values for basal expression level mutations were drawn from an uniform distribution $\mathcal{U}(0, 100)$. Recombination was implemented by drawing a recombining individual with a probability $r = 0.05$, randomly drawing two genome breakpoints, and exchanging the recombinant fragments with another randomly drawn individual in the population. The network topology was immutable.

3.3.3 Pipeline

We generated 2,000 unique 40-gene network topology samples using the *igraph* package (Csardi and Nepusz, 2006) in R. Each network topology sample was used to simulate the evolution of one population of networks. The regulatory interaction values that yield random gene expression levels were established in the network establishment process, as described in section 2.3.2. To have a neutral starting point before applying directional or fluctuating selection for the main experiment, stabilizing selection was applied on all genes in each network population for 5,000 generations, until mutation-selection-drift balance was reached (burn-in phase). During this phase, the intrinsic noise and basal expression levels were allowed to mutate. The burn-in phase ensured that all 40 genes in each network had already been evolving under stabilizing selection for long enough that mutation-selection-drift balance had been reached, and any selection response observed after an environmental shift would be due to directional or fluctuating instead of stabilizing selection. The heterogeneous population of genotypes at the end of the burn-in phase was the starting point for the main simulation, the expression noise and mean evolution under an environmental shift. One gene in each network was randomly chosen to undergo an environmental shift, in which its optimum expression level s_i^{opt} was increased or decreased by 20% of s_{max} relative to its previous value. The remaining 39 genes remained under stabilizing selection, with their respective optimum expression level values unchanged. Three selective scenarios were simulated: positive and negative directional selection, in which the optimum expression level was increased or decreased by by 20% of s_{max}, respectively, and one fluctuating selection, in which optimum expression level was oscillating between an increase and a decrease of 20% of s_{max} every other generation. The intrinsic noise and basal expression levels were mutable, and the network topology was immutable. The populations were evolved for 10,000 generations.

To distinguish the effect the network topology had on the evolvability of gene expression in changing environments, we also simulated the evolution of single, isolated genes under directional and fluctuating selection. A dataset of 1,000 genes with a random basal expression level value drawn from $\mathcal{U}(20, 80)$ was used to generate populations of 1,000 individuals and

their evolution was simulated using the previously described pipeline. Each population was evolved for 5,000 generations under stabilizing selection, after which directional or fluctuating selection was applied, or stabilizing selection maintained, for an additional 30,000 generations. Directional selection was applied by setting the optimum expression level to its basal expression level value $s^{basal} + 20\% s_{max}$, and fluctuating selection was applied by setting the optimum expression level to alternate between basal expression level $s^{basal} + 20\% s_{max}$ and $s^{basal} - 20\% s_{max}$ every other generation. To distinguish the response of the mean expression level and intrinsic noise, and their co-evolutionary dynamics, three scenarios were simulated: 1) with immutable basal expression level and mutable intrinsic noise; 2) with both basal expression level and intrinsic noise mutable; 3) with mutable basal expression level and immutable intrinsic noise.

The gene regulatory network model and evolutionary framework was implemented in C++ and the simulations results analysed in R 3.6.3 (Team, 2021).

3.4 Results

To study the evolution of gene expression noise and mean expression level under directional and fluctuating selection, we simulated the evolution of single, isolated genes and genes connected in model gene regulatory networks with unique network topologies. In each network, one gene was randomly chosen to be under directional or fluctuating selection, while the rest remained under stabilizing selection. The populations of model networks were evolved with mutable gene intrinsic noise and basal expression level, and a fixed network topology. We first present the results of the single genes and genes in regulatory networks in populations evolved under directional selection.

3.4.1 Gene expression noise is transiently increased as a response to directional selection

We studied the response of intrinsic expression noise and mean expression level of genes under directional selection on gene expression level in single genes by simulating the evolution of 1,000 genes with random basal expression level values. To distinguish the response of the mean expression level and intrinsic noise, and their co-evolutionary dynamics, we simulated three scenarios: first, in which the intrinsic noise was mutable, but the basal expression level was not (Fig. 3.2A-F); second, in which both intrinsic noise and basal expression level were mutable (Fig. 3.2G-L); third, in which the basal expression level was mutable, but intrinsic noise was not (Fig. 3.2M-R).

In the first scenario, in which only the intrinsic noise was mutable, but basal expression level was not, we found an increase in expression variance and intrinsic noise after directional selection was applied. An example of the evolutionary dynamics is shown in Fig. 3.2A-D. The mean expression level did not change after directional selection was applied, as basal expression level was immutable (Fig. 3.2 A, D). However, the expression variance and intrinsic noise increase (Fig. 3.2B, C), indicating that noise was beneficial in genes whose expression level was far from the optimum. In the dataset of 1,000 simulated genes, the average increase of the intrinsic noise was significantly higher in genes under directional selection than genes which remained under stabilizing selection (p-value $< 2.2 \times 10^{-16}$, Wilcoxon's test, Fig. 3.2E).

In the second scenario, in which both basal expression level and intrinsic noise were

mutable, two distinct evolutionary phases were observed: the adaptive phase, during which the basal expression level was evolving towards the new optimum expression level; and the postadaptive phase, after the basal expression level has reached the optimum expression level. An example of the evolutionary dynamics is shown in Fig. 3.2G-L. A co-evolutionary pattern was observed - expression level variance, determined by the intrinsic noise, was elevated while the basal expression level was evolving towards the new optimum. Expression noise, and intrinsic noise, was reduced once the mean expression level reaches the optimum expression level. The average increase of the intrinsic noise during the adaptive phase was significantly higher in genes under directional selection than genes which remained under stabilizing selection (p-value = 1.25×10^{-10}, Wilcoxon's test) or genes under directional selection in the postadaptive phase (p-value = 8.19×10^{07}, Wilcoxon's test, Fig. 3.2K).

In the third scenario, in which the basal expression was mutable, but the intrinsic noise was not, we again observe the adaptive and postadaptive phase of expression mean evolution. An example of the evolutionary dynamics is shown in Fig. 3.2M-R. However, since the intrinsic noise could not evolve, there is no elevation of expression level variance during the adaptive phase (Fig. 3.2N). Importantly, there is significant increase in expression level variance during the adaptive phase when the noise cannot evolve, indicating that the basal expression level variants segregating in the population did not cause the elevated expression variance signal in the scenario in which both mean and noise could evolve (Fig. 3.2H).

These results showed that gene expression noise has a fitness benefit if the mean expression level is not near the optimal level, *i.e.* the gene is under directional selection. We observed that expression noise remains elevated if the mean expression level cannot change due to constraints. However, if the mean expression level can evolve towards a new optimum, once it has reached the new optimum a selection switch happens - the expression level is under stabilizing selection again, and expression noise is deleterious. Therefore, expression noise is beneficial transiently while the expression mean is evolving towards a new optimum, but becomes deleterious once it has reached the new peak. In case the expression mean cannot evolve, expression noise has a constant fitness benefit. The evolution of the expression mean may be constrained by the position of the gene in the gene regulatory network, which is what we explore in the next section.

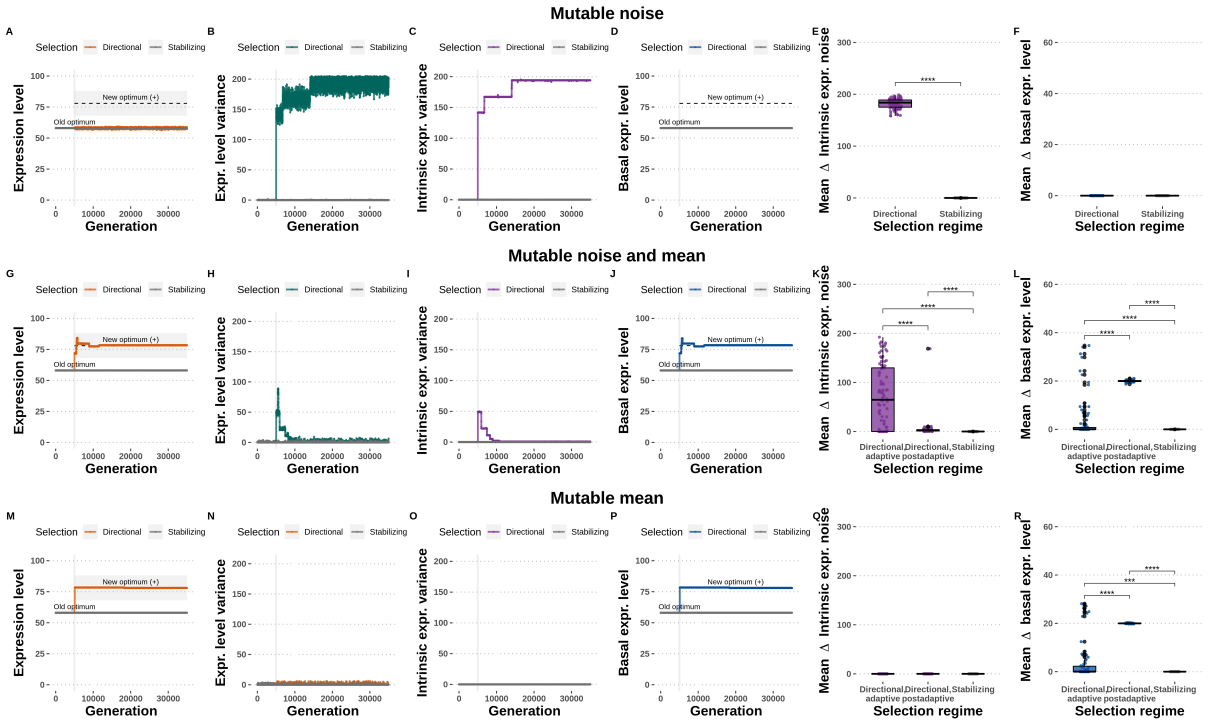

Figure 3.2: **Expression noise is beneficial under directional selection if the mean expression level is fixed, or transiently while the mean is evolving to a new optimum.** **A-D** - Evolutionary dynamics of an example gene evolving under directional selection with mutable noise levels (A - mean expression level, B - expression level variance, C - intrinsic noise, D - basal expression level). **E-F** Average change of intrinsic noise (E) and average change of basal expression levels (F) relative to the average pre-selection levels in genes with mutable noise that undergo directional selection or remain under stabilizing selection. **G-J** - Evolutionary dynamics of an example gene evolving under directional selection with mutable noise and mean expression levels (G - mean expression level, H - expression level variance, I - intrinsic noise, J - basal expression level). **K-L** Average change of intrinsic noise (K) and average change of basal expression levels (L) relative to the average pre-selection levels in genes with mutable noise and mean expression levels that undergo directional selection or remain under stabilizing selection. **M-P** - Evolutionary dynamics of an example gene evolving under directional selection with mutable noise levels (M - mean expression level, N - expression level variance, O - intrinsic noise, P - basal expression level). **Q-R** Average change of intrinsic noise (Q) and average change of basal expression levels (R) relative to the average pre-selection levels in genes with mutable noise that undergo directional selection or remain under stabilizing selection. Dataset consists of 1,000 genes evolved for 30,000 generations. Asterisks indicate statistical significance of Wilcoxon's tests: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

3.4.2 Regulator genes in gene regulatory networks are less adaptable to directional selection than non-regulator genes

Next, we studied the response of intrinsic expression noise and mean expression level of genes under directional selection in gene regulatory networks to investigate whether the network background has an effect on the evolvability of genes connected in gene networks. The evolutionary trajectory of a gene under directional selection from an example 40-gene network is shown in (Fig 3.3A-D). Before directional selection is applied, in the mutation-selection-drift balance phase, the expression level of the focal gene is under stabilizing selection and, consequently, the population-wide mean expression level shows little variation (Fig 3.3A). After the optimal expression level is increased, the population mean expression starts increasing until it reaches the new optimum. We categorized genes as adapted if their mean expression level reached $s^{opt} \pm 10\% s_{max}$. For genes whose expression level adapted to directional selection, we distinguish two evolutionary phases: i) adaptation, during which the expression mean is evolving towards the new optimum, and ii) postadaptation, after the expression mean has reached the new optimum. During these two phases, we track the changes in the phenotypic noise (Fig 3.3B), the intrinsic noise genotypes (Fig 3.3C) and the basal expression level genotypes (Fig 3.3D). In this example (gene G16, network 1765), the gene expression level managed to adapt to a new optimum and during the adaptation phase, the phenotypic and intrinsic noise was increased. In the postadaptation phase, the phenotypic and intrinsic noise decrease to their average values before the optimum shift.

Figure 3.3 *(next page)*: **Adaptation to directional selection on gene expression level is constrained by regulatory interactions in the gene regulatory network. A-D** - Evolutionary dynamics of an example gene evolved under directional selection (A - mean expression level, B - expression level variance, C - intrinsic noise, D - basal expression level). Black vertical lines indicate the initiation of the environmental shift, *i.e.* directional selection, and the ending of the adaptive phase, respectively. Dashed horizontal lines indicate the optimal expression levels under directional selection. **E** Proportion of genes in each regulatory category that responded or did not respond to directional selection. **F** Time to adaptation of adapted genes in each regulatory category. **G-I** Evolutionary metrics of genes in each regulatory category that underwent directional selection or remained under stabilizing selection after the environmental shift. Relative change of expression variance, phenotypic noise (G), relative change of intrinsic noise (H), relative change of basal expression level (I) show different pattern in adaptive and postadaptive phases. Expression noise was increased during the adaptive phase, but was reduced after adaptation. The dataset consists of 2,000 40-gene networks evolved for 10,000 generations. Acronyms: MSD balance - Mutation-selection-drift balance. Asterisks indicate statistical significance of Wilcoxon's tests against a default value of $\mu = 0$: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

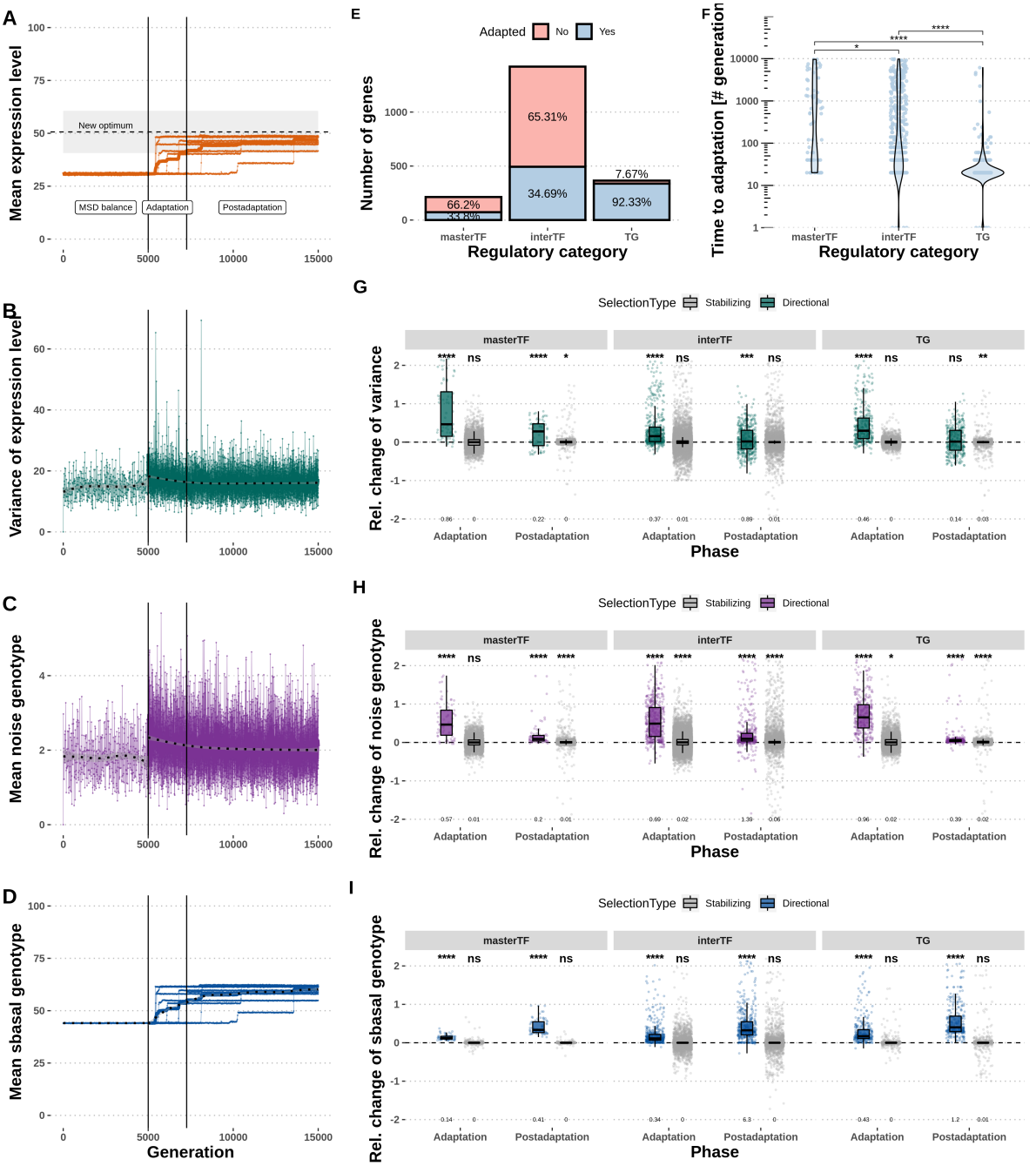

Figure 3.3: see previous page.

Out of 2000 genes under directional selection, 899 (44%) adapted after 10,000 generations. The regulatory category had a significant effect on the adaptivity and the time to adapt. We categorized genes into three categories: i) master transcription factors, which regulate one or more downstream genes, but are not regulated by any upstream gene; ii) intermediate transcription factors, which are both regulated and regulate one or more genes; and iii) target genes, which are regulated by one or more upstream genes, but do not regulate any downstream gene themselves. Target genes are more likely to adapt to directional selection than intermediate transcription factors and master transcription factors (Fig 3.3E). The time to adapta-

tion, defined as the number of generations until the new optimum expression level is reached, is longer in master transcription factors (p-value < 2.2×10^{-16}, Wilcoxon's test) and intermediate transcription factors (p-value < 2.2×10^{-16}, Wilcoxon's test) than target genes (Fig 3.3F).

We then investigated the response of the expression variance (phenotypic noise), intrinsic expression noise and basal expression level during adaptation and postadaptation phase. During the adaptation phase, genes under directional selection had significantly increased their expression variance relative to the average expression variance before the optimum shift, under mutation-selection balance (Fig 3.3G). After adaptation, the expression variance returned to the average expression variance values before the optimum shift. Genes that remained under stabilizing selection did not change their expression variance. The same pattern was observed for intrinsic noise (Fig 3.3H), where genes under directional selection increased their intrinsic noise during adaptation, but returned to preadaptive values after adaptation. Genes that remained under stabilizing selection did not change their intrinsic noise. Lastly, as expected, genes under directional selection increased their basal expression level (Fig 3.3I), as opposed to genes under stabilizing selection.

Next, we looked at the effect the gene's position in the regulatory network had on the propensity of the gene to adapt to directional selection and the strength of its response. Target genes had a significantly higher probability of responding to directional selection than master transcription factors (logistic regression model with coefficient β_{TG} = 3.159, p-value < 2.2×10^{-16}). Intermediate transcription factors did not have a significantly higher probability of responding to directional selection than master transcription factors (logistic regression model with coefficient β_{iTF} = 0.039, p-value = 0.79). Node-level network metrics, absolute instrength and absolute outstrength, had a significant effect on the adaptation probability. Absolute instrength, a metric of how strongly a gene is being regulated by other genes, had a significant positive effect on the adaptation probability (logistic regression with coefficients $\beta_{\text{instrength}}$ = 0.0936, p-value = 2.01×10^{-6}; $\beta_{\text{outstrength}}$ = -0.5970, p-value < 2.2×10^{-16}). The time to adaptation, defined as the number of generations it took to reach the new optimal expression level, was significantly lower in target genes (logistic regression model with coefficient β_{TG} = -1632.0, p-value = 9.72×10^{-14}) and intermediate transcription factors (logistic regression model with coefficient β_{iTF} = 0.039, p-value = 0.79).

3.4.3 Gene expression noise is increased as a response to fluctuating selection

To investigate the effect the gene network has on the evolvability of gene expression in changing environments, we simulated the evolution of single genes and genes in regulatory networks under fluctuating selection on gene expression level. The fluctuating selection was implemented by imposing an alternating optimum expression level that switches between an increased and decreased optimum expression level every other generation. We first report results from the simulations of single, isolated genes.

As with directional selection, we simulated three scenarios to disentangle the response of expression mean and expression noise levels: first, in which the intrinsic noise was mutable, but the basal expression level was not (Fig. 3.4A-F); second, in which both intrinsic noise and basal expression level were mutable (Fig. 3.4G-L); third, in which the basal expression level was mutable, but intrinsic noise was not (Fig. 3.4M-R).

In the first scenario, in which the intrinsic noise is mutable, but the basal expression level

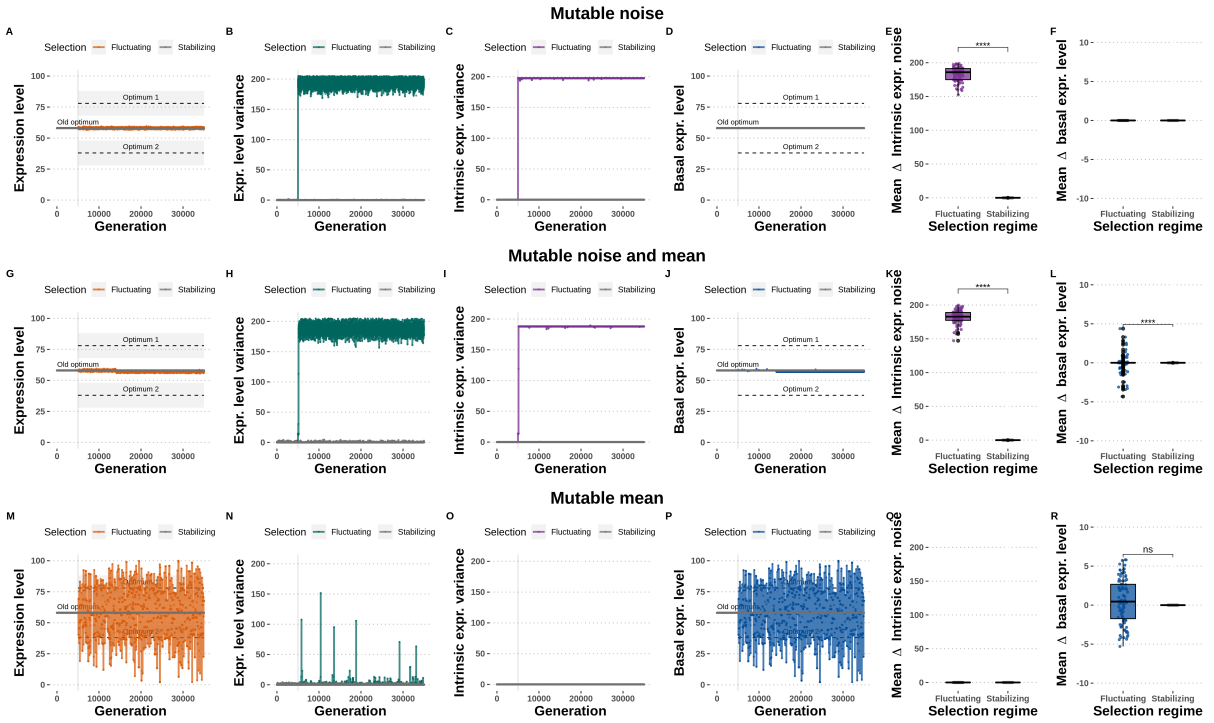

Figure 3.4: **Expression noise is beneficial under fluctuating selection, because the mean expression level is under constraint.** **A-D** - Evolutionary dynamics of an example gene evolving under fluctuating selection with mutable noise levels (A - mean expression level, B - expression level variance, C - intrinsic noise, D - basal expression level). **E-F** Average change of intrinsic noise (E) and average change of basal expression levels (F) relative to the average pre-selection levels in genes with mutable noise that undergo fluctuating selection or remain under stabilizing selection. **G-J** - Evolutionary dynamics of an example gene evolving under fluctuating selection with mutable noise and mean expression levels (G - mean expression level, H - expression level variance, I - intrinsic noise, J - basal expression level). **K-L** Average change of intrinsic noise (K) and average change of basal expression levels (L) relative to the average pre-selection levels in genes with mutable noise and mean expression levels that undergo fluctuating selection or remain under stabilizing selection. **M-P** - Evolutionary dynamics of an example gene evolving under fluctuating selection with mutable noise levels (M - mean expression level, N - expression level variance, O - intrinsic noise, P - basal expression level). **Q-R** Average change of intrinsic noise (Q) and average change of basal expression levels (R) relative to the average pre-selection levels in genes with mutable noise that undergo directional selection or remain under stabilizing selection. Dataset consists of 1,000 genes evolved for 30,000 generations. Asterisks indicate statistical significance of Wilcoxon's tests: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

is not, the mean expression level and basal expression levels expectedly did not significantly change under fluctuating selection (Fig. 3.4A, D). However, expression variance and intrinsic noise were increased (Fig. 3.4B, C). In the dataset of 1,000 genes, the average change of intrinsic noise relative to pre-selection values was significantly higher in genes under fluctuating selection than in genes that remained under stabilizing selection (p-value < 2.2×10^{-16}, Wilcoxon's

test, Fig. 3.4E). Since the optimum expression level alternates every other generation, there is no distinct time point of reaching adaptation under fluctuating selection, unlike directional selection.

In the second scenario, in which both intrinsic noise and basal expression levels are mutable, the same pattern is observed as in the first scenario. The mean expression level and basal expression levels did not significantly change under fluctuating selection (Fig. 3.4G, J), even though the basal expression level was mutable in this scenario. Since fluctuating selection imposes a new optimal expression level every other generation, it is, assumingly, not beneficial for the population to evolve towards any single optimum. However, this likely depends on the frequency of environmental shifts, *i.e.* the amount of time the population spends under selection in each optimum. If instead of changing the environment every second generation, the environment changed less frequently, the population would have more time to adapt to the new optimum and a series of adaptations would be likely. In this scenario, however, the environment shifts very often, and therefore the mean expression level cannot evolve. Since the evolution of mean expression level is constrained, the increased expression noise as a response to fluctuating selection is maintained, as in the case of single genes under directional selection with immutable mean expression level (Fig. 3.2E). In the dataset of 1,000 genes, the average change of intrinsic noise relative to pre-selection values was significantly higher in genes under fluctuating selection than in genes that remained under stabilizing selection (p-value $< 2.2 \times 10^{-16}$, Wilcoxon's test, Fig. 3.4E).

In the third scenario, in which the basal expression level was mutable, but the intrinsic noise was not, the mean expression level and basal expression level showed large fluctuations between the two expression level optima (Fig. 3.4M, P). However, the expression variance was not increased as in the second scenario, in which the mean and noise could evolve, indicating that the observed increase in expression variance as a response to fluctuating selection in the second scenario was not due to the heterogeneity of basal expression level genotypes while the population was evolving towards a new peak. Instead, the increase of expression variance in the second scenario reflects the increase of intrinsic noise as a fitness benefit in response to fluctuating selection on gene expression level.

3.4.4 Target genes in gene regulatory networks respond more strongly to fluctuating selection than non-target genes

Lastly, we studied the response of intrinsic expression noise and mean expression level of genes under fluctuating selection in gene regulatory networks to investigate whether the network background has an effect on the evolvability of genes connected in gene networks. The evolutionary trajectory of a gene under fluctuating selection from an example 40-gene network (gene G37, network 2) is shown in (Fig 3.5A-D). In the mutation-selection-drift balance phase, before fluctuating selection was applied, the expression level of the focal gene is under stabilizing selection. Consequently, the population-wide mean expression level shows little variation (Fig 3.5A), which persisted after fluctuating selection was applied and the optimal expression level started changing every second generation. However, the phenotypic noise (Fig 3.5B) and the intrinsic noise (Fig 3.5C) increased as a response to fluctuating selection.

The strength of the response to fluctuating selection by increasing intrinsic noise may be affected by the position of the gene in the gene regulatory network. In the entire dataset, consisting of 80,000 genes from 2,000 40-gene network topologies, target genes had a significantly

higher increase in relative expression variance and intrinsic noise under fluctuating selection than genes under stabilizing selection in the rest of the network (p-value < 2.2×10^{-16} and p-value < 2.2×10^{-16}, respectively, Wilcoxon's test, Fig 3.5E, F). The relative expression variance and intrinsic noise in intermediate transcription factors and master transcription factors under directional selection was also significantly higher than rest of the genes under stabilizing selection in their respective networks (p-value < 2.2×10^{-16} and p-value = 1.68×10^{-7} for relative expression variance, p-value < 2.2×10^{-16} and p-value = 0.0001 for intrinsic noise, Wilcoxon's test), but the difference was smaller than for target genes. This effect can be explained by noise propagation within the gene regulatory networks. Namely, genes in gene regulatory networks propagate their expression noise downstream, and increasing the intrinsic noise of one gene would increase noise propagated to downstream elements in the network, which would be deleterious in case the rest of the network is under stabilizing selection. Since intermediate and master transcription factors activate or repress the expression level of downstream genes and propagate their expression noise, the increase of their intrinsic noise is constrained by stabilizing selection acting on the downstream genes. Therefore, the adaptability of transcription factors to fluctuating selection is constrained by their regulatory function. Target genes do not have the same constraint, as they do not have downstream elements, and are free to adapt to fluctuating selection by increasing their intrinsic expression noise.

Figure 3.5 *(next page)*: **Fluctuating selection. A-D** - Evolutionary dynamics of an example gene evolved under fluctuating selection (A - mean expression level, B - expression level variance, C - intrinsic noise, D - basal expression level). Black vertical lines indicate the initiation of the environmental shift, *i.e.* fluctuating selection. Dashed horizontal lines indicate the optimal expression levels in the two alternating environments. **E-G** Evolutionary metrics of genes in each regulatory category that underwent fluctuating selection or remained under stabilizing selection after the environmental shift. Relative change of expression variance, phenotypic noise (E), relative change of intrinsic noise (F), relative change of basal expression level (G) show different response strength to fluctuating selection depending on the gene regulatory category. Target genes increase intrinsic noise as a response to fluctuating selection to a higher degree than non-target genes. Dataset consists of 2,000 40-gene networks evolved for 10,000 generations. Acronyms: MSD balance - Mutation-selection-drift balance. Asterisks indicate statistical significance of Wilcoxon's tests against a default value of $\mu = 0$: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

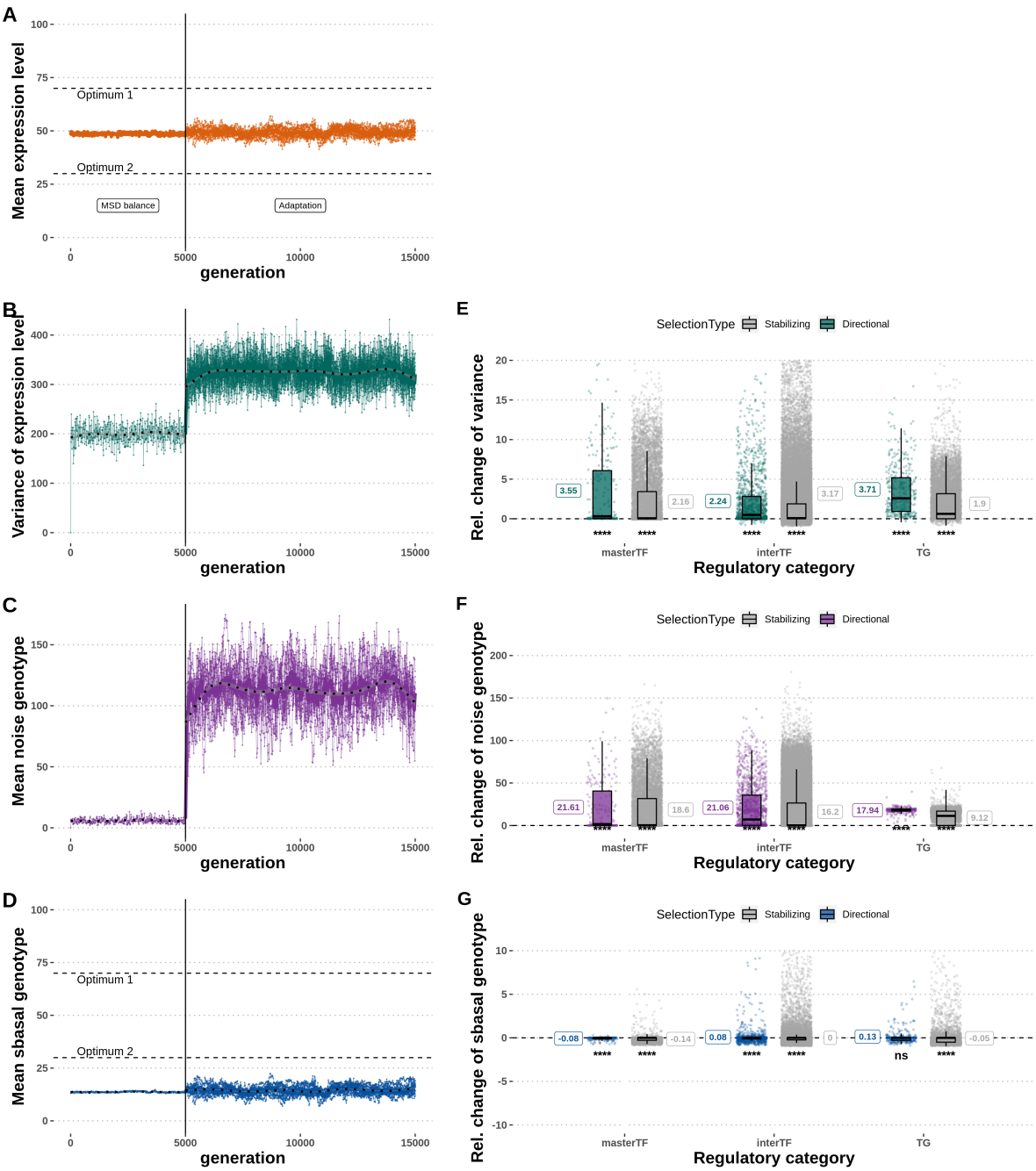

Figure 3.5: see previous page.

3.5 Discussion

In this work, we aimed to investigate how natural selection in changing environments affects the gene expression mean and noise levels in the context of gene regulatory networks. We hypothesized that selection might favor an increase of expression noise as a mechanism of increasing population heterogeneity after an environmental shift, and that there might be differences in adaptability between constituent genes in the gene network. To test this hypothesis, we modified the gene regulatory network model introduced in Chapter 2 to include an evolvable mean expression level as well as expression noise level. We then simulated the evolution of gene expression mean and noise in populations of isolated genes and genes in gene regulatory networks under directional and fluctuating selection, and used genes evolved under stabilizing selection as a control. We found that expression noise was increased during the adaptive phase under directional selection, and was reduced after adaptation to the new mean expression had been reached. Under fluctuating selection, expression noise was consistently increased after selection had been applied. Lastly, in both cases, regulator genes had a lower probability of responding to selection, and responded less strongly than non-regulator genes, indicating a constraining effect of the gene regulatory network on the adaptability of constituent genes.

3.5.1 Limitations of the model

In the simulation framework in this study the population size is kept constant, *i.e.* the individuals are reproduced into the next generation by sampling until the fixed population size is reached. Consequently, the population will never go extinct, even if the individuals have extremely low fitness, and the population will have time to accumulate potentially beneficial mutations and, potentially, adapt. This is not biologically realistic, as dramatic environmental shifts can drive species extinct and survival is not guaranteed if the population is far away from its environmental range. Bet-hedging, the increase of the phenotypic heterogeneity in the population, is a strategy of ensuring that some proportion of the population has a fit phenotype whichever the environmental conditions it finds itself in. Observing the evolution of bet-hedging might be more apparent if the populations who did not adapt through bet-hedging went extinct, instead of being propagated to a constant population size in every generation. The simulation framework used here can be modified to include the possibility of extinction events by adding a non-fixed population size and a viability threshold. Including the possibility of extinction events would enable the simulation of evolutionary rescue, where a declining population manages to survive an environmental shift and recover.

3.5.2 Epistasis slows down adaptation and is affected by network architecture

Adaptation, the process by which organisms adjust to their environment, has often been analysed using the framework of adaptive landscapes, first introduced by Sewall Wright in 1932 (Wright, 1931, 1932). In the adaptive landscape framework, adaptation is depicted as a population of genotypes climbing a landscape consisting of fitness peaks and valleys, corresponding to genotype configurations that are more or less fit to the given environment. The adaptability of a population, in terms of speed of adaptation and the possibility of reaching a fitness

peak, depends on many factors, such as population genetics parameters (e.g. effective population size), initial frequency of adaptive alleles, and the genetic architecture of the selected trait (Olson-Manning et al., 2012). A known factor that slows down adaptation is epistasis. The non-independence of gene expression levels of genes connected in a gene regulatory network can be seen as a form of epistasis. Here, we report evidence of selection at the network level limiting the adaptability of genes in gene regulatory networks, specifically, the gene expression level of transcription factors in changing environments. Reducing the number and strength of interactions between genes, *i.e.* reducing the connectivity of the gene regulatory network, would reduce epistasis and thereby increase adaptability to potential selective pressure in the future. Real biological gene regulatory networks are, indeed, sparse (Leclerc, 2008), and network sparsity was shown to be an emergent property resulting from optimising the explorability of new phenotypes (Busiello et al., 2017).

3.6 Conclusion

Our results showed that intrinsic gene expression noise is beneficial when the gene expression level is under directional or fluctuating selection, such as after an environmental shift. Under directional selection, increased gene expression noise confers a fitness benefit while the mean expression level is evolving towards the new optimal expression level, and it becomes detrimental after the new optimum is reached. Under fluctuating selection, expression noise confers a fitness benefit as a bet-hedging strategy in unpredictable environments. Furthermore, the regulatory function of the gene in the gene regulatory network had a constraining effect on adaptability to directional or fluctuating selection. Transcription factors were less likely to adapt than target genes, because changes in their gene expression levels had direct downstream effects in the network. These results suggest that peripheral genes (target genes) in gene regulatory networks are more adaptable to changing environments than central genes (transcription factors). They further indicate that the gene network background must be taken into consideration when studying the adaptability of gene expression level of individual genes, because the gene network might impose constraints on constituent genes, which would be invisible if genes were considered as isolated components.

Chapter 4

General Discussion

"As biologists, we must grapple with, and reconcile, two very different views of cellular behaviour. On the one hand, we frequently think of cellular functions as being determined by 'circuits' of interacting genes and proteins. In a loosely analogous way to electronic circuits, these chemical circuits encode genetic programmes that underlie differentiation, the cell cycle and other behaviours. They accurately respond to stimuli and generate precise behavioural programmes in individual cells. On the other hand, there is the 'noisy' view of the cell we get when we actually look at cells : they exist in squishy, dynamic and heterogeneous populations, the morphologies, gene expression patterns and differentiated states of which differ from one another, even when environment and genotype are fixed."

– James Locke and Michael Elowitz (Locke and Elowitz, 2009)

In this thesis I have studied the evolution of gene expression noise in gene networks and have found that the evolution of gene expression noise is affected by the gene network background and that expression noise may be deleterious or adaptive. These results have implications on our understanding of the cell, the fundamental unit of organisms. In the following, I would like to discuss the most common conceptualization of the organism, which does not account for noise, as well as some advantages of investigating noise in biological systems.

4.1 Organism as a machine, and why it isn't

4.1.1 History of the organism as a machine concept

"What is the nature of organisms?" is one of the oldest questions in biology, and one of the most pervasive notions is the concept of the organism as an organic machine. The concept of the organism as a machine dates back to the natural philosophy of Descartes in the 17th century. Analogies between man-made tools and organisms have been made since antiquity, but Descartes' metaphysics makes the assumption that organisms are not just similar to machines, but are, in fact, organic machines designed by a divine being (Descartes, 1972). He argued that the only fundamental difference between a human being, or any living being, and a clockwork machine is the degree of complexity. This view has grounded the tone of reductionism in natural sciences in general, and particularly influenced biology to frame living beings as complicated contraptions with specialized parts intentionally assembled to serve a specific purpose. Over the ages the analogies have changed - after the industrial revolution

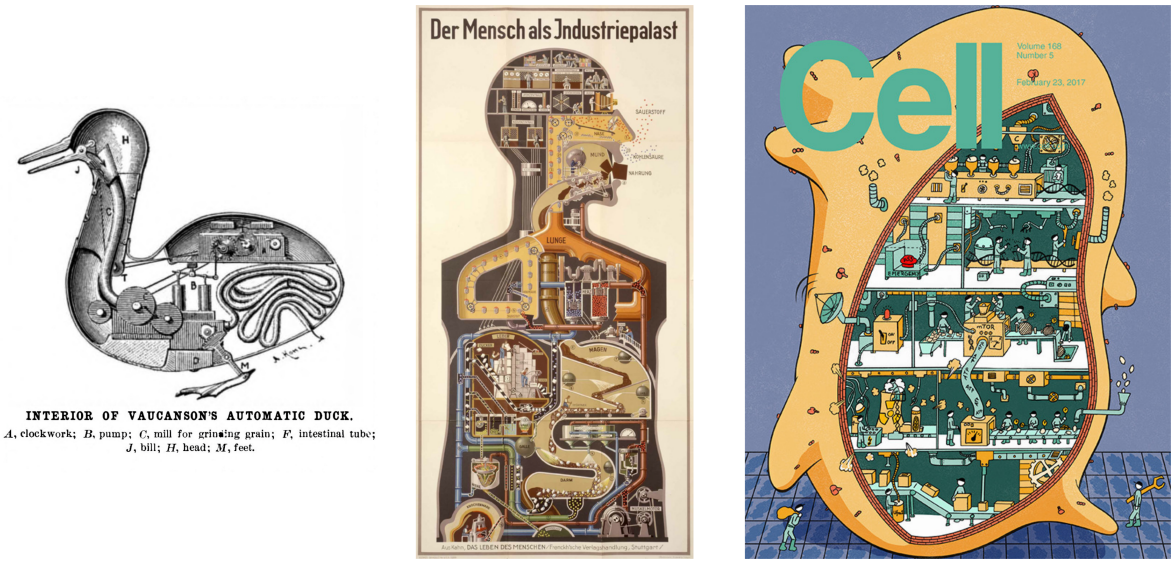

Figure 4.1: **Illustrations of organisms as machines.** From left to right: Schematic of Le Canard digérateur (The Digesting Duck), an automaton created by Jacques de Vaucanson in 1764; Der Mensch als Industriepalast (Man as Industrial Palace) by Fritz Kahn (1926). National Library of Medicine, Stuttgart; Cell as a factory, cover of *Cell* Feb 23, 2017 Volume 168 Issue 5 p743-946. Credit: Sigrid Knemeyer.

organisms have been compared to steam engines, miniature factories, electric circuits, and it seems the latest analogy is of living beings as organic computers. Interestingly, the mechanical devices organisms have been likened to over the ages have always been updated to the most complicated contraption of the time.

The conceptualization of organisms as machines is one of the most ubiquitous ideas in modern biology, particularly prevalent in developmental and molecular biology, but not missing from evolutionary biology. Molecular biology is filled with terms such as protein *machines*, ATP *pumps*, gene *circuits*, molecular *motors*. In developmental biology we have developmental *programs*; in neuroscience the brain as a *computer*; in systems biology the pathway as an *electronic circuit*; in evolutionary biology the phenomenon of hidden variation as evolutionary *capacitance*. Example illustrations of organisms framed as machines can be found in Fig 4.1.

4.1.2 How adequate is framing organisms as organic machines?

Similarities between man-made machines and living beings can be found, but there are many dissimilarities, which is concerning for such an widespread concept in modern biology and can lead to careless assumptions. Multiple criticisms of the machine concept of the organism have been raised (Rosen, 1991). The thermodynamical argument focuses on the thermodynamical states of living beings and machines. Organisms exist in a far-from-equilibrium state, as opposed to machines. Another argument is based on scale. The world at a microscopic scale is vastly different than at macroscopic scale, because physical forces scale differently with the object size. For example, at the molecular scale at which proteins operate, the forces of gravity are relatively minor, but thermal noise and viscous forces from the cytoplasm are significant. At such scales, moving through a solution has been described as "swimming through molasses" or "walking in a hurricane" (Astumian, 2007). There is also the teleological argument, which

posits that the purpose of the machine is defined extrinsically, while organisms do not have an outside purpose and their activity serves only to maintain themselves (Nicholson, 2019). Also, machines are designed to deterministically perform their functions, while organisms display a heterogeneity in their responses to environmental stimuli and due to their stochastic nature. Gene expression, arguably the most important process in the cell, is a probabilistic process, resulting in phenotypic heterogeneity even among isogenic cells in identical environments.

What are the flaws of the machinistic description of organisms? Assuming living beings are machines consciously or unconsciously imparts a bias of intentional design and purpose of each part of the organism. Consequently, it can lead to reasoning that most, if not every, phenotypic trait are the way they are because they serve a specific purpose, and therefore, have been selected for at some point in evolutionary history. This view is known as adaptationism in evolutionary biology, and has been criticized for many decades (Gould and Lewontin, 1979) and fueling exhausting selectionist vs. neutralist debates.

The machine metaphor of the organism does have utility, however. Three categories of usefulness of scientific metaphor can be distinguished: theoretical, heuristic, and rhetorical functions (Bradie, 1999). Metaphors with a theoretical function are useful to help conceptualize, represent and explain the natural phenomenon. Metaphors with a rhetorical function are useful for science communication to non-scientists, and metaphors with a heuristic function are methodological tools used to frame an empirical investigation. Dan Nicholson argues that the machine concept of the organism is an inadequate theoretical and rhetorical metaphor, but has heuristic value (Nicholson, 2013). Namely, for reasons mentioned above, the machine is not an apt representation of real biological organisms, and therefore cannot function as a theoretical metaphor, nor as a rhetorical one. However, bearing in mind that organisms are not truly machines, like Descartes envisioned, the machine concept can be used when studying their parts. The key to the heuristic usefulness is that while organisms are not machines, their parts approximately behave like ones. Parts of organisms, such as organs or organelles, cannot self-organize and self-replicate like whole organisms can, and are dependent on outside influence to maintain themselves, just like machines. Parts also have specific functions in the larger whole, as machines do, as opposed to the entire organism, which doesn't have a purpose. When studying organismal parts, it is convenient to use the machinistic framework to downscale the complexity of the entire organism. Echoing George Box ("All models are wrong, but some are useful") – metaphors can be wrong, but useful. They can be harmful, as well, so they must be employed consciously and cautiously.

4.1.3 A better metaphor for organisms

If the machine metaphor of the organism is incorrect and potentially harmful, what would be a more adequate metaphor for organisms? Several authors have called for looking to physics, instead of engineering, for more fitting analogies. Rather than using human-designed devices as templates for organisms, we should look for thermodynamically open systems operating far from equilibrium with their environment, with which they exchange energy and matter. Such systems are known as dissipative systems in physics. Dissipative systems, particularly dissipative structures, have been proposed as metaphors better suited for explaining living beings (Goldbeter, 2018). Dissipative structures are dissipative systems, so thermodynamically open systems operating far-from-equilibrium, which also have a self-reproducible steady state. Examples of dissipatives structures include a flame, a stream, a tornado, or a hurricane. These

natural phenomena exchange energy and matter with their environment, maintain their existence and steady state through some internal reactions, and can self-reproduce. As such, they have much more in common with a bacteria than a machine does. Perhaps it would be more useful to think of organisms as flame-like or hurricane-like structures instead of MacBook-like structures.

4.1.4 What not being blind to randomness in organisms reveals

Cells are complex systems in which many genes are expressed and interact with each other in what is often represented as gene networks. Gene networks are also very often represented as electric circuits, whether it's a metabolic network, gene regulatory network, or a signalling pathway, and analyzed from the framework of the organism as a machine. This framework implies a deterministic, machine-like behaviour of gene networks. However, gene expression noise is inevitable, making the entire system noisy and affecting its function. This, in turn, imposes differential selective pressure on the constituent genes based on their position in the network. Adapting to noise can affect the structure of the gene network, as well, as shown in Chapter 2. Global network features affect the average noise level of genes in the network. Features of the network architecture can be changed to lower or increase average expression noise in the network, showcasing that the higher level of organization (gene network) has a distinct effect on the lower level of organization (the genes).

Noise can also be exploited by organisms to their benefit, as was shown in Chapter 3, indicating that noise is not just an unavoidable nuisance, but a helpful and adaptive trait, as well.

In this thesis, I have presented evidence of selection on expression noise on a system level shaping the evolution of gene expression. These effects would not have been found had the system not been studied as a complex system using network models, and had the inherent randomness of gene expression not been modelled. Often, biological systems are studied from a reductionist perspective and the complexities of real biological systems are overlooked. Systems biology as a field aims to avoid this simplification and instead integrate different levels of organization of biological systems, and it has made significant progress in the past two to three decades. However, it is still a fairly young field, being established as a field of science at the turn of the millennium.

4.2 Why should a non-biologist care about expression noise?

Biology is the scientific study of living beings. Living beings, such as bacteria, rats, or humans, are challenging to study because many factors shape them. Whether a bacterium can eat a certain food or not, what color a rat's fur is, or how tall a child grows up to be, depends on many factors. It is common knowledge that some of these factors are inherent to each organism, such as the genetic material they inherited from their parents, while some factors are inherent to the environment the organism is living in, such as food availability or the surrounding climate. Apart from these factors, there is an inescapable randomness to how organisms turn out, which stems from their internal chemistry and not from any genetic or environmental factor. The internal noise yields differences in organisms, which affects their body and behaviour.

Evolutionary biology is the scientific study of how living beings change in time. There are sources of randomness in the evolutionary process itself. Living beings exist as populations in the wild which have a finite number of individuals and because of it their evolution is subject to random process called genetic drift. Mutations in the genetic material are innately random. Even if organisms themselves wouldn't be noisy, their evolution would be.

The development of the organism, and it's evolutionary past and future, are, therefore, subject to noise and randomness. If we want to understand how organisms are shaped in short or longer timescales, we have to study the effect the inherent randomness of living beings has on their evolution. This inherent randomness is often viewed as a nuisance, but can be exploited for the organism's benefit, as well. In this thesis, I have studied an aspect of the inherent randomness of organisms, expression noise, and investigated in which scenarios it is unfavourable or favourable, and the constraints on its evolvability.

Expression noise is, simply put, a phenomenon that makes differences between organisms, even when they are genetically identical and sharing an identical environment. How does discovering anything about expression noise affect the life of an ordinary, law-abiding, tax-paying citizen, who is financing this research in the first place?

The most obvious benefits are seen in medical applications. Expression noise was shown to be an actor in bacterial drug-resistance (Sánchez-Romero and Casadesús, 2014; Sun et al., 2020), an increasingly worrying problem in medicine. Elevated levels of expression noise have also been found in human cancer cells (Han et al., 2016), indicating a link between deregulation of gene expression noise and cancer development. Less obvious long-term benefits would be in wildlife conservation efforts, as phenotypic variation, which is created by expression noise, has been shown to help species adapt to a wider ecological niche by making them ecological generalists (Draghi, 2020).

Importantly, noise is a basic aspect of all life on Earth, and studying it brings deeper understanding of the fundamental principles of life. Accounting for noise brings us closer to having a better idea of how organisms function and what they are. It shows that noise can be suppressed or elevated for the benefit of the organism.

Lastly, even if there are no imaginable benefits to studying a natural phenomenon, it should not be discarded. Often, the biggest breakthroughs in technology came from "useless" fundamental research, which by definition has no foreseeable application. For example, a long time ago, some curious people have studied rocks that had happened to stick to each other, which set the baseline for some other people in the 17th century to study magnetism for no good reason, which ended up with the harnessing of electricity and yielded all electronic devices we have today, including the computer I am writing this thesis on. A lack of obvious applications of a scientific investigation today does not mean a lack of applications tomorrow, or in the next decade or century. Unfortunately, we cannot predict which research findings will end up having a practical application down the line, which is why it is important to recognize, as Abraham Flexner puts it, the usefulness of useless knowledge (as well as fund it!).

Bibliography

P. Alberch. From genes to phenotype: dynamical systems and evolvability. *Genetica*, 84(1): 5–11, May 1991. ISSN 1573-6857. doi: 10.1007/BF00123979. URL https://doi.org/10.1007/BF00123979.

U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6): 450–461, June 2007. ISSN 1471-0064. doi: 10.1038/nrg2102. URL https://www.nature.com/articles/nrg2102. Number: 6 Publisher: Nature Publishing Group.

R. D. Astumian. Design principles for Brownian molecular machines: how to swim in molasses and walk in a hurricane. *Physical Chemistry Chemical Physics*, 9(37):5067–5083, Sept. 2007. ISSN 1463-9084. doi: 10.1039/B708995C. URL https://pubs.rsc.org/en/content/articlelanding/2007/cp/b708995c. Publisher: The Royal Society of Chemistry.

O. T. Avery, C. M. MacLeod, and M. McCarty. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *Journal of Experimental Medicine*, 79(2):137–158, Feb. 1944. ISSN 0022-1007. doi: 10.1084/jem.79.2.137. URL https://doi.org/10.1084/jem.79.2.137.

R. B. R. Azevedo, R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*, 440(7080): 87–90, Mar. 2006. ISSN 1476-4687. doi: 10.1038/nature04488. URL https://www.nature.com/articles/nature04488. Number: 7080 Publisher: Nature Publishing Group.

E. Azpeitia and A. Wagner. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System. *Frontiers in Molecular Biosciences*, 7:67, Apr. 2020. ISSN 2296-889X. doi: 10.3389/fmolb.2020.00067. URL https://www.frontiersin.org/article/10.3389/fmolb.2020.00067/full.

A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai. Noise in protein expression scales with natural protein abundance. *Nature Genetics*, 38(6):636–643, June 2006. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1807. URL http://www.nature.com/articles/ng1807.

R. Barbuti, R. Gori, P. Milazzo, and L. Nasti. A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models. *Journal of Membrane Computing*, 2(3):207–226, Oct. 2020. ISSN 2523-8914. doi: 10.1007/s41965-020-00046-y. URL https://doi.org/10.1007/s41965-020-00046-y.

G. V. Barroso, N. Puzovic, and J. Y. Dutheil. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics*, 208(1):173–189, Jan. 2018. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.117.300467. URL https://academic.oup.com/genetics/article/208/1/173-189/6066506.

K. Bartoń. MuMIn: Multi-Model Inference, Apr. 2020. URL https://CRAN.R-project.org/package=MuMIn.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67:1–48, Oct. 2015. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL https://doi.org/10.18637/jss.v067.i01.

H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462(7269):90–93, Nov. 2009. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature08504. URL http://www.nature.com/articles/nature08504.

A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, June 2000. ISSN 0028-0836, 1476-4687. doi: 10.1038/35014651. URL http://www.nature.com/articles/35014651.

A. Bergman and M. L. Siegal. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424(6948):549–552, July 2003. ISSN 1476-4687. doi: 10.1038/nature01765. URL https://www.nature.com/articles/nature01765. Number: 6948 Publisher: Nature Publishing Group.

W. J. Blake, M. KÆrn, C. R. Cantor, and J. J. Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–637, Apr. 2003. ISSN 1476-4687. doi: 10.1038/nature01546. URL https://www.nature.com/articles/nature01546. Number: 6932 Publisher: Nature Publishing Group.

M. Bradie. Science and Metaphor. *Biology and Philosophy*, 14(2):159–166, Apr. 1999. ISSN 1572-8404. doi: 10.1023/A:1006601214943. URL https://doi.org/10.1023/A:1006601214943.

D. M. Busiello, S. Suweis, J. Hidalgo, and A. Maritan. Explorability and the origin of network sparsity in living systems. *Scientific Reports*, 7(1):12323, Sept. 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-12521-1. URL https://www.nature.com/articles/s41598-017-12521-1. Number: 1 Publisher: Nature Publishing Group.

Z. Bódi, Z. Farkas, D. Nevozhay, D. Kalapis, V. Lázár, B. Csörgő, Nyerges, B. Szamecz, G. Fekete, B. Papp, H. Araújo, J. L. Oliveira, G. Moura, M. A. S. Santos, T. S. Jr, G. Balázsi, and C. Pál. Phenotypic heterogeneity promotes adaptive evolution. *PLOS Biology*, 15(5):e2000644, May 2017. ISSN 1545-7885. doi: 10.1371/journal.pbio.2000644. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2000644. Publisher: Public Library of Science.

B. Camellato, I. J. Roney, A. Azizi, D. Charlebois, and M. Kaern. Engineered gene networks enable non-genetic drug resistance and enhanced cellular robustness. *Engineering Biology*, 3(4):72–79, 2019. ISSN 2398-6182. doi: 10.1049/enb.2019.0009. URL

https://onlinelibrary.wiley.com/doi/abs/10.1049/enb.2019.0009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1049/enb.2019.0009.

J. N. Carey, E. L. Mettert, M. Roggiani, K. S. Myers, P. J. Kiley, and M. Goulian. Regulated Stochasticity in a Bacterial Signaling Network Permits Tolerance to a Rapid Environmental Change. *Cell*, 173(1):196–207.e14, Mar. 2018. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2018.02.005. URL https://www.cell.com/cell/abstract/S0092-8674(18)30149-1. Publisher: Elsevier.

G. Chalancon, C. N. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, and M. M. Babu. Interplay between gene expression noise and regulatory network architecture. *Trends in Genetics*, 28(5):221–232, May 2012. ISSN 01689525. doi: 10.1016/j.tig.2012.01.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0168952512000157.

M. Chapal, S. Mintzer, S. Brodsky, M. Carmi, and N. Barkai. Resolving noise–control conflict by gene duplication. *PLOS Biology*, 17(11):e3000289, Nov. 2019. ISSN 1545-7885. doi: 10.1371/journal.pbio.3000289. URL https://dx.plos.org/10.1371/journal.pbio.3000289.

D. A. Charlebois. Effect and evolution of gene expression noise on the fitness landscape. *Physical Review E*, 92(2):022713, Aug. 2015. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.92.022713. URL https://link.aps.org/doi/10.1103/PhysRevE.92.022713.

D. A. Charlebois, G. Balázsi, and M. Kærn. Coherent feedforward transcriptional regulatory motifs enhance drug resistance. *Physical Review E*, 89(5):052708, May 2014. doi: 10.1103/PhysRevE.89.052708. URL https://link.aps.org/doi/10.1103/PhysRevE.89.052708. Publisher: American Physical Society.

S. R. Chepyala, Y.-C. Chen, C.-C. S. Yan, C.-Y. D. Lu, Y.-C. Wu, and C.-P. Hsu. Noise propagation with interlinked feed-forward pathways. *Scientific Reports*, 6(1):23607, July 2016. ISSN 2045-2322. doi: 10.1038/srep23607. URL http://www.nature.com/articles/srep23607.

S. Ciliberti, O. C. Martin, and A. Wagner. Innovation and robustness in complex regulatory gene networks. *Proceedings of the National Academy of Sciences*, 104(34):13591–13596, Aug. 2007. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0705396104. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0705396104.

F. H. Crick. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12:138–163, 1958. ISSN 0081-1386.

G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal, complex systems*, 1695(5):1–9, 2006.

M. Delbrück. The Burst Size Distribution in the Growth of Bacterial Viruses (Bacteriophages)1. *Journal of Bacteriology*, 50(2):131–135, Aug. 1945. ISSN 0021-9193. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC374120/.

R. Descartes. *Treatise of Man: French Text with Translation and Commentary, Trans. Thomas Steele Hall.* Cambridge, Mass.: Harvard University Press, 1972.

S. Dey, M. Soltani, and A. Singh. Enhancement of gene expression noise from transcription factor binding to genomic decoy sites. *Scientific Reports*, 10(1):9126, June 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-65750-2. URL https://www.nature.com/articles/s41598-020-65750-2. Number: 1 Publisher: Nature Publishing Group.

J. Draghi. Developmental noise and ecological opportunity across space can release constraints on the evolution of plasticity. *Evolution & Development*, 22(1-2):35–46, 2020. ISSN 1525-142X. doi: 10.1111/ede.12305. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ede.12305. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ede.12305.

S. Dray and A.-B. Dufour. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22:1–20, Sept. 2007. ISSN 1548-7660. doi: 10.18637/jss.v022.i04. URL https://doi.org/10.18637/jss.v022.i04.

Y. Dublanche, K. Michalodimitrakis, N. Kümmerer, M. Foglierini, and L. Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular Systems Biology*, 2(1):41, Jan. 2006. ISSN 1744-4292. doi: 10.1038/msb4100081. URL https://www.embopress.org/doi/full/10.1038/msb4100081. Publisher: John Wiley & Sons, Ltd.

F. Duveau, A. Hodgins-Davis, B. P. Metzger, B. Yang, S. Tryban, E. A. Walker, T. Lybrook, and P. J. Wittkopp. Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. *eLife*, 7:e37272, Aug. 2018. ISSN 2050-084X. doi: 10.7554/eLife.37272. URL https://elifesciences.org/articles/37272.

M. B. Elowitz. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, Aug. 2002. ISSN 00368075, 10959203. doi: 10.1126/science.1070919. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1070919.

C. Espinosa-Soto. On the role of sparseness in the evolution of modularity in gene regulatory networks. *PLOS Computational Biology*, 14(5):e1006172, May 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006172. URL https://dx.plos.org/10.1371/journal.pcbi.1006172.

C. Espinosa-Soto and A. Wagner. Specialization Can Drive the Evolution of Modularity. *PLOS Computational Biology*, 6(3):e1000719, Mar. 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000719. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000719. Publisher: Public Library of Science.

K. S. Farquhar, D. A. Charlebois, M. Szenk, J. Cohen, D. Nevozhay, and G. Balázsi. Role of network-mediated stochasticity in mammalian drug resistance. *Nature Communications*, 10(1):2766, Dec. 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-10330-w. URL http://www.nature.com/articles/s41467-019-10330-w.

J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition, 2019. URL https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biology*, 2(6):e137, Apr. 2004. ISSN 1545-7885. doi: 10.1371/journal.pbio.0020137. URL https://dx.plos.org/10.1371/journal.pbio.0020137.

Y. Gilad, A. Oshlack, and S. A. Rifkin. Natural selection on gene expression. *Trends in Genetics*, 22(8):456–461, Aug. 2006. ISSN 0168-9525. doi: 10.1016/j.tig.2006.06.002. URL https://www.cell.com/trends/genetics/abstract/S0168-9525(06)00170-3. Publisher: Elsevier.

D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, Dec. 1977. ISSN 0022-3654, 1541-5740. doi: 10.1021/j100540a008. URL https://pubs.acs.org/doi/abs/10.1021/j100540a008.

D. T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, July 2000. ISSN 0021-9606. doi: 10.1063/1.481811. URL https://doi.org/10.1063/1.481811.

D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, July 2001. ISSN 0021-9606. doi: 10.1063/1.1378322. URL https://doi.org/10.1063/1.1378322.

A. Goldbeter. Dissipative structures in biological systems: bistability, oscillations, spatial patterns and waves. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2124):20170376, June 2018. doi: 10.1098/rsta.2017.0376. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0376. Publisher: Royal Society.

S. J. Gould and R. C. Lewontin. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 205(1161,):581–598, 1979. URL http://www.jstor.org/stable/77447.

S. F. Greenbury, I. G. Johnston, M. A. Smith, J. P. K. Doye, and A. A. Louis. The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. *Journal of Theoretical Biology*, 267(1):48–61, Nov. 2010. ISSN 0022-5193. doi: 10.1016/j.jtbi.2010.08.006. URL https://www.sciencedirect.com/science/article/pii/S0022519310004054.

A. Grönlund, P. Lötstedt, and J. Elf. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nature Communications*, 4(1):1864, May 2013. ISSN 2041-1723. doi: 10.1038/ncomms2867. URL https://www.nature.com/articles/ncomms2867. Number: 1 Publisher: Nature Publishing Group.

J. B. S. Haldane. A Defense of Beanbag Genetics. *Perspectives in Biology and Medicine*, 7(3):343–360, 1964. ISSN 1529-8795. doi: 10.1353/pbm.1964.0042. URL http://muse.jhu.edu/content/crossref/journals/perspectives_in_biology_and_medicine/v007/7.3.haldane.html.

R. Han, G. Huang, Y. Wang, Y. Xu, Y. Hu, W. Jiang, T. Wang, T. Xiao, and D. Zheng. Increased gene expression noise in human cancers is correlated with low p53 and immune activities as well as late stage cancer. *Oncotarget*, 7(44):72011–72020, Oct. 2016. ISSN 1949-2553. doi: 10.18632/oncotarget.12457. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5342140/.

J. Hausser, A. Mayo, L. Keren, and U. Alon. Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*, 10(1):68, Dec. 2019. ISSN 2041-1723. doi: 10.1038/s41467-018-07391-8. URL http://www.nature.com/articles/s41467-018-07391-8.

C. Hens, U. Harush, S. Haber, R. Cohen, and B. Barzel. Spatiotemporal signal propagation in complex networks. *Nature Physics*, 15(4):403–412, Apr. 2019. ISSN 1745-2481. doi: 10.1038/s41567-018-0409-0. URL https://www.nature.com/articles/s41567-018-0409-0. Number: 4 Publisher: Nature Publishing Group.

A. D. Hershey and M. Chase. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *Journal of General Physiology*, 36(1):39–56, Sept. 1952. ISSN 0022-1295. doi: 10.1085/jgp.36.1.39. URL https://doi.org/10.1085/jgp.36.1.39.

B. Huang, M. Lu, D. Jia, E. Ben-Jacob, H. Levine, and J. N. Onuchic. Interrogating the topological robustness of gene regulatory circuits by randomization. *PLOS Computational Biology*, 13(3):e1005456, Mar. 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005456. URL https://dx.plos.org/10.1371/journal.pcbi.1005456.

E. Huerta-Sanchez and R. Durrett. Wagner's canalization model. *Theoretical Population Biology*, 71(2):121–130, Mar. 2007. ISSN 0040-5809. doi: 10.1016/j.tpb.2006.10.006. URL https://www.sciencedirect.com/science/article/pii/S0040580906001286.

D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of statistical software*, 24(3):nihpa54860, May 2008. ISSN 1548-7660. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/.

C. A. Jackson, D. M. Castro, G.-A. Saldi, R. Bonneau, and D. Gresham. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *eLife*, 9:e51254, Jan. 2020. ISSN 2050-084X. doi: 10.7554/eLife.51254. URL https://elifesciences.org/articles/51254.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer US, New York, NY, 2013. ISBN 978-1-07-161417-4 978-1-07-161418-1. doi: 10.1007/978-1-0716-1418-1. URL https://link.springer.com/10.1007/978-1-0716-1418-1.

E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology*, 23(8):975–982, Aug. 2005. ISSN 1546-1696. doi: 10.1038/nbt1114. URL https://www.nature.com/articles/nbt1114. Number: 8 Publisher: Nature Publishing Group.

A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner. Phase separation provides a mechanism to reduce noise in cells. page 6, 2020.

C. Koch, J. Konieczka, T. Delorey, A. Lyons, A. Socha, K. Davis, S. A. Knaack, D. Thompson, E. K. O'Shea, A. Regev, and S. Roy. Inference and Evolutionary Analysis of Genome-Scale Regulatory Networks in Large Phylogenies. *Cell Systems*, 4(5):543–558.e8, May 2017. ISSN 24054712. doi: 10.1016/j.cels.2017.04.010. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471217301783.

T. Laarits, P. Bordalo, and B. Lemos. Genes under weaker stabilizing selection increase network evolvability and rapid regulatory adaptation to an environmental shift. *Journal of Evolutionary Biology*, 29(8):1602–1616, Aug. 2016. ISSN 1010061X. doi: 10.1111/jeb.12897. URL http://doi.wiley.com/10.1111/jeb.12897.

S. J. Larsen, R. Röttger, H. H. Schmidt, and J. Baumbach. E. coli gene regulatory networks are inconsistent with gene expression data. *Nucleic Acids Research*, 47(1):85–92, Jan. 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1176. URL https://doi.org/10.1093/nar/gky1176.

R. D. Leclerc. Survival of the sparsest: robust gene networks are parsimonious. *Molecular Systems Biology*, 4(1):213, Jan. 2008. ISSN 1744-4292, 1744-4292. doi: 10.1038/msb.2008.52. URL https://onlinelibrary.wiley.com/doi/10.1038/msb.2008.52.

B. Lehner. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular Systems Biology*, 4(1):170, Jan. 2008. ISSN 1744-4292. doi: 10.1038/msb.2008.11. URL https://www.embopress.org/doi/full/10.1038/msb.2008.11. Publisher: John Wiley & Sons, Ltd.

F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101(14):4781–4786, Apr. 2004. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0305937101. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0305937101.

G. A. Linneweber, M. Andriatsilavo, S. B. Dutta, M. Bengochea, L. Hellbruegge, G. Liu, R. K. Ejsmont, A. D. Straw, M. Wernet, P. R. Hiesinger, and B. A. Hassan. A neurodevelopmental origin of behavioral individuality in the Drosophila visual system. *Science*, 367(6482):1112–1119, Mar. 2020. doi: 10.1126/science.aaw7182. URL https://www.science.org/doi/10.1126/science.aaw7182. Publisher: American Association for the Advancement of Science.

M. Liu, S. A. West, and G. A. Cooper. Relatedness and the evolution of mechanisms to divide labor in microorganisms. *Ecology and Evolution*, 11(21):14475–14489, 2021. ISSN 2045-7758. doi: 10.1002/ece3.8067. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.8067. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.8067.

J. C. W. Locke and M. B. Elowitz. Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology*, 7(5):383–392, May 2009. ISSN 1740-1534. doi: 10.1038/nrmicro2056. URL https://www.nature.com/articles/nrmicro2056. Number: 5 Publisher: Nature Publishing Group.

M. Lynch and J. S. Conery. The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, 290(5494):1151–1155, Nov. 2000. doi: 10.1126/science.290.5494.1151. URL https://www.science.org/doi/10.1126/science.290.5494.1151. Publisher: American Association for the Advancement of Science.

H. Maamar, A. Raj, and D. Dubnau. Noise in Gene Expression Determines Cell Fate in Bacillus subtilis. *Science*, 317(5837):526–529, July 2007. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1140818. URL https://science.sciencemag.org/content/317/5837/526. Publisher: American Association for the Advancement of Science Section: Report.

P. C. Maloney and B. Rotman. Distribution of suboptimally induces -D-galactosidase in Escherichia coli. The enzyme content of individual cells. *Journal of Molecular Biology*, 73(1): 77–91, Jan. 1973. ISSN 0022-2836. doi: 10.1016/0022-2836(73)90160-5.

G. Mao, R. Zeng, J. Peng, K. Zuo, Z. Pang, and J. Liu. Reconstructing gene regulatory networks of biological function using differential equations of multilayer perceptrons. *BMC Bioinformatics*, 23(1):503, Nov. 2022. ISSN 1471-2105. doi: 10.1186/s12859-022-05055-5. URL https://doi.org/10.1186/s12859-022-05055-5.

E. Mayr. *Animal Species and Evolution*. Belknap of Harvard University Press, 1963.

G. Mendel. Versuche über Pflanzen-Hybriden" [Experiments Concerning Plant Hybrids]. *Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn]*, IV:3–47, 1865.

P. E. Meyer. infotheo: Information-Theoretic Measures, 2014. URL https://cran.r-project.org/package=infotheo.

P. Mignone, G. Pio, D. D'Elia, and M. Ceci. Exploiting transfer learning for the reconstruction of the human gene regulatory network. *Bioinformatics*, 36(5):1553–1561, Mar. 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz781. URL https://doi.org/10.1093/bioinformatics/btz781.

N. Molina and E. van Nimwegen. The evolution of domain-content in bacterial genomes. *Biology Direct*, 3:51, Dec. 2008. ISSN 1745-6150. doi: 10.1186/1745-6150-3-51.

A. Mugler, M. Kittisopikul, L. Hayden, J. Liu, C. H. Wiggins, G. M. Süel, and A. M. Walczak. Noise Expands the Response Range of the Bacillus subtilis Competence Circuit. *PLOS Computational Biology*, 12(3):e1004793, Mar. 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004793. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004793. Publisher: Public Library of Science.

Nature. Method of the Year 2013. *Nature Methods*, 11(1):1–1, Jan. 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2801. URL https://www.nature.com/articles/nmeth.2801. Number: 1 Publisher: Nature Publishing Group.

D. Nevozhay, R. M. Adams, E. V. Itallie, M. R. Bennett, and G. Balázsi. Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit. *PLOS Computational Biology*, 8(4):e1002480, Apr. 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002480. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002480. Publisher: Public Library of Science.

J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–846, June 2006. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature04785. URL http://www.nature.com/articles/nature04785.

M. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, 2010. ISBN 978-0-19-920665-0. doi: 10.1093/acprof:oso/9780199206650.001.0001. URL https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650.

D. J. Nicholson. OrganismsMachines. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4, Part B):669–678, Dec. 2013. ISSN 1369-8486. doi: 10.1016/j.shpsc.2013.05.014. URL https://www.sciencedirect.com/science/article/pii/S1369848613000824.

D. J. Nicholson. Is the Cell Really a Machine? *Journal of Theoretical Biology*, 477:108–126, 2019.

A. Novick and M. Weiner. ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON*. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7):553–566, July 1957. ISSN 0027-8424. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC528498/.

A. Odorico, E. Rünneburger, and A. Le Rouzic. Modelling the influence of parental effects on gene-network evolution. *Journal of Evolutionary Biology*, 31(5):687–700, 2018. ISSN 1420-9101. doi: 10.1111/jeb.13255. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jeb.13255. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jeb.13255.

S. Ohno. *Evolution by Gene Duplication*. Springer Berlin, Heidelberg, 1970. ISBN 978-0-387-05225-0. Google-Books-ID: sxUDAAAAMAAJ.

P. Oikonomou and P. Cluzel. Effects of topology on network evolution. *Nature Physics*, 2 (8):532–536, Aug. 2006. ISSN 1745-2473, 1745-2481. doi: 10.1038/nphys359. URL http://www.nature.com/articles/nphys359.

C. F. Olson-Manning, M. R. Wagner, and T. Mitchell-Olds. Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, 13(12):867–877, Dec. 2012. ISSN 1471-0064. doi: 10.1038/nrg3322. URL https://www.nature.com/articles/nrg3322. Number: 12 Publisher: Nature Publishing Group.

M. Oubounyt, M. L. Elkjaer, T. Laske, A. B. Grønning, M. Moeller, and J. Baumbach. De-novo reconstruction and identification of transcriptional gene regulatory network modules differentiating single-cell clusters. *NAR Genomics and Bioinformatics*, 5(1):lqad018, Mar. 2023. ISSN 2631-9268. doi: 10.1093/nargab/lqad018. URL https://doi.org/10.1093/nargab/lqad018.

J. M. Pedraza. Noise Propagation in Gene Networks. *Science*, 307(5717):1965–1969, Mar. 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1109090. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1109090.

M. Pigliucci. Genotype–phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540):557–566, Feb. 2010. ISSN 0962-8436. doi: 10.1098/rstb.2009.0241. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2817137/.

J. Pinheiro, D. Bates, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. 2022. URL https://CRAN.R-project.org/package=nlme.

R. Pinho, E. Borenstein, and M. W. Feldman. Most Networks in Wagner's Model Are Cycling. *PLoS ONE*, 7(4):e34285, Apr. 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0034285. URL https://dx.plos.org/10.1371/journal.pone.0034285.

R. Pinho, V. Garcia, and M. W. Feldman. Phenotype Accessibility and Noise in Random Threshold Gene Regulatory Networks. *PLOS ONE*, 10(4):e0119972, Apr. 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0119972. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119972. Publisher: Public Library of Science.

J. M. Raser and E. K. O'Shea. Noise in Gene Expression: Origins, Consequences, and Control. *Science (New York, N.Y.)*, 309(5743):2010–2013, Sept. 2005. ISSN 0036-8075. doi: 10.1126/science.1105891. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360161/.

B. Rhoné, J.-T. Brandenburg, and F. Austerlitz. Impact of selection on genes involved in regulatory network: a modelling study. *Journal of Evolutionary Biology*, 24(10):2087–2098, 2011. ISSN 1420-9101. doi: 10.1111/j.1420-9101.2011.02335.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2011.02335.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2011.02335.x.

M. M. Riehle, A. F. Bennett, and A. D. Long. Genetic architecture of thermal adaptation in Escherichia coli. *Proceedings of the National Academy of Sciences*, 98(2):525–530, Jan. 2001. doi: 10.1073/pnas.98.2.525. URL https://www.pnas.org/doi/full/10.1073/pnas.98.2.525. Publisher: Proceedings of the National Academy of Sciences.

G. Rodrigo and M. A. Fares. Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach. *eLife*, 7:e29739, Jan. 2018. ISSN 2050-084X. doi: 10.7554/eLife.29739. URL https://elifesciences.org/articles/29739.

R. Rosen. *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*. Columbia University Press, Oct. 1991. ISBN 978-0-231-07564-0. Pages: 285 Pages.

N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology*, 323(5):785–793, Nov. 2002. ISSN 0022-2836. doi: 10.1016/S0022-2836(02)00994-4. URL https://www.sciencedirect.com/science/article/pii/S0022283602009944.

R. Röttger, U. Rückert, J. Taubert, and J. Baumbach. How Little Do We Actually Know? On the Size of Gene Regulatory Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(5):1293–1300, Sept. 2012. ISSN 1557-9964. doi: 10.1109/TCBB.2012.71. Conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics.

J. M. Schmiedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Blüthgen, D. S. Marks, and A. van Oudenaarden. MicroRNA control of protein expression noise. *Science*, 348(6230):128–132, Apr. 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa1738. URL https://www.sciencemag.org/lookup/doi/10.1126/science.aaa1738.

J. M. Schmiedel, L. B. Carey, and B. Lehner. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nature Communications*, 10(1):3180, Dec. 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-11116-w. URL http://www.nature.com/articles/s41467-019-11116-w.

M. Schmutzer and A. Wagner. Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. *PLOS Computational Biology*, 16(10):e1007727, Oct. 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007727. URL https://dx.plos.org/10.1371/journal.pcbi.1007727.

E. Schrödinger. *What is Life? The Physical Aspect of the Living Cell.* Cambridge University Press, 1944.

E. Sharon, D. van Dijk, Y. Kalma, L. Keren, O. Manor, Z. Yakhini, and E. Segal. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Research*, 24(10):1698–1706, Oct. 2014. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.168773.113. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.168773.113.

H. Shu, J. Zhou, Q. Lian, H. Li, D. Zhao, J. Zeng, and J. Ma. Modeling gene regulatory networks using neural network architectures. *Nature Computational Science*, 1(7):491–501, July 2021. ISSN 2662-8457. doi: 10.1038/s43588-021-00099-8. URL https://www.nature.com/articles/s43588-021-00099-8. Number: 7 Publisher: Nature Publishing Group.

O. M. Sigalova, A. Shaeiri, M. Forneris, E. E. Furlong, and J. B. Zaugg. Predictive features of gene expression variation reveal mechanistic link with differential expression. *Molecular Systems Biology*, 16(8), Aug. 2020. ISSN 1744-4292, 1744-4292. doi: 10.15252/msb.20209539. URL https://onlinelibrary.wiley.com/doi/10.15252/msb.20209539.

J. L. Spudich and D. E. Koshland. Non-genetic individuality: chance in the single cell. *Nature*, 262(5568):467–471, Aug. 1976. ISSN 1476-4687. doi: 10.1038/262467a0. URL https://www.nature.com/articles/262467a0. Number: 5568 Publisher: Nature Publishing Group.

L. J. Steggles, R. Banks, O. Shaw, and A. Wipat. Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics (Oxford, England)*, 23(3):336–343, Feb. 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btl596.

L. Sun, P. Ashcroft, M. Ackermann, and S. Bonhoeffer. Stochastic Gene Expression Influences the Selection of Antibiotic Resistance Mutations. *Molecular Biology and Evolution*, 37(1):58–70, Jan. 2020. ISSN 0737-4038. doi: 10.1093/molbev/msz199. URL https://doi.org/10.1093/molbev/msz199.

M. Sun and J. Zhang. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLOS Genetics*, 15(9):e1008389, Sept. 2019. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008389. URL https://dx.plos.org/10.1371/journal.pgen.1008389.

M. A. Sánchez-Romero and J. Casadesús. Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proceedings of the National Academy of Sciences*, 111(1):355–360, Jan. 2014. doi: 10.1073/pnas.1316084111. URL https://www.pnas.org/doi/10.1073/pnas.1316084111. Publisher: Proceedings of the National Academy of Sciences.

F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, and M. A. Surani. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382, May 2009. ISSN 1548-7105. doi: 10.1038/nmeth.1315.

R. C. Team. R: A Language and Environment for Statistical Computing, 2021. URL https://www.R-project.org/.

M. E. Tsuda and M. Kawata. Evolution of Gene Regulatory Networks by Fluctuating Selection and Intrinsic Constraints. *PLOS Computational Biology*, 6(8):e1000873, Aug. 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000873. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000873. Publisher: Public Library of Science.

A. Urchueguía, L. Galbusera, D. Chauvin, G. Bellement, T. Julou, and E. v. Nimwegen. Genome-wide gene expression noise in Escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network. *PLOS Biology*, 19(12):e3001491, Dec. 2021. ISSN 1545-7885. doi: 10.1371/journal.pbio.3001491. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001491. Publisher: Public Library of Science.

J.-W. Veening, O. A. Igoshin, R. T. Eijlander, R. Nijland, L. W. Hamoen, and O. P. Kuipers. Transient heterogeneity in extracellular protease production by Bacillus subtilis. *Molecular Systems Biology*, 4:184, 2008. ISSN 1744-4292. doi: 10.1038/msb.2008.18.

E. Ventre, U. Herbach, T. Espinasse, G. Benoit, and O. Gandrillon. One model fits all: Combining inference and simulation of gene regulatory networks. *PLOS Computational Biology*, 19(3):e1010962, Mar. 2023. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1010962. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010962. Publisher: Public Library of Science.

B. Verd, N. A. Monk, and J. Jaeger. Modularity, criticality, and evolvability of a developmental gene regulatory network. *eLife*, 8:e42832, June 2019. ISSN 2050-084X. doi: 10.7554/eLife.42832. URL https://elifesciences.org/articles/42832.

M. Vlková and O. K. Silander. Gene regulation in Escherichia coli is commonly selected for both high plasticity and low noise. *Nature Ecology & Evolution*, pages 1–15, June 2022. ISSN 2397-334X. doi: 10.1038/s41559-022-01783-2. URL https://www.nature.com/articles/s41559-022-01783-2. Publisher: Nature Publishing Group.

A. Wagner. Does Evolutionary Plasticity Evolve? *Evolution*, 50(3):1008–1023, 1996. ISSN 1558-5646. doi: 10.1111/j.1558-5646.1996.tb02342.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.1996.tb02342.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.1996.tb02342.x.

Z. Wang and J. Zhang. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proceedings of the National Academy of Sciences*, 108(16):E67–E76, Apr. 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1100059108. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1100059108.

J. D. Watson and F. H. C. Crick. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738, Apr. 1953. ISSN 1476-4687. doi: 10.1038/171737a0. URL https://www.nature.com/articles/171737a0. Number: 4356 Publisher: Nature Publishing Group.

S. Wright. EVOLUTION IN MENDELIAN POPULATIONS. *Genetics*, 16(2):97–159, Mar. 1931. ISSN 1943-2631. doi: 10.1093/genetics/16.2.97. URL https://doi.org/10.1093/genetics/16.2.97.

S. Wright. The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *Proceedings of the Sixth International Congress on Genetics*, 1:356–366, 1932. URL https://bibbase.org/network/publication/wright-therolesofmutationinbreedingcrossbreedingandselectioninevolution-1932.

R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, and J. B. Hogenesch. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences*, 111(45):16219–16224, Nov. 2014. doi: 10.1073/pnas.1408886111. URL https://www.pnas.org/doi/10.1073/pnas.1408886111. Publisher: Proceedings of the National Academy of Sciences.

List of manuscripts

Peer-reviewed publications

Chapter 2:
Puzović, N., Madaan, T. Dutheil, J. Y. (2023). Being noisy in a crowd: Differential selective pressure on gene expression noise in model gene regulatory networks. PLOS Computational Biology, doi:10.1101/2022.08.01.502352

Author contributions

Thesis title:
Living with noise: The evolution of gene expression noise in gene regulatory networks

Chapter 2:
Nataša Puzović and Julien Y. Dutheil were responsible for the conceptualization, investigation, methodology, and writing (review & editing). Nataša Puzović and Tanvi Madaan were responsible for the formal analysis and visualization. Nataša Puzović was responsible for the software development, validation, and writing (original draft). Julien Y. Dutheil was responsible for supervision.

Chapter 3:
Nataša Puzović and Julien Y. Dutheil were responsible for the conceptualization, investigation, and methodology. Nataša Puzović was responsible for the formal analysis, visualization, software development, validation, and writing (original draft). Julien Y. Dutheil was responsible for supervision and writing (review & editing).

Place and date

Signature (Nataša Puzović)

Signature (Julien Yann Dutheil)

Acknowledgements

First and foremost, I would like to thank my incredible supervisor, Julien Dutheil. For his guidance, endless patience, discussions and everything he has taught me over the years, which ranges from linear models in statistics, career advice, to always reminding me that research is fun. His enthusiasm for science and kind demeanor were a tonic when times were difficult and becoming cynical about academia was easy. Thank you for the advice and the example you set, which I will try to follow and emulate wherever I go.

I would like to thank members of my Thesis Advisory Committee, Arne Traulsen and Tal Dagan, for the helpful advice and fruitful discussion over the past three years. And another thanks to Tal for agreeing to be an examiner of this thesis. Thank you to Jan Baumbach for agreeing to be a member of the defense committee, and Matthias Leippe for being the chair of the defense.

I would also like to thank members of the Molecular Systems Evolution research group: Fernanda Trancoso, Jinyang Liang, Diyar Hamidi, and all the past members, for the nice atmosphere in the lab. A special thanks to Derk Wachsmuth and Angela Donner, for helping with the nitty-gritty and always being cheerful. Thanks to Tanvi Madaan and Barbara D'Albis, who I had the honor of supervising during their internships, and from whom I learned through teaching. Thanks to Nikhil Sharma, who I had the pleasure of working with, and would happily do again in the future. Thanks to Guy Reeves for the stochastic and eclectic, but always captivating, discussions. Thank you to Diethard Tautz, for creating a great working environment in the department, and the Max Planck Society.

A special thanks to the Plönies - Andrea Bours, Artemis Efstratiou and Carolina Peralta, without whom I can't imagine going through this ordeal. Thank you for the countless dinners, daytrips, two-hour coffee breaks, the serious conversations, the random banter, emergency hugs, non-emergency hugs, and for being there for me always. You are a gift. Thanks to my stepbrother Aleksa Čepić, for randomly barging into my office to share nonsense that always cheered me up, and having my back whenever I needed. Thanks to Ana Garoña and Vaibhvi, for adopting me. Thanks to Gustavo Barroso, aka Gman. Whether we are partying, discussing the implementation of networks in genotype-phenotype maps, or nursing emotional wounds over cocktails in Kunkel's, it's always a blast. I also want to thank Bilal Haider, whose endless energy and cheer rubbed off on everyone around him, and Filipa Moutinho, whose endless chill radiated in the office. I hope to see you in conferences in the future. I also want to thank all the students and postdocs at the institute, especially the IMPRS gang, who made working in the office pleasant, beer hours fun and doing a PhD less stressful.

I am thankful to Petnica Science Center, for the first steps in science in general, and the long-term friendships gained there. Thank you to Vladimir Gluhović, my hype man and inspiration. Thanks to Ana Radojičić and Ljubica Mihaljević, for the understanding, emotional support and also the less serious chats throughout the years. Thanks to Igor Asanović, for

the heartfelt conversations and movie recommendations. Thank you to Ana Vasić, Mina Obradović, Tara Vujović and Dunja Ilić, the musketeer gang. Talking to you girls feels like home.

I would like to thank my family: my grandma Gordana, my mother Branka, my father Jovan, my aunt Natasha, and my uncle Jim. Thanks for teaching me what a mitochondrion is when I was a kid, and setting me up to end up here.

Lastly, I would like to thank Danica Despotović, whose friendship I hope I deserve some day. Listing all the things I am thankful for in the more than a decade that we have been friends would be an exercise in futility of the absurd kind and waste of time, time better spent talking until the sunrise and eating some really good food. Though I suspect we are getting too old for staying up that late.

Affidavit

Hiermit erkläre ich, dass:

i) Abgesehen von der Anleitung meines Betreuers sind Inhalt und Gestaltung dieser Doktorarbeit „Leben mit Lärm: Die Entwicklung des Genexpressionsrauschens in genregulatorischen Netzwerken“ ist das Produkt meiner eigenen Arbeit. Alle Zitate aus anderen Werken sowie Paraphrasen oder Zusammenfassungen anderer Werke sind als solche gekennzeichnet und in der Arbeit entsprechend gekennzeichnet.

ii) Diese Arbeit oder Teile davon wurden nicht im Rahmen einer Prüfung oder eines Abschlusses an einer Bildungseinrichtung im In- oder Ausland eingereicht.

iii) Die Erstellung dieser Arbeit unterliegt den Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft.

iv) Ich habe sichergestellt, dass die schriftliche Fassung dieser Arbeit mit der Fassung, welche auf dem beiliegenden Speichermedium zu finden ist, übereinstimmt.

I hereby declare that:

i) Apart from my supervisor's guidance, the content and design of this Ph-D thesis "Living with noise: The evolution of gene expression noise in gene regulatory networks" is the product of my own work. All quotations from other works as well as paraphrases or summaries of other works have been identified as such and properly acknowledged in the thesis.

ii) This thesis or parts thereof have not been submitted to an educational institution in Germany or abroad as part of an examination or degree qualification.

iii) The preparation of this thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation.

iv) I have ensured that the written version of this thesis is identical to the version saved on the enclosed storage medium.

Signature (Nataša Puzović)

Date and location

Appendix A

Supplementary information

A.1 Gene regulatory network model and evolutionary model

A.1.1 Parameters

The parameter values used for the gene regulatory network model and evolutionary simulations and their descriptions are shown in Table A.1.

Table A.1: **Parameters used in the simulations.**

Parameter	Symbol	Value	Description
Number of nodes in the network	n	40	Number of genes in the gene regulatory network
Network density	d	0.05	Proportion of potential connections in the network
Regulatory matrix	$W = (w_{ij})_{1 \le i \le n,\ 1 \le j \le n}$	see Supp. Data	Regulatory relationships in the gene regulatory network
Intrinsic noise	$\{\eta_i^{\text{int}}\}_{1 \le i \le n}$	100	Gene-specific noise of each gene
Basal expression levels	$\{S_i^{basal}\}_{1 \le i \le n}$	$\{20, ..., 20\}$	Constitutive expression level
Number of timesteps for genotype realization	T_r	50	Number of timesteps the expression levels are updated
Minimal expression level	s_{min}	0	Minimal expression level
Maximal expression level	s_{max}	100	Minimal expression level
Number of timesteps to check oscillatory dynamics	τ	10	Time window to apply oscillation criterion
Maximal allowed fluctuation in gene expression levels	ϵ	$1e^{-06}$	Criterion to check oscillatory dynamics
Population size	N	1,000	Number of individuals in a population
Number of generations	T	10,000	Length of the evolutionary simulation in generations
Optimal expression levels (for network establishment)	$\{s_i^{opt}\}_{1 \le i \le n}$	$\{50, ..., 50\}$	Expression levels that correspond to maximum fitness
Mutation rate (regulatory interactions)	μ_w	0.05	Mutation probability of a regulatory interaction, per interaction, per repl. event
Mutation value mean (regulatory interactions)	m_w	0	Mean of normal distribution from which mutation values are drawn
Mutation value variance (regulatory interactions)	v_w	2	Variance of normal distribution from which mutation values are drawn
Selective pressures	$\{\rho_i\}_{1 \le i \le n}$	$\{1, ..., 1\}$	Contribution of expression level to fitness
Mutation rate (intrinsic noise)	μ_η	0.01	Mutation probability of intrinsic noise, per gene, per repl. event
Mutation value mean (intrinsic noise)	m_η	100	Mean of normal distribution from which mutation values are drawn
Mutation value variance (intrinsic noise)	v_η	40	Variance of normal distribution from which mutation values are drawn
Recombination rate	r	0.05	Probability of offspring entering the recombination process

A.1.2 Robustness of network realization

In the simulations performed in this study the gene regulatory network model was realized into the phenotype by synchronously updating the expression levels of all genes in every time step. To test whether the time of updating changes the steady state expression levels, we compared steady state expression levels of 200 samples of 20-gene random networks realized without noise synchronously and asynchronously. Synchronous updating was performed by updating the expression level of all genes in the network at every time step for T_r time steps ($T_r = 50$). Asynchronous updating was performed by randomly choosing a gene in every time step and updating only its expression levels, for $n \times T_r = 1000$ time steps. We realized each network topology 1000 times synchronously and asynchronously and measured the Hamming

distance between the expression level vectors in the last time step. Two expression level values were deemed identical if their difference was less than 0.001. In 92% of cases the synchronous and asynchronous realizations of the same network configuration had a Hamming distance of 0 (Fig A.1A). The mean expression level, expression variance, CV, noise and Fano factor were highly correlated between the synchronous and asynchronous realizations (Fig A.1B-F). Examples of expression level dynamics of deterministic and stochastic realizations with synchronous and asynchronous updating schemes in one network are shown in Fig A.2.

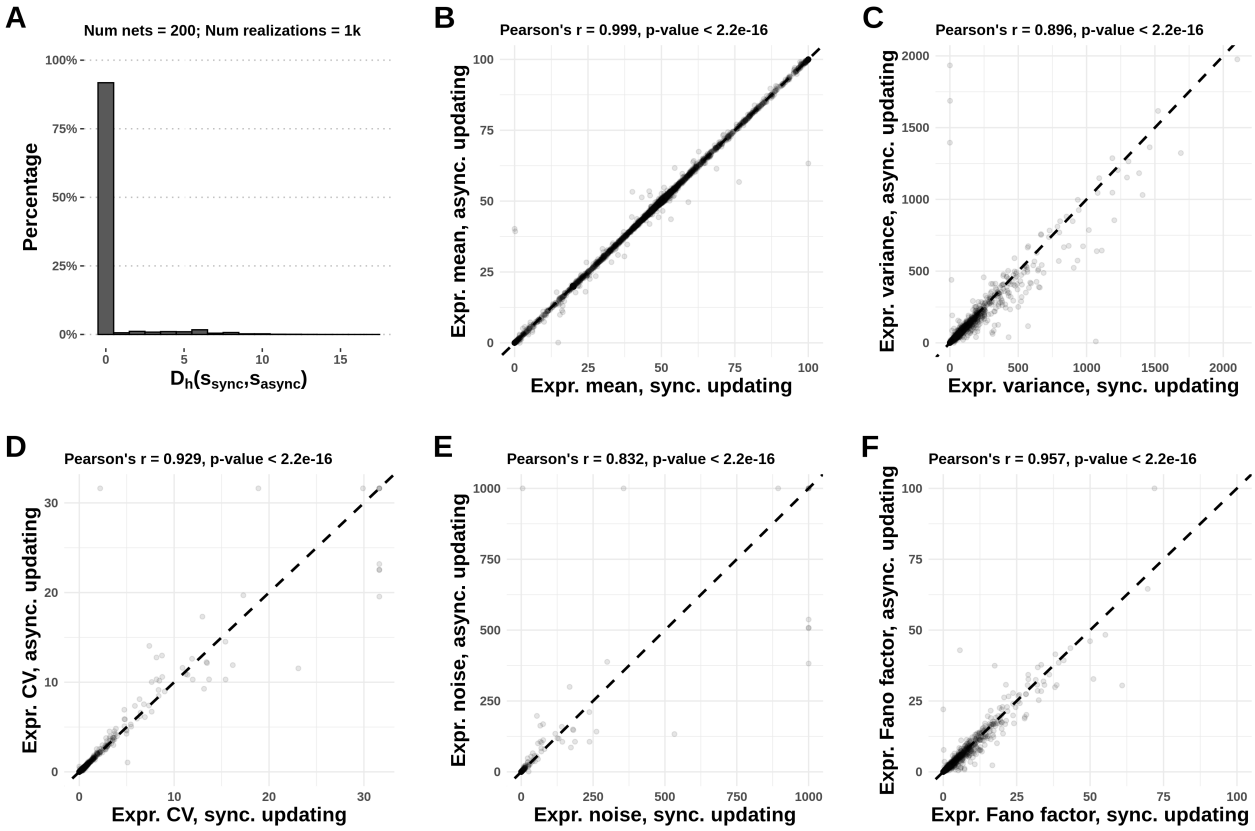

Figure A.1: **Network realization is robust to synchronous or asynchronous expression level updating mode during network realization. A** - Hamming distance between the steady states of synchronously and asynchronously realized networks. Expression level values between synchronous and asynchronous realizations were deemed identical if their difference was less than 0.001. **B** - Mean expression level of populations of synchronously and asynchronously realized networks. **C-F** - Expression level variance, CV, noise and Fano factor of genes from populations of synchronously and asynchronously realized networks.

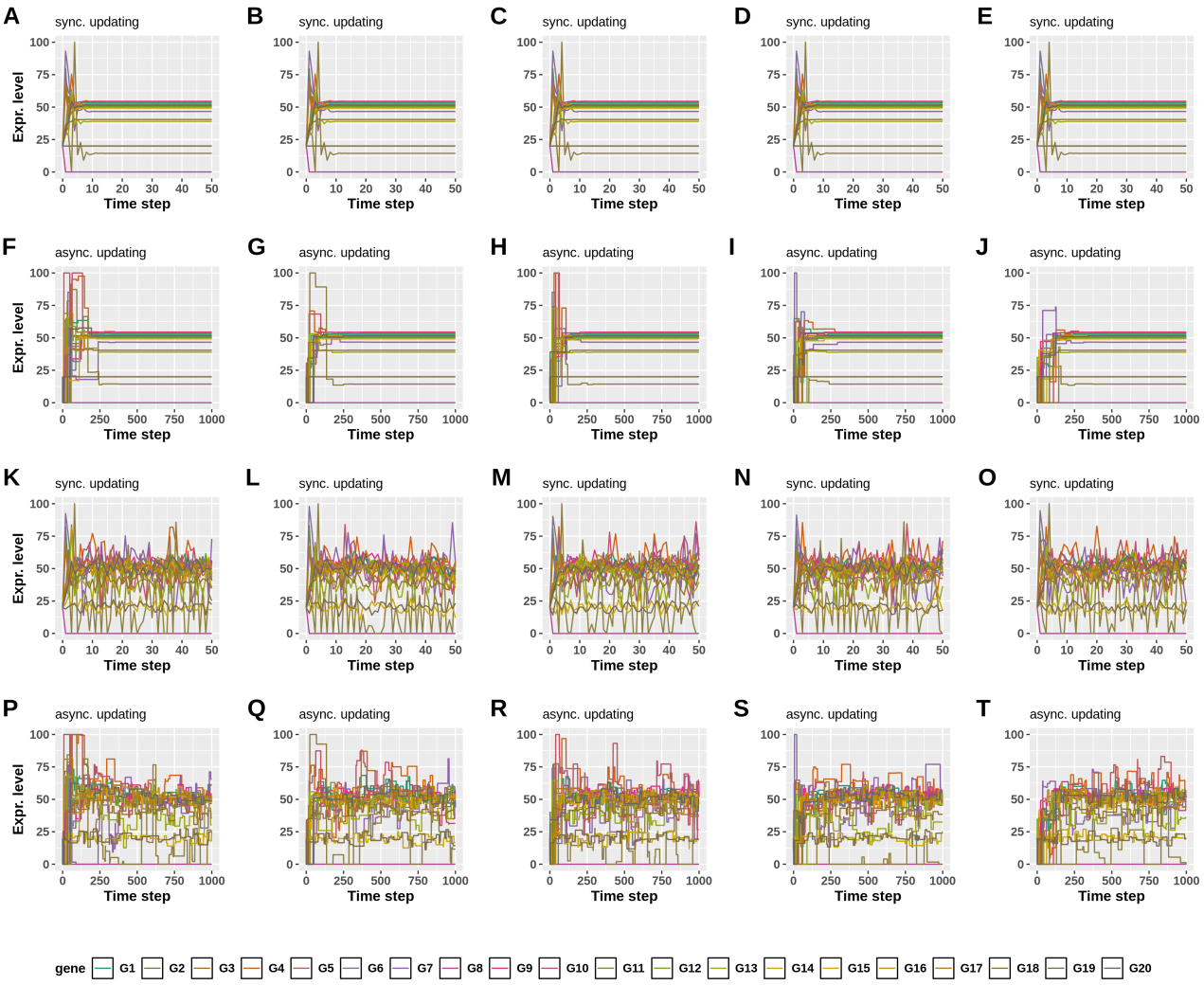

Figure A.2: **Examples of expression level dynamics in realizations of the same network with different expression level updating modes and noise levels. A-E** Five non-noisy realizations with synchronous expression level updating. Since there is no random component in the realization, there are no differences between the realizations. **F-J** Five non-noisy realizations with asynchronous expression level updating. The expression level of a randomly chosen gene is updated in each timestep. Consequently, even though there is no intrinsic expression noise, the dynamics differ between the five realizations, but they reach the same steady state as in the synchronously updated realizations. **K-O** Five noisy realizations with synchronous expression level updating. The mean of the gene expression levels equals the steady state expression levels of the realizations without noise. **P-T** Five noisy realizations with asynchronous expression level updating.

A.1.3 Convergence of expression levels during network establishment

An optimal expression level of $s_{opt} = 50$ is imposed on all genes during network establishment to find network configurations which have intermediate expression levels. At the end of the network establishment process, 68% (54333/80000) of genes in the fittest networks had a steady state expression level in the range of $s_{opt} \pm \frac{s_{opt}}{2}$ (Fig. A.3).

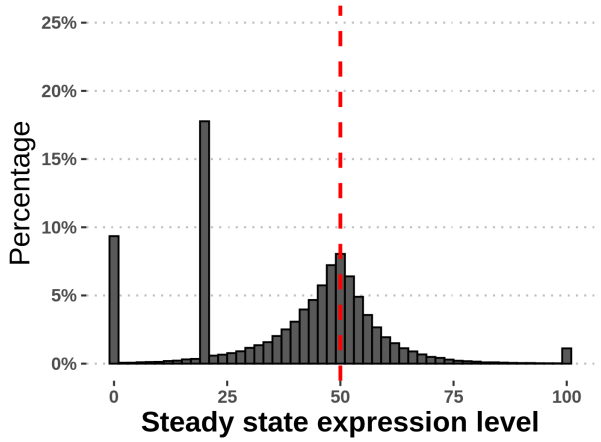

Figure A.3: **Most genes have intermediate steady state expression level after the network establishment process.** Histogram of steady state expression levels after the network establishment process. Dataset consists of 80,000 genes from 2000 random 40-gene network topologies. The peak at $s = 20$ indicates genes which are not activated by other genes and are expressed only at the basal level of $\{S_i^{basal}\}_{1 \le i \le n} = \{20, ..., 20\}$. Red dashed line indicates the optimal expression level.

A.1.4 Population size

We tested the effect of the population size on selective pressure by simulating the evolution of a dataset of 500 network topologies with different population sizes. We find that increasing the population size increases the selective pressure acting on constituent genes (Fig A.4). A population size of 1000 was chosen for the main simulations in this study.

A.1.5 Stability of mean expression level

We imposed stabilizing selection on gene expression levels and observed a repeatable pattern of reduction of gene expression variance. The mean expression level was stable (Fig A.5A) throughout evolution, meaning that the adapted populations had a higher fitness due to a reduction of gene expression variance, not changes in expression mean.

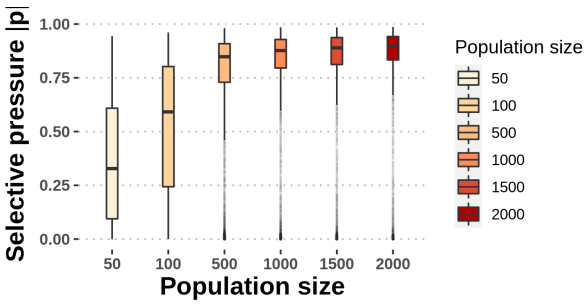

Figure A.4: **Increasing the population size increases the selective pressure on genes under stabilizing selection on gene expression level.** Dataset consists of 20,000 genes from 500 random 40-gene network topologies.

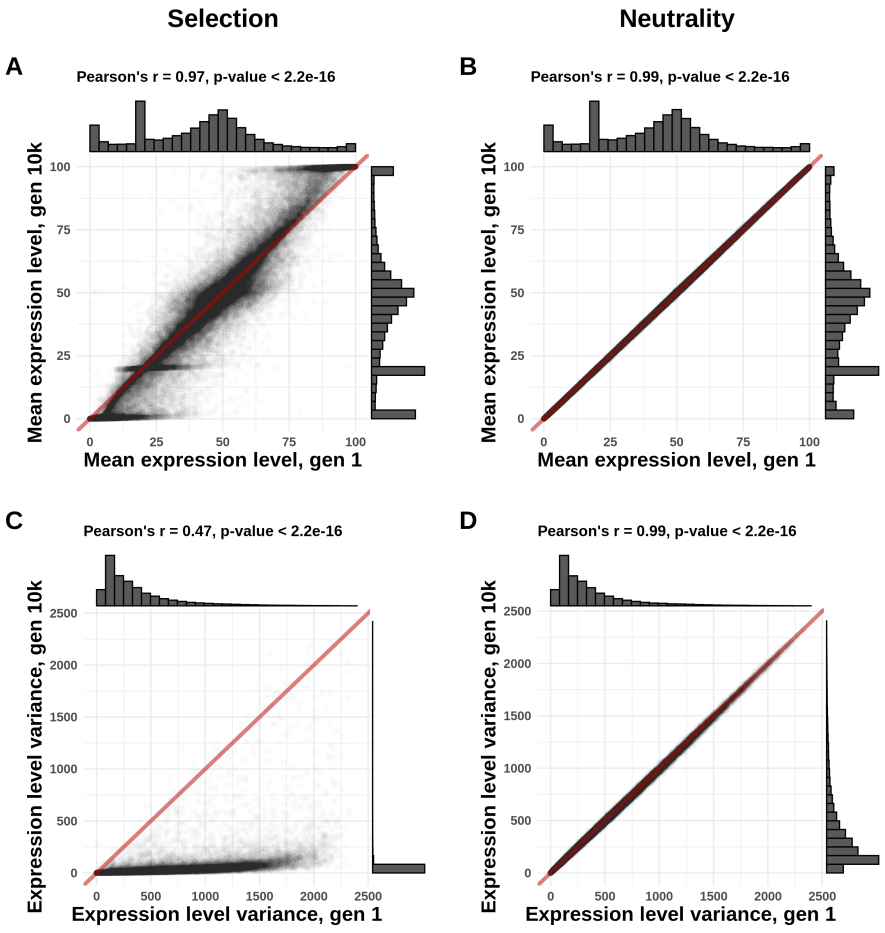

Figure A.5: **Mean expression level does not change after noise evolution under stabilizing selection on gene expression levels. A, B** - Mean expression level in the first and last generation of populations evolved under selection (A) and neutrality (B). **C, D** - Expression variance in the first and last generation of populations evolved under selection (C) and neutrality (D). Red lines indicate lines with a slope of 1.

A.2 Network centrality metrics

To measure the centrality of nodes in the gene networks, we computed 19 node-level centrality measures. These centrality measures are: degree, indegree, outdegree, closeness, betweenness, eigenvector centrality, node strength, instrength, outstrength, hub score, authority including weights, authority excluding weights, absolute node strength, absolute instrength, and absolute outstrength, flow betweenness, load centrality, information centrality, and stress centrality. These measures were heavily intercorrelated and correlated with the expression noise metrics - expression noise, change of expression noise after selection, and selective pressure (Fig A.6). We also computed 12 graph-level centrality measures to study the effects of the global topology on the average selective pressure. These measures are: diameter, mean path distance, degree assortativity, degree centralization, indegree centralization, outdegree centalization, closeness centralization, betweenness centralization, average degree, average indegree, and average outdegree. The global network metrics were intercorrelated, as well (Fig A.7). In the study of the effects of network centrality on evolvability of gene-specific expression noise we focused on instrength and outstrength as node-level centrality measures, and summarized the 12 graph-level measures into two synthetic independent variables using principal component analysis.

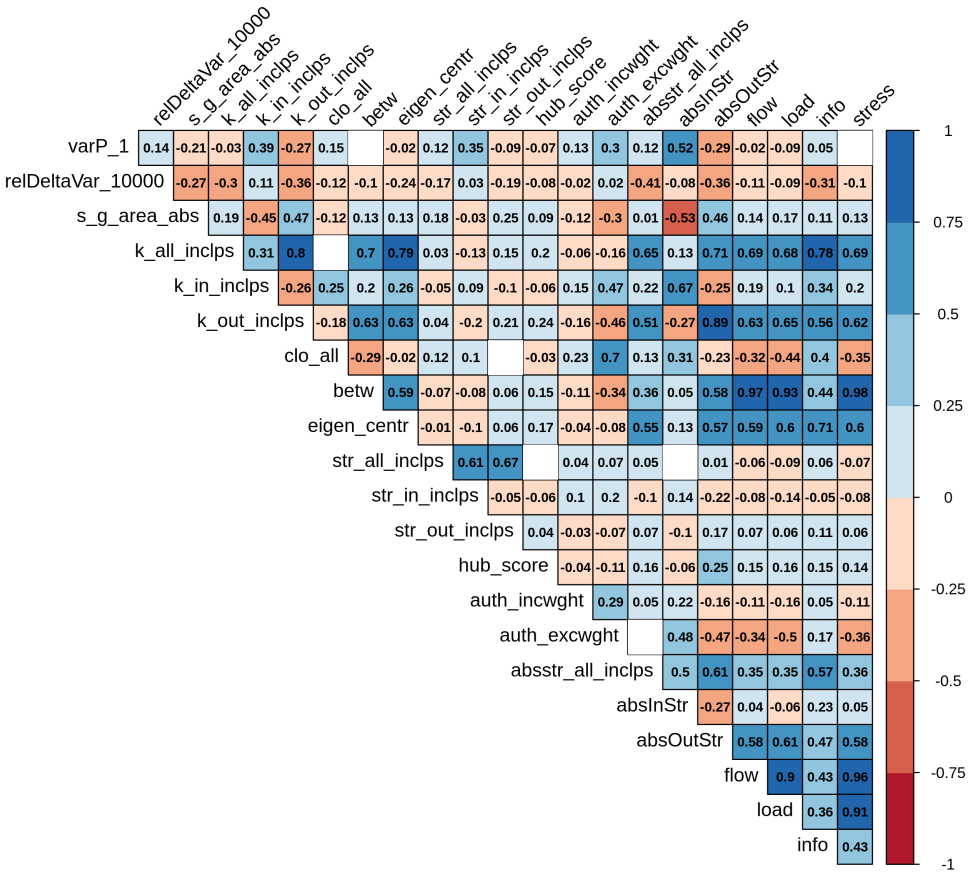

Figure A.6: **Correlation matrix of node-level network centrality metrics and expression noise metrics.** Spearman's rank correlation coefficients shown in the cells. Empty cells indicate a non-significant p-value (p-value > 0.05). Abbreviations: varP_1 - expression variance in the first generation; relDeltaVar_10000 - relative change of expression variance between the first and generation 10,000; s_g_area_abs - selective pressure on each node; k_all_inclps - degree; k_in_inclps - indegree; k_out_inclps - outdegree; clo_all - closeness; betw - betweenness; eigen_centr - eigenvector centrality; str_all_inclps - node strength; str_in_inclps - instrength; str_out_inclps - outstrength; hub_score - hub score; auth_incwght - authority including weights; auth_excwght - authority excluding weights; absstr_all_inclps - absolute strength; absInStr - absolute instrength; absOutStr - absolute outstrength; flow - flow betweenness; load - load centrality; info - information centrality; stress - stress centrality. Dataset consists of 148,886 genes from 2,000 random network topologes.

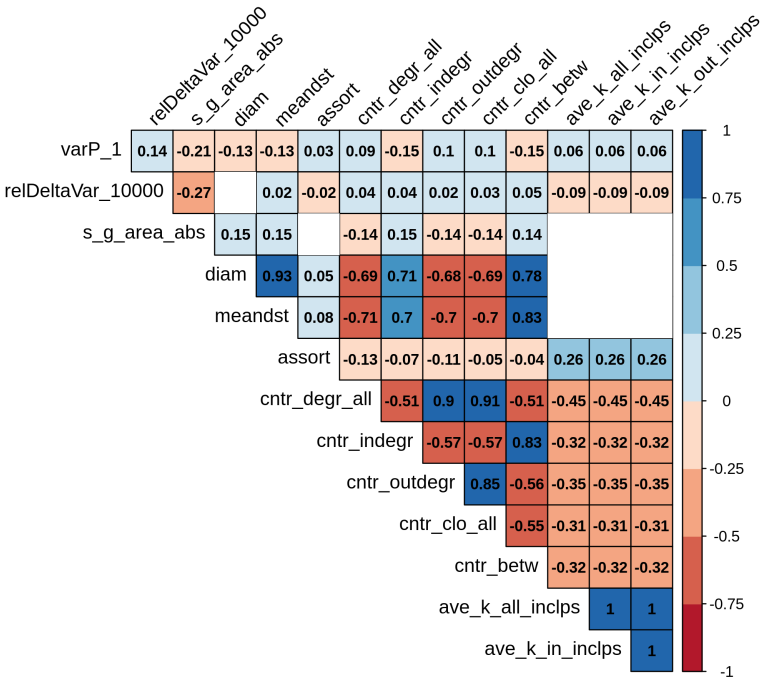

Figure A.7: **Correlation matrix of graph-level network centrality metrics and expression noise metrics.** Spearman's rank correlation coefficients shown in the cells. Empty cells indicate a non-significant p-value (p-value > 0.05). Abbreviations: varP_1 - expression variance in the first generation; relDeltaVar_10000 - relative change of expression variance between the first and generation 10,000; s_g_area_abs - selective pressure on each node; diam - diameter; meandst - mean path distance; assort - degree assortativity; cntr_degr_all - degree centralization; cntr_indegr - indegree centralization; cntr_outdegr - outdegree centralization; cntr_clo_all - closeness centralization; cntr_betw - betweenness centralization; ave_k_all_inclps - average degree; ave_k_in_inclps - average indegree; ave_k_out_inclps - average outdegree. Dataset consists of 148,886 genes from 2,000 random network topologes..

A.2.1 Colinearity between instrength and outstrength

The predictor variables used in statistical modelling in the main results, node instrength and outstrength, were correlated (Spearman's ρ = -0.17, p-value < 2.2×10^{-16}, Fig A.8A-B). This correlation is due to the distributions of in and out nodes being non independent: the more in-connections has, the less out-connections. We also observed that a part of the residuals non-normality may be due to points with a value of zero in one of the two predictor variables. As a control, we rerun the entire analysis on two additional filtered datasets. In the first one, we kept only genes with zero values of either instrength or outstrength, *i.e.* this dataset consisted of only regulators and target genes. Instrength and outstrength were more correlated in the first filtered dataset (Spearman's ρ = -0.86, p-value < 2.2×10^{-16}, Fig A.8C) than in the unfiltered dataset. In the second filtered dataset, we removed all genes that had a zero value of either instrength or outstrength, *i.e.* this dataset consisted of genes that are both regulators and regulated. Instrength and outstrength were less correlated in the second filtered dataset (Spearman's ρ = -0.03, p-value < 2.2×10^{-16}, Fig A.8C) than in the unfiltered dataset, and this filtering somewhat reduced the heteroskedasticity of the Pearson's residuals in the statistical models. The same pattern of effects and significance of instrength and outstrength was observed in the filtered datasets as in the main dataset, indicating that our conclusions are robust to the heteroskedasticity of Pearson's residuals and collinearity between the explanatory variables. The results of all statistical models are summarized in Table A.7 in Section 5.

A.2.2 PCA of global network metrics

To investigate the effects of the intercorrelated graph-level network centrality metrics on noise propagation and noise evolution, we performed a principal component analysis (PCA) to construct independent summary variables representing graph-level network centrality metrics. The first two dimensions of the PCA expressed 85.4% of the total data inertia (Fig A.9A), so we chose the first two principal components (PCs) as synthetic explanatory variables in linear mixed-effects models in the main results. The loadings of the first two PCs are shown in Fig A.9B. The loading of the first synthetic variable (PC1) is dominated by negative loadings of diameter and mean path distance, and the centralization measures, namely positive loadings of outdegree and closeness centralization and negative loadings of indegree and betweenness centralization. The loading of the second synthetic variable (PC2) is dominated by the negative loading of the average degree, average indegree and average outdegree measures. For easier interpretation, the sign of the PCs has been switched in the statistical modelling shown in the main text.

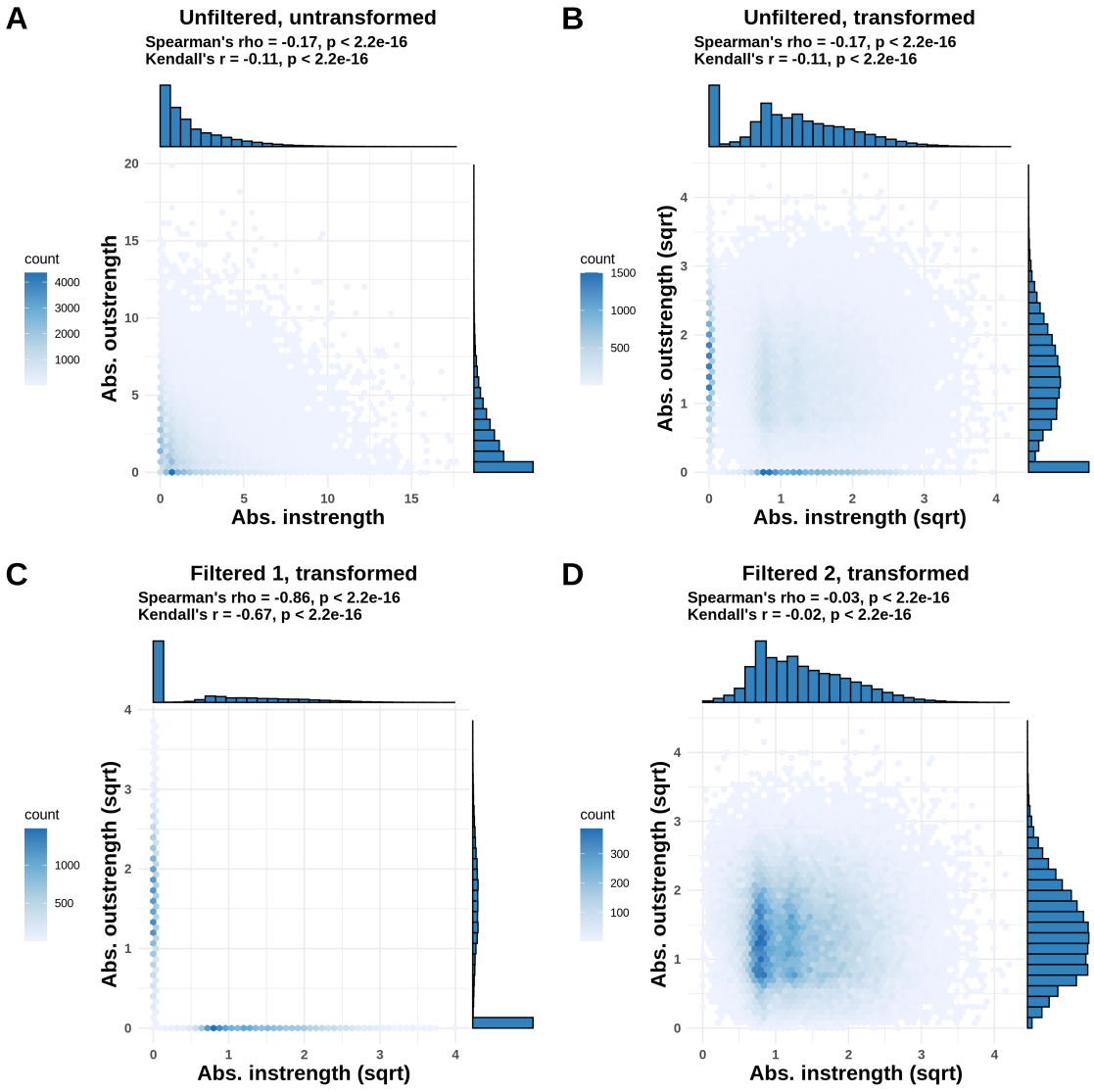

Figure A.8: **Correlations between node instrength and outstrength in unfiltered and filtered datasets.** **A** - Correlation between instrength and outstrength in the unfiltered dataset. Dataset consists of 148,886 genes from 2,000 random network topologies. **B** - Correlation between square-root transformed instrength and outstrength in the unfiltered dataset. **C** - Correlation between square-root transformed instrength and outstrength in the filtered dataset. Dataset consists of 43,214 genes from 2,000 random network topologies. **D** - Correlation between square-root transformed instrength and outstrength in the filtered dataset. The dataset consists of 105,672 genes from 2,000 random network topologies.

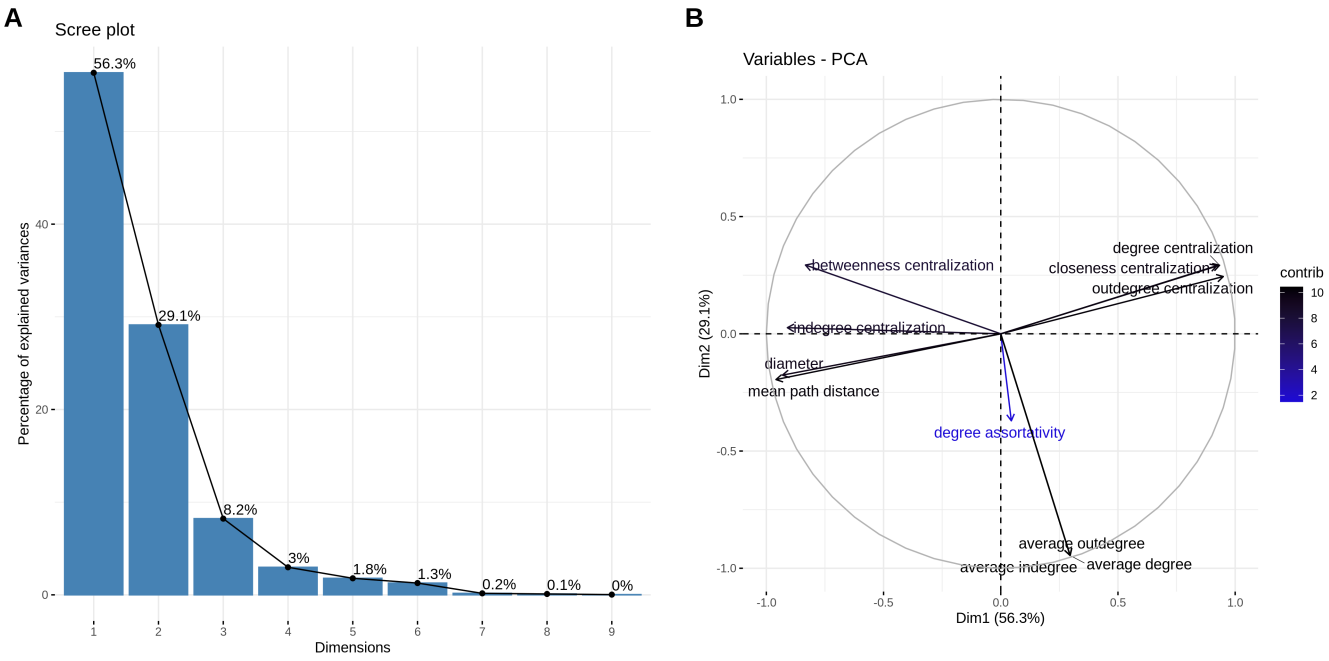

Figure A.9: **Principal component analysis of the graph-level network centrality metrics. A** - Scree plot depicting the percentage of total variance explained by each principal component. The first two principal components express 85.4% of the total inertia. **B** - Correlation circle showing the loadings of the first two principal components.

A.3 Diagnostics of statistical models

A.3.1 GLMM: Noise propagation

To investigate whether noise propagation is captured by our gene regulatory network model, we fitted a linear mixed-effects model with the following formula:

$$y = X\beta + Zu + \epsilon, \tag{A.1}$$

where the outcome variable, y, is a column vector of the expression variance of each node; X is a matrix of two explanatory fixed-effects variables, node instrength and node outstrength, β is a column vector of the two fixed-effects coefficients; Z is the column vector for the design of the random effect variable, network topology sample, and the number of groups equivalent to the number of network topology samples; u is a column vector with the random-effects coefficient for each group (network topology sample); ϵ is a column vector with the residuals. When fitting a model with the assumption of constant variance, the Pearson's residuals were heteroskedastic (Fig A.10A). We fitted models with different variance structures and based on Akaike's Information Criterion chose the model with the exponential function of the node instrength as the variance structure. Pearson's residuals of the chosen and all other fitted models are shown in Fig A.10B-E. Changing the variance structure did not change the significance or the effect of the fixed variables (Table A.2). The variance inflation factor (VIF), a measure of collinearity of explanatory variables, was 1.08. A VIF value lower than 3 indicates that the statistical significance of the inferred effects is reliable in spite of collinearity (James et al., 2013).

A.3.2 GLMM: Relative change of expression variance

To investigate whether local network centrality measures affect the evolution of expression variance, we fitted a linear mixed-effects model with the same formula as Eq. 1. with the normalized change of expression variance as the outcome variable. When fitting a model with the assumption of constant variance, the Pearson's residuals were heteroskedastic (Fig A.11A). We fitted models with different variance structures and based on Akaike's Information Criterion chose the model with the exponential function of the node abstrength as the variance structure. Pearson's residuals of the chosen and all other fitted models are shown in Fig A.11B-E. Changing the variance structure did not change the significance or the effect of the fixed variables (Table A.3). The variance inflation factor (VIF), a measure of collinearity of explanatory variables, was 1.06. A VIF value lower than 3 indicates that the statistical significance of the inferred effects is reliable in spite of collinearity.

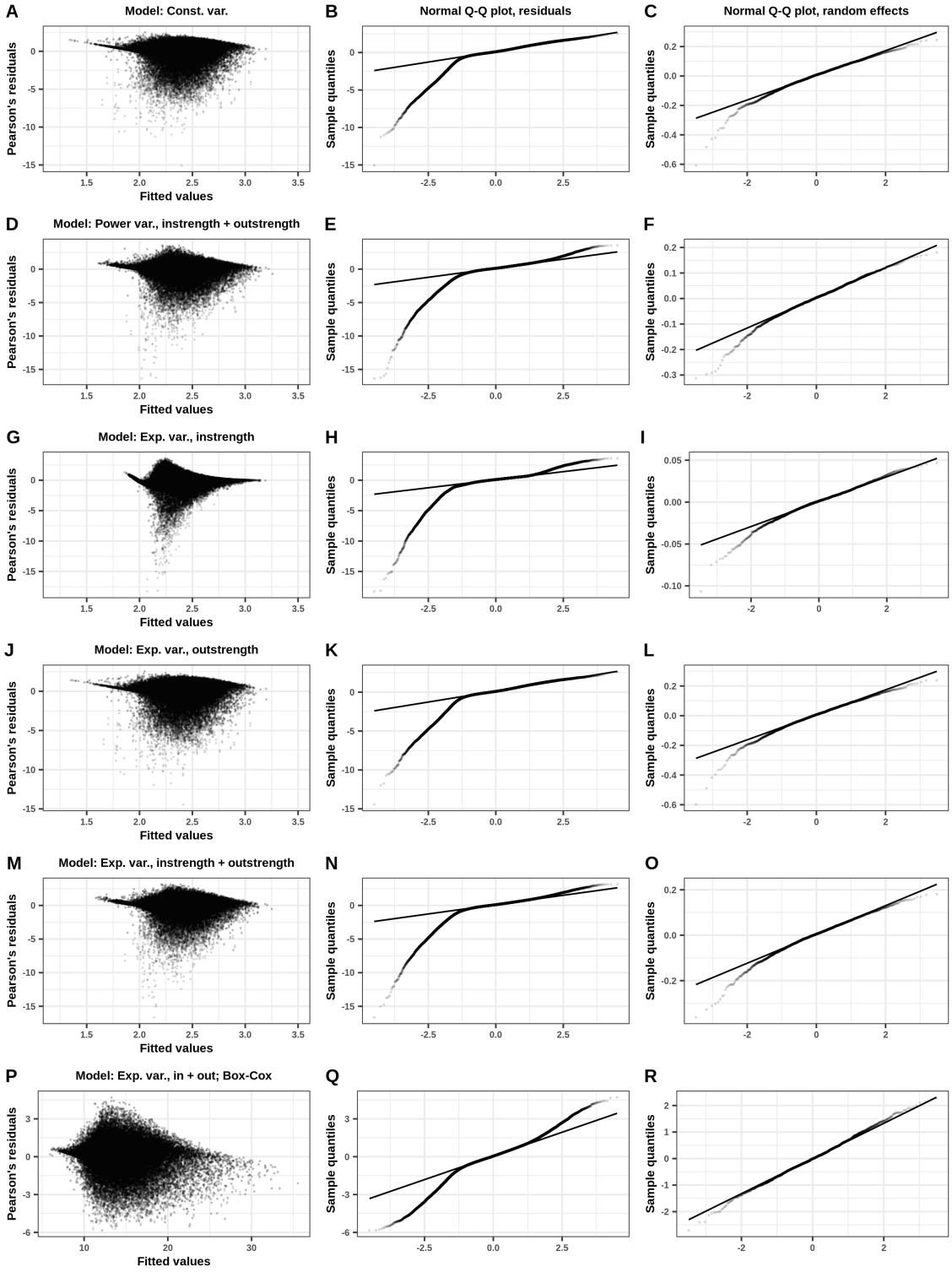

Figure A.10: **Diagnostics of linear mixed-effects models with expression variance as a response variable and different variance structures.**

Figure A.10: Continued. **A-C** Plot of Pearson's residuals vs. fitted values (A), Q-Q plot of Pearson's residuals (B), Q-Q plot of random effects (C) of a model with no variance structure. **D-F** Pearson's residuals vs. fitted values (D), Q-Q plot of standardized Pearson's residuals (E), Q-Q plot of random effects (C) of a model with a variance structure modelled as a power function of instrength and outstrength. **G-I** Pearson's residuals vs. fitted values (G), Q-Q plot of standardized Pearson's residuals (H), Q-Q plot of random effects (I) of a model with a variance structure modelled as an exponential function of instrength. **J-L** Pearson's residuals vs. fitted values (J), Q-Q plot of standardized Pearson's residuals (K), Q-Q plot of random effects (L) of a model with a variance structure modelled as an exponential function of outstrength. **M-O** Pearson's residuals vs. fitted values (M), Q-Q plot of standardized Pearson's residuals (N), Q-Q plot of random effects (O) of a model with a variance structure modelled as an exponential function of instrength and outstrength. **P-R** Pearson's residuals vs. fitted values (P), Q-Q plot of standardized Pearson's residuals (Q), Q-Q plot of random effects (R) of a model with a variance structure modelled as an exponential function of instrength and outstrength and with the explanatory variables transformed with the Box-Cox transform.

Table A.2: **Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with expression variance as a response variable.** The results of models with different variance structures are shown in the table. The effect size differs by a small margin, but the sign and significance remain the same regardless of variance structure. The model with the variance structure as an exponential function of instrength had the lowest Akaike's Information Criterion and was chosen as the best model. Abbreviations: const. var. - constant variance; power var., in + out; variance as a power function of instrength; exp. var., in - variance as an exponential function of instrength; exp. var., out - variance as an exponential function of outstrength; exp. var., in + out - variance as an exponential function of instrength and outstrength; absInStrT_sqrt - absolute instrength, square-root transformed; absOutStrT_sqrt - absolute outstrength, square-root transformed.

Model	Predictors	Value	Std.Error	p.value	p.significant	AIC
const. var.	(Intercept)	2.12252974	0.0037040699	0.000000e+00	✔ Yes	173025.99
const. var.	absInStrT_sqrt	0.23326569	0.0014753454	0.000000e+00	✔ Yes	173025.99
const. var.	absOutStrT_sqrt	-0.07906913	0.0014530464	0.000000e+00	✔ Yes	173025.99
power var., in + out	(Intercept)	2.03654963	0.0024398194	0.000000e+00	✔ Yes	144787.06
power var., in + out	absInStrT_sqrt	0.27234172	0.0012837816	0.000000e+00	✔ Yes	144787.06
power var., in + out	absOutStrT_sqrt	-0.04355256	0.0011948879	1.415734e-289	✔ Yes	144787.06
★ exp. var., in	(Intercept)	2.01421296	0.0014226309	0.000000e+00	✔ Yes	96143.99
★ exp. var., in	absInStrT_sqrt	0.27887716	0.0011158593	0.000000e+00	✔ Yes	96143.99
★ exp. var., in	absOutStrT_sqrt	-0.02214644	0.0007327439	4.731613e-200	✔ Yes	96143.99
exp. var., out	(Intercept)	2.11543714	0.0036669065	0.000000e+00	✔ Yes	172518.76
exp. var., out	absInStrT_sqrt	0.23610694	0.0014788851	0.000000e+00	✔ Yes	172518.76
exp. var., out	absOutStrT_sqrt	-0.07599671	0.0014553914	0.000000e+00	✔ Yes	172518.76
exp. var., in + out	(Intercept)	2.05479175	0.0028207627	0.000000e+00	✔ Yes	144773.85
exp. var., in + out	absInStrT_sqrt	0.26581550	0.0013998517	0.000000e+00	✔ Yes	144773.85
exp. var., in + out	absOutStrT_sqrt	-0.04811173	0.0013049873	3.496844e-296	✔ Yes	144773.85

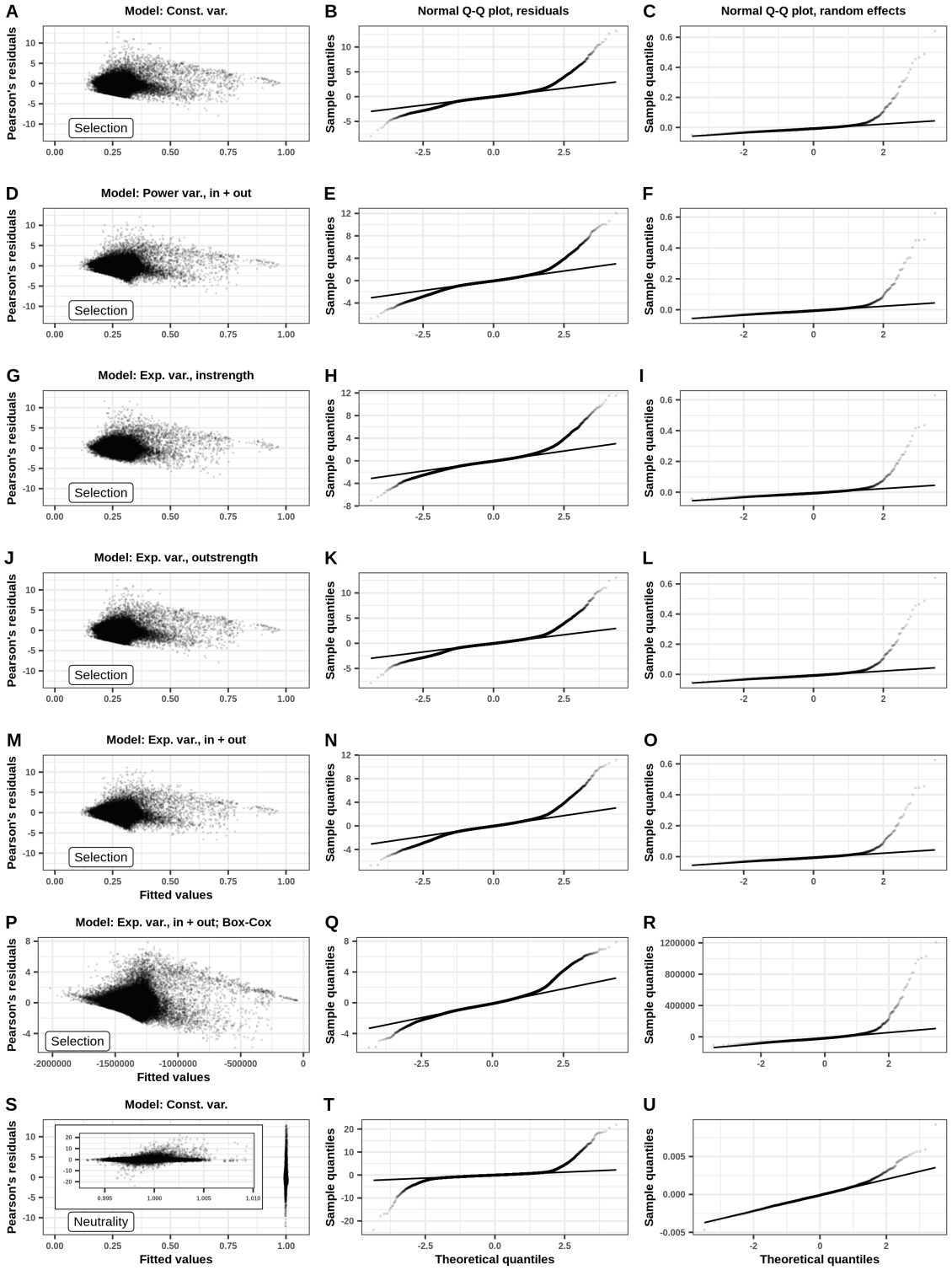

Figure A.11: **Diagnostics of linear mixed-effects models with relative change of expression variance after selection as a response variable and different variance structures.**

Figure A.11: Continued. **A-C** Plot of Pearson's residuals vs. fitted values (A), Q-Q plot of Pearson's residuals (B), Q-Q plot of random effects (C) of a model with no variance structure. **D-F** Pearson's residuals vs. fitted values (D), Q-Q plot of standardized Pearson's residuals (E), Q-Q plot of random effects (C) of a model with a variance structure modelled as a power function of instrength and outstrength. **G-I** Pearson's residuals vs. fitted values (G), Q-Q plot of standardized Pearson's residuals (H), Q-Q plot of random effects (I) of a model with a variance structure modelled as an exponential function of instrength. **J-L** Pearson's residuals vs. fitted values (J), Q-Q plot of standardized Pearson's residuals (K), Q-Q plot of random effects (L) of a model with a variance structure modelled as an exponential function of outstrength. **M-O** Pearson's residuals vs. fitted values (M), Q-Q plot of standardized Pearson's residuals (N), Q-Q plot of random effects (O) of a model with a variance structure modelled as an exponential function of instrength and outstrength. **P-R** Pearson's residuals vs. fitted values (P), Q-Q plot of standardized Pearson's residuals (Q), Q-Q plot of random effects (R) of a model with a variance structure modelled as an exponential function of instrength and outstrength and with the explanatory variables transformed with the Box-Cox transform. **S-U** Pearson's residuals vs. fitted values (M), Q-Q plot of standardized Pearson's residuals (N), Q-Q plot of random effects (O) of a model with constant variance structure fitted on the dataset of populations evolved under neutrality.

Table A.3: **Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with relative change of expression variance after selection as a response variable.** The results of models with different variance structures are shown in the table. The effect size differs by a small margin, but the sign and significance remain the same regardless of variance structure. The model with the variance structure as an exponential function of instrength had the lowest Akaike's Information Criterion and was chosen as the best model. Abbreviations: const. var. - constant variance; power var., in + out; variance as a power function of instrength; exp. var., in - variance as an exponential function of instrength; exp. var., out - variance as an exponential function of outstrength; exp. var., in + out - variance as an exponential function of instrength and outstrength; absInStrT_sqrt - absolute instrength, square-root transformed; absOutStrT_sqrt - absolute outstrength, square-root transformed.

Model	Predictors	Value	Std.Error	DF	t.value	p.value	AIC	p.significant
const. var.	(Intercept)	0.337916277	0.0013430352	72373	251.60641	0.000000e+00	-144342.1	✔ Yes
const. var.	absInStrT_sqrt	-0.008901309	0.0004334489	72373	-20.53600	1.898873e-93	-144342.1	✔ Yes
const. var.	absOutStrT_sqrt	-0.039686466	0.0004269809	72373	-92.94669	0.000000e+00	-144342.1	✔ Yes
power var., in + out	(Intercept)	0.353325888	0.0011743341	72373	300.87339	0.000000e+00	-149555.2	✔ Yes
power var., in + out	absInStrT_sqrt	-0.013152971	0.0004173796	72373	-31.51321	1.686444e-216	-149555.2	✔ Yes
power var., in + out	absOutStrT_sqrt	-0.047848893	0.0003958800	72373	-120.86718	0.000000e+00	-149555.2	✔ Yes
★ exp. var., in	(Intercept)	0.339920854	0.0011538478	72373	294.59766	0.000000e+00	-153714.6	✔ Yes
★ exp. var., in	absInStrT_sqrt	-0.002698555	0.0004280085	72373	-6.30491	2.900180e-10	-153714.6	✔ Yes
★ exp. var., in	absOutStrT_sqrt	-0.046068532	0.0003902311	72373	-118.05450	0.000000e+00	-153714.6	✔ Yes
exp. var., out	(Intercept)	0.338768771	0.0013394715	72373	252.91227	0.000000e+00	-144376.3	✔ Yes
exp. var., out	absInStrT_sqrt	-0.009392597	0.0004342054	72373	-21.63169	1.926129e-103	-144376.3	✔ Yes
exp. var., out	absOutStrT_sqrt	-0.039910911	0.0004274997	72373	-93.35893	0.000000e+00	-144376.3	✔ Yes
exp. var., in + out	(Intercept)	0.348060641	0.0012043388	72373	289.00558	0.000000e+00	-150295.3	✔ Yes
exp. var., in + out	absInStrT_sqrt	-0.011395088	0.0004347183	72373	-26.21258	9.703240e-151	-150295.3	✔ Yes
exp. var., in + out	absOutStrT_sqrt	-0.046273751	0.0004127882	72373	-112.10047	0.000000e+00	-150295.3	✔ Yes

A.3.3 GLMM: Selective pressure

To investigate whether local network centrality measures affect the strength of selective pressure acting on genes, we fitted a linear mixed-effects model with the same formula as Eq. 1. with the selective pressure as the outcome variable. When fitting a model with the assumption of constant variance, the Pearson's residuals were heteroskedastic (Fig A.12A). We fitted models with different variance structures and based on Akaike's Information Criterion chose the model with the exponential function of the node abstrength as the variance structure. Pearson's residuals of the chosen and all other fitted models are shown in Fig A.12B-E. Changing the variance structure did not change the significance or the effect of the fixed variables (Table A.4). The variance inflation factor (VIF), a measure of collinearity of explanatory variables, was 1.03. A VIF value lower than 3 indicates that the statistical significance of the inferred effects is reliable in spite of collinearity.

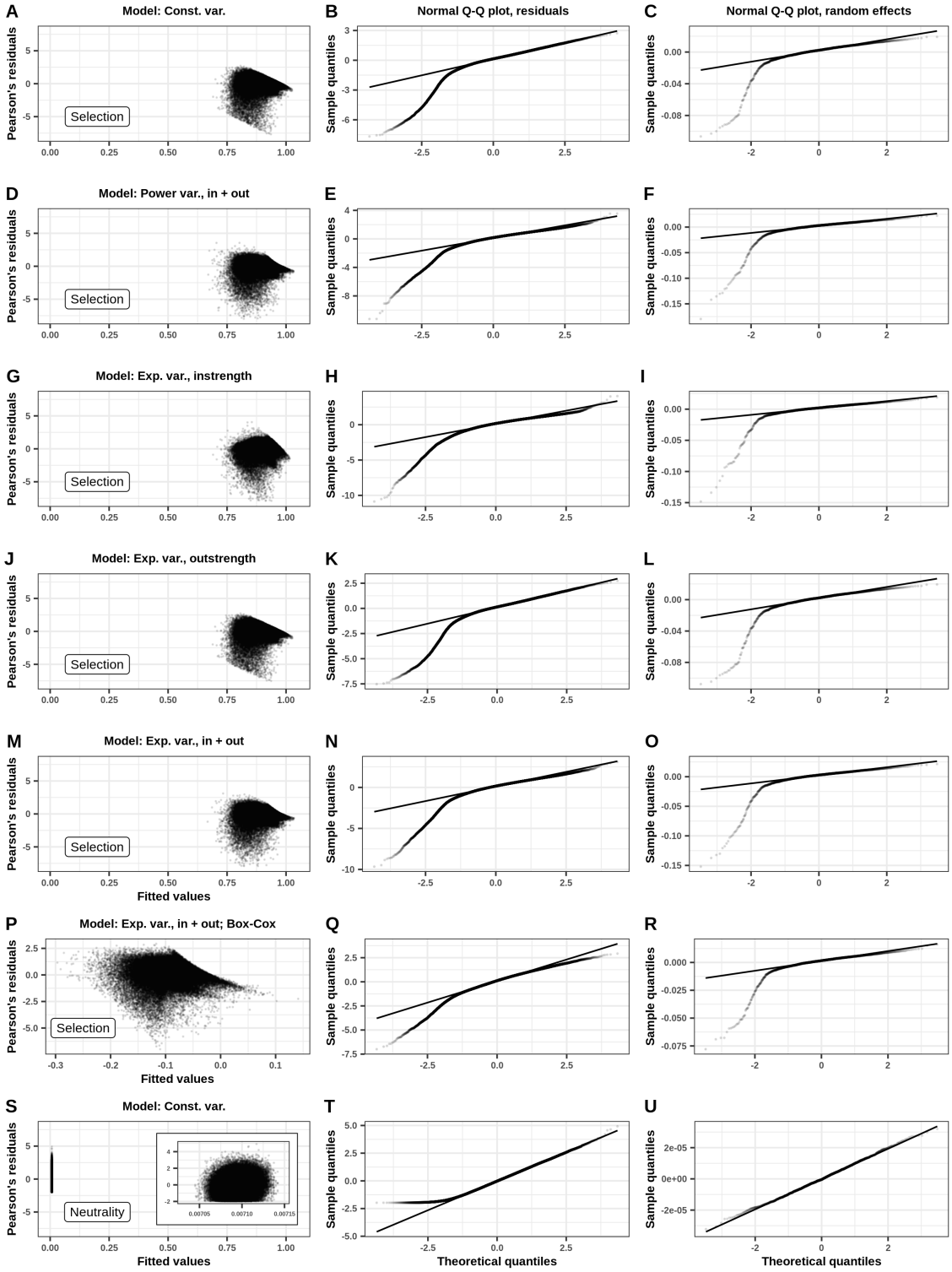

Figure A.12: **Diagnostics of linear mixed-effects models with selective pressure as a response variable and different variance structures.**

Figure A.12: Continued. **A-C** Plot of Pearson's residuals vs. fitted values (A), Q-Q plot of Pearson's residuals (B), Q-Q plot of random effects (C) of a model with no variance structure. **D-F** Pearson's residuals vs. fitted values (D), Q-Q plot of standardized Pearson's residuals (E), Q-Q plot of random effects (C) of a model with a variance structure modelled as a power function of instrength and outstrength. **G-I** Pearson's residuals vs. fitted values (G), Q-Q plot of standardized Pearson's residuals (H), Q-Q plot of random effects (I) of a model with a variance structure modelled as an exponential function of instrength. **J-L** Pearson's residuals vs. fitted values (J), Q-Q plot of standardized Pearson's residuals (K), Q-Q plot of random effects (L) of a model with a variance structure modelled as an exponential function of outstrength. **M-O** Pearson's residuals vs. fitted values (M), Q-Q plot of standardized Pearson's residuals (N), Q-Q plot of random effects (O) of a model with a variance structure modelled as an exponential function of instrength and outstrength. **P-R** Pearson's residuals vs. fitted values (P), Q-Q plot of standardized Pearson's residuals (Q), Q-Q plot of random effects (R) of a model with a variance structure modelled as an exponential function of instrength and outstrength and with the explanatory variables transformed with the Box-Cox transform. **S-U** Pearson's residuals vs. fitted values (M), Q-Q plot of standardized Pearson's residuals (N), Q-Q plot of random effects (O) of a model with constant variance structure fitted on the dataset of populations evolved under neutrality.

Table A.4: **Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with selective pressure as a response variable.** The results of models with different variance structures are shown in the table. The effect size differs by a small margin, but the sign and significance remain the same regardless of variance structure. The model with the variance structure as an exponential function of instrength had the lowest Akaike's Information Criterion and was chosen as the best model. Abbreviations: const. var. - constant variance; power var., in + out; variance as a power function of instrength; exp. var., in - variance as an exponential function of instrength; exp. var., out - variance as an exponential function of outstrength; exp. var., in + out - variance as an exponential function of instrength and outstrength; absInStrT_sqrt - absolute instrength, square-root transformed; absOutStrT_sqrt - absolute outstrength, square-root transformed.

Model	Predictors	Value	Std.Error	DF	t.value	p.value	p.significant	AIC
const. var.	(Intercept)	0.88415564	0.0006748052	63167	1310.2382	0	✔ Yes	-189809.3
const. var.	absInStrT_sqrt	-0.04096498	0.0003106663	63167	-131.8617	0	✔ Yes	-189809.3
const. var.	absOutStrT_sqrt	0.03588198	0.0002908981	63167	123.3490	0	✔ Yes	-189809.3
power var., in + out	(Intercept)	0.87597512	0.0005514233	63167	1588.5710	0	✔ Yes	-199304.3
power var., in + out	absInStrT_sqrt	-0.03786164	0.0002801003	63167	-135.1717	0	✔ Yes	-199304.3
power var., in + out	absOutStrT_sqrt	0.03995255	0.0002456307	63167	162.6529	0	✔ Yes	-199304.3
★ exp. var., in	(Intercept)	0.88278537	0.0005377403	63167	1641.6576	0	✔ Yes	-207009.0
★ exp. var., in	absInStrT_sqrt	-0.03708753	0.0002878319	63167	-128.8513	0	✔ Yes	-207009.0
★ exp. var., in	absOutStrT_sqrt	0.03434249	0.0002328101	63167	147.5129	0	✔ Yes	-207009.0
exp. var., out	(Intercept)	0.88370457	0.0006732199	63167	1312.6537	0	✔ Yes	-189839.4
exp. var., out	absInStrT_sqrt	-0.04101737	0.0003111198	63167	-131.8379	0	✔ Yes	-189839.4
exp. var., out	absOutStrT_sqrt	0.03629719	0.0002910387	63167	124.7160	0	✔ Yes	-189839.4
exp. var., in + out	(Intercept)	0.87761056	0.0005922562	63167	1481.8090	0	✔ Yes	-199570.5
exp. var., in + out	absInStrT_sqrt	-0.03864219	0.0002999411	63167	-128.8326	0	✔ Yes	-199570.5
exp. var., in + out	absOutStrT_sqrt	0.03974518	0.0002648410	63167	150.0719	0	✔ Yes	-199570.5

A.3.4 GLM: Average selective pressure

To investigate whether global network centrality measures affect the average selective pressure acting on genes in networks, we fitted a linear model with the average selective pressure per network as the outcome variable, and the first two principal components as explanatory variables, after performing a principal component analysis on 12 graph-level centrality metrics. Diagnostical plots are shown in Fig A.13.

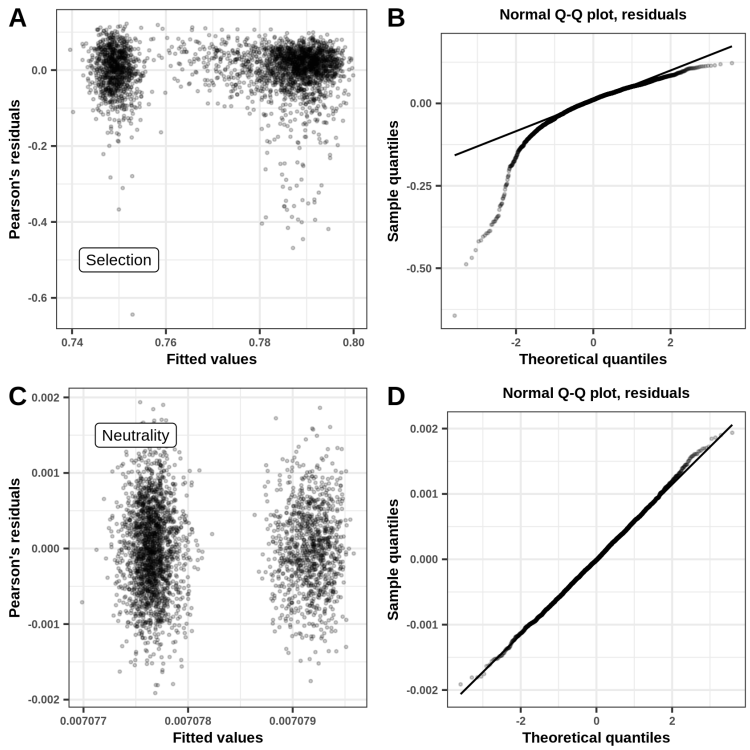

Figure A.13: **Diagnostics of linear model with average selective pressure per network as a response variable. A, B** - Pearson's residuals vs. fitted values (A) and Q-Q plot of standardized Pearson's residuals (B) in the model fitted on selected populations. **C, D** - Pearson's residuals vs. fitted values (C) and Q-Q plot of standardized Pearson's residuals (D) in the model fitted on neutral populations.

A.4 Robustness of results in different topology structures

The analysis of the effects of local network metrics on the evolution of expression noise was performed on a dataset of 2,000 random (Erdős–Rényi) network topologies. To check whether our results hold for other network topology types, we performed the same analysis of two additional datasets: 1,000 scale-free (Barabási–Albert) networks, and 1,000 small-world (Watts–Strogatz model) networks. The results of all generalized linear mixed-effects models and mutual information tests are consistent and summarized in Table A.5.

Table A.5: **The effects and significance of local network centrality metrics are consistent across different topological structures.** The effect size differs by a small margin, but the sign and significance remain the same across different topological structures. Dataset consists of 113,274 genes from 3,000 network topologies.

Response	Topology	Expl. var.	Beta	p-value (GLMM)[1]	MI	p-value (MI)[2]
Expression variance	ER	Instrength	0.28	$< 2.2 \times 10^{-16}$ ***	0.67	10^{-4} ***
		Outstrength	-0.02	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
	BA	Instrength	0.21	$< 2.2 \times 10^{-16}$ ***	0.34	10^{-4} ***
		Outstrength	-0.06	$< 2.2 \times 10^{-16}$ ***	0.11	10^{-4} ***
	WS	Instrength	0.25	$< 2.2 \times 10^{-16}$ ***	0.43	10^{-4} ***
		Outstrength	-0.05	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
Rel. change of expr. variance	ER	Instrength	-0.003	2.9×10^{-10} ***	0.09	10^{-4} ***
		Outstrength	-0.046	$< 2.2 \times 10^{-16}$ ***	0.14	10^{-4} ***
	BA	Instrength	-0.041	$< 2.2 \times 10^{-16}$ ***	0.19	10^{-4} ***
		Outstrength	-0.027	$< 2.2 \times 10^{-16}$ ***	0.26	10^{-4} ***
	WS	Instrength	0.004	$< 2.6 \times 10^{-7}$ ***	0.09	10^{-4} ***
		Outstrength	-0.039	$< 2.2 \times 10^{-16}$ ***	0.08	10^{-4} ***
Probability of responding to selection	ER	Instrength	-1.87	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.08	$< 2.2 \times 10^{-16}$ ***	—	—
	BA	Instrength	-2.01	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.4	$< 2.2 \times 10^{-16}$ ***	—	—
	WS	Instrength	-1.88	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.16	3.9×10^{-11} ***	—	—
Gene-specific selective pressure	ER	Instrength	-0.04	$< 2.2 \times 10^{-16}$ ***	0.19	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.31	10^{-4} ***
	BA	Instrength	-0.02	$< 2.2 \times 10^{-16}$ ***	0.14	10^{-4} ***
		Outstrength	0.02	$< 2.2 \times 10^{-16}$ ***	0.54	10^{-4} ***
	WS	Instrength	-0.04	$< 2.2 \times 10^{-16}$ ***	0.17	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.2	10^{-4} ***

[1] Coefficients and their significance were computed using linear mixed-effects model (see Methods). [2] Mutual information p-values were computed using a Monte Carlo permutation test with 10,000 permutations. Asterisks indicate statistical significance: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

A.5 Robustness of results to unequal fitness contribution of genes

In most simulations performed in this study we assumed for simplicity an equal fitness contribution of all genes, which is not biologically realistic. To check whether our results are robust with different parametrization of fitness contribution, we performed the same analysis of an additional dataset: 500 of the same (Erdős–Rényi) networks used in the main results, but with the values of fitness contribution of all genes $\{\rho_i\}_{1 \le i \le n}$ drawn from an uniform distribution $\mathcal{U}(0, 2)$. The results of all generalized linear mixed-effects models and mutual information tests are consistent and summarized in Table A.6.

Table A.6: **The effects and significance of local network centrality metrics are consistent between assumptions of equal and unequal fitness contributions of genes.** The effect size differs by a small margin, but the sign and significance remain between the two parametrizations of fitness contribution.

Response	FC[2]	Expl. var.	Beta	p-value (GLMM)[1]	MI	p-value (MI)[2]
Expression variance	equal	Instrength	0.28	$< 2.2 \times 10^{-16}$ ***	0.67	10^{-4} ***
		Outstrength	-0.02	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
	unequal	Instrength	0.26	$< 2.2 \times 10^{-16}$ ***	0.67	10^{-4} ***
		Outstrength	-0.02	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
Rel. change of expr. variance	equal	Instrength	-0.003	2.9×10^{-10} ***	0.09	10^{-4} ***
		Outstrength	-0.046	$< 2.2 \times 10^{-16}$ ***	0.14	10^{-4} ***
	unequal	Instrength	-0.006	$< 2.2 \times 10^{-16}$ ***	0.02	10^{-4} ***
		Outstrength	-0.018	$< 2.2 \times 10^{-16}$ ***	0.03	10^{-4} ***
Probability of responding to selection	equal	Instrength	-1.87	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.08	$< 6.67 \times 10^{-7}$ ***	—	—
	unequal	Instrength	-1.96	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.13	$< 2.2 \times 10^{-16}$ ***	—	—
Gene-specific selective pressure	equal	Instrength	-0.04	$< 2.2 \times 10^{-16}$ ***	0.19	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.31	10^{-4} ***
	unequal	Instrength	-0.06	$< 2.2 \times 10^{-16}$ ***	0.07	10^{-4} ***
		Outstrength	0.05	$< 2.2 \times 10^{-16}$ ***	0.06	10^{-4} ***

[1] Coefficients and their significance were computed using linear mixed-effects model (see Methods). [2] Mutual information p-values were computed using a Monte Carlo permutation test with 10,000 permutations. Asterisks indicate statistical significance: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001. [2] Fitness contribution.

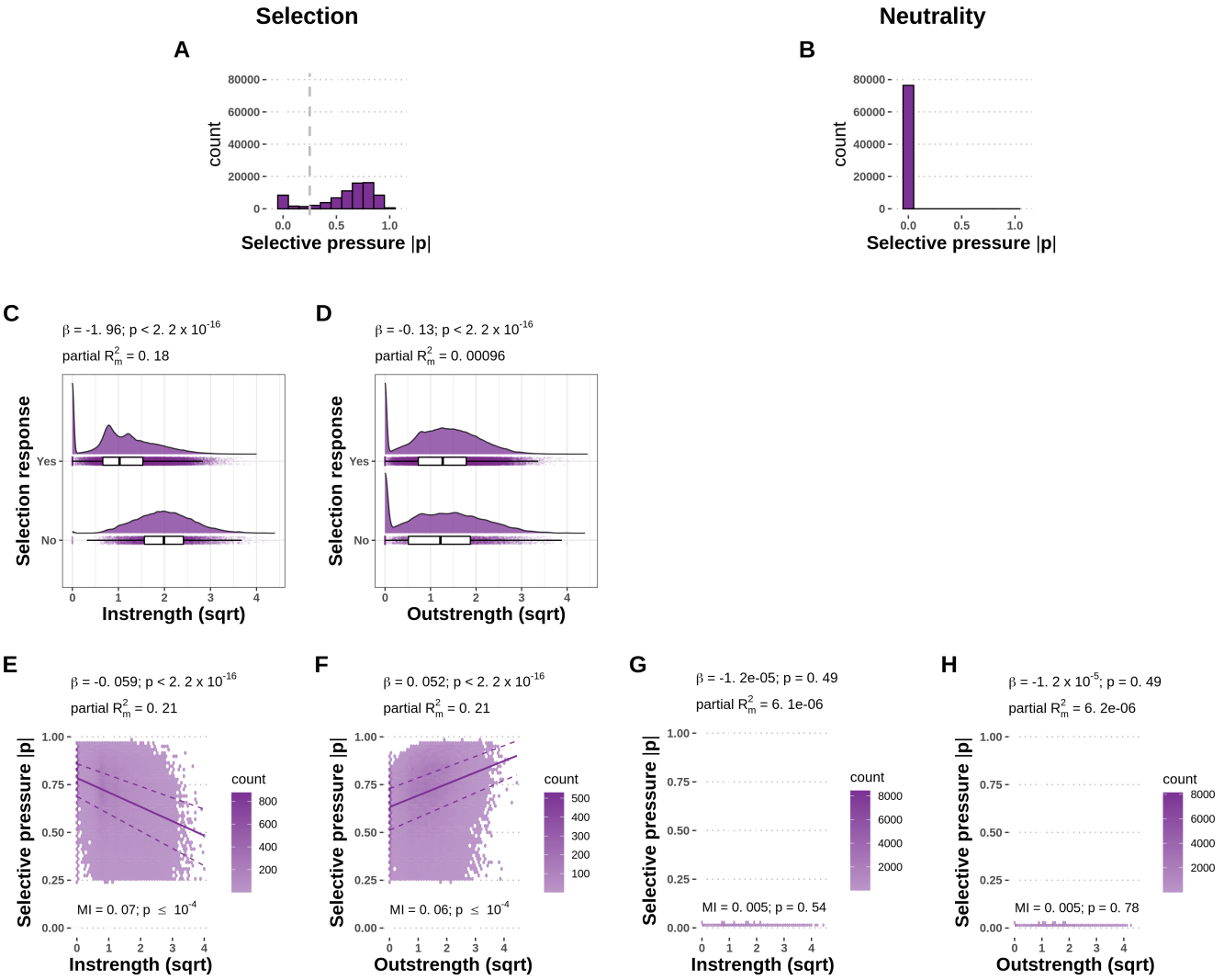

Figure A.14: **Differential selective pressure is acting on genes based on their centrality, regardless of the fitness contribution parametrization. A, B** - Distributions of the measured selective pressure in selected (A) and neutral (B) populations. Genes with a selective pressure above 0.25 were categorized as responsive to selection. **C, D** - High instrength genes are less likely to respond to selection. Absolute instrength (C) has a strong significant negative effect on the probability of selection response. Absolute outstrength (D) has a weak significant negative effect on the probability of selection response. **E, F** - In the subset of genes that responded to selection, high instrength (E) decreases the selective pressure, while high outstrength (F) increases the selective pressure acting on individual genes. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **G, H** - Absolute instrength (G) and outstrength (H) have no significant effect on the selective pressure in the non-selected populations. The dataset consists of 74,443 genes from 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Coefficients, p-values and partial marginal R^2 measures are estimated using logistic regression and linear mixed-effects models with selection responsiveness or selective pressure as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed using 10,000 permutations.

A.6 Filtered datasets

We performed the analyses on two additional datasets to get a clearer picture of the effects of instrength and outstrength on the expression noise metrics. The first filtered dataset is a dataset in which genes that are both regulators and regulated were removed, *i.e.* it consists exclusively of pure regulators (genes that regulate others and are not being regulated) and purely regulated genes (genes that are being regulated and do not regulate other genes). The second filtered dataset is a dataset that consists exclusively of genes that are both regulators and regulated, *i.e.* in this dataset pure regulators and purely regulated genes have been removed. The effects and significance of the two local centrality metrics are consistent in analyses of expression variance, relative change of expression variance and gene-specific selective pressure (Table A.7). In the unfiltered and second filtered dataset we found a significant small negative effect of outstrength on the probability of responding to selection. However, the negative effect is lost when we analysed only the genes that are either regulators or regulated genes (Filtered 1 dataset), and we observed a significant strong positive effect of outstrength on the probability of responding to selection. We concluded that instrength in genes that are both regulators and regulated influences the effect of outstrength on the probability of responding to selection and that there are complex interactions between the two centrality metrics. However, when there are no genes that have both instrength and outstrength the effects are clear and instrength has a strongly negative effect, while outstrength has a strongly positive effect on the probability of a gene to respond to selection.

Table A.7: **Filtered and unfiltered datasets.** Filtered 1 dataset is dataset in which genes that are both regulators and regulated were removed, *i.e.* it consists exclusively of pure regulators (genes that regulate others and are not being regulated) and purely regulated genes (genes that are being regulated and do not regulate other genes). Filtered 1 dataset consists of 43,214 genes from 2,000 random network topologes. Unfiltered dataset consists of 148,886 genes from 2,000 random network topologes. Filtered 2 dataset consists exclusively of genes that are both regulators and regulated, *i.e.* in this dataset pure regulators and purely regulated genes have been removed. Filtered 2 dataset consists of 105,672 genes from 2,000 random network topologes.

Response	Dataset	Expl. var.	Beta	p-value (GLMM)[1]	MI	p-value (MI)[2]
Expression variance	Filtered 1	Instrength	0.29	$< 2.2 \times 10^{-16}$ ***	0.8	10^{-4} ***
		Outstrength	7.8×10^{-4}	$< 2.6 \times 10^{-9}$ ***	0.61	10^{-4} ***
	Unfiltered	Instrength	0.28	$< 2.2 \times 10^{-16}$ ***	0.67	10^{-4} ***
		Outstrength	-0.022	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
	Filtered 2	Instrength	0.24	$< 2.2 \times 10^{-16}$ ***	0.43	10^{-4} ***
		Outstrength	-0.08	$< 2.2 \times 10^{-16}$ ***	0.02	10^{-4} ***
Rel. change of expr. variance	Filtered 1	Instrength	-0.033	$< 2.2 \times 10^{-16}$ ***	0.29	10^{-4} ***
		Outstrength	-0.073	$< 2.2 \times 10^{-16}$ ***	0.34	10^{-4} ***
	Unfiltered	Instrength	-0.003	2.9×10^{-10} ***	0.09	10^{-4} ***
		Outstrength	-0.046	$< 2.2 \times 10^{-16}$ ***	0.14	10^{-4} ***
	Filtered 2	Instrength	-0.017	$< 2.2 \times 10^{-16}$ ***	0.07	10^{-4} ***
		Outstrength	-0.035	$< 2.2 \times 10^{-16}$ ***	0.08	10^{-4} ***
Probability of responding to selection	Filtered 1	Instrength	-1.94	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	1.55	$< 9.79 \times 10^{-11}$ ***	—	—
	Unfiltered	Instrength	-1.87	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.08	$< 6.67 \times 10^{-7}$ ***	—	—
	Filtered 2	Instrength	-1.79	$< 2.2 \times 10^{-16}$ ***	—	—
		Outstrength	-0.25	$< 2.2 \times 10^{-16}$ ***	—	—
Gene-specific selective pressure	Filtered 1	Instrength	-0.05	$< 2.2 \times 10^{-16}$ ***	0.63	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.72	10^{-4} ***
	Unfiltered	Instrength	-0.04	$< 2.2 \times 10^{-16}$ ***	0.1	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.05	10^{-4} ***
	Filtered 2	Instrength	-0.04	$< 2.2 \times 10^{-16}$ ***	0.1	10^{-4} ***
		Outstrength	0.03	$< 2.2 \times 10^{-16}$ ***	0.17	10^{-4} ***

[1] Coefficients and their significance were computed using linear mixed-effects model (see Methods). [2] Mutual information p-values were computed using a Monte Carlo permutation test with 10,000 permutations. Asterisks indicate statistical significance: n.s. - p-value > 0.05; * - p-value ≤ 0.05; ** - p-value ≤ 0.01; *** - p-value ≤ 0.001; **** - p-value ≤ 0.0001.

Appendix B

Published manuscripts

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Being noisy in a crowd: Differential selective pressure on gene expression noise in model gene regulatory networks

Nataša Puzović[1]*, **Tanvi Madaan**[1], **Julien Y. Dutheil**[1,2]

1 Molecular Systems Evolution Research Group, Max Planck Institute for Evolutionary Biology, Plön, Schleswig-Holstein, Germany, **2** Institut des sciences de l'évolution, Montpellier, Languedoc-Roussillon, France

* puzovic@evolbio.mpg.de

OPEN ACCESS

Citation: Puzović N, Madaan T, Dutheil JY (2023) Being noisy in a crowd: Differential selective pressure on gene expression noise in model gene regulatory networks. PLoS Comput Biol 19(4): e1010982. https://doi.org/10.1371/journal.pcbi.1010982

Editor: Marija Cvijovic, Chalmers University of Technology and University of Gothenburg, SWEDEN

Received: August 8, 2022

Accepted: February 27, 2023

Published: April 20, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1010982

Data Availability Statement: The simulation results data and the code necessary to reproduce all figures is available at https://doi.org/10.5281/

Abstract

Expression noise, the variability of the amount of gene product among isogenic cells grown in identical conditions, originates from the inherent stochasticity of diffusion and binding of the molecular players involved in transcription and translation. It has been shown that expression noise is an evolvable trait and that central genes exhibit less noise than peripheral genes in gene networks. A possible explanation for this pattern is increased selective pressure on central genes since they propagate their noise to downstream targets, leading to noise amplification. To test this hypothesis, we developed a new gene regulatory network model with inheritable stochastic gene expression and simulated the evolution of gene-specific expression noise under constraint at the network level. Stabilizing selection was imposed on the expression level of all genes in the network and rounds of mutation, selection, replication and recombination were performed. We observed that local network features affect both the probability to respond to selection, and the strength of the selective pressure acting on individual genes. In particular, the reduction of gene-specific expression noise as a response to stabilizing selection on the gene expression level is higher in genes with higher centrality metrics. Furthermore, global topological structures such as network diameter, centralization and average degree affect the average expression variance and average selective pressure acting on constituent genes. Our results demonstrate that selection at the network level leads to differential selective pressure at the gene level, and local and global network characteristics are an essential component of gene-specific expression noise evolution.

Author summary

"No man is an island, entire of itself. Each is a piece of the continent, a part of the main." declares John Donne in his poem *For Whom the Bell Tolls*, emphasizing that no individual human is entirely separate from humanity as a whole interconnected system. Organisms are biological systems constituted of many interacting components that also interact with each other and the environment. Understanding the evolution of single components such

zenodo.6939845, together with the code necessary to generate all raw simulation files.

Funding: NP is funded by the International Max Planck Research School (IMPRS) for Evolutionary Biology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

as individual cells or genes can only be fully achieved by considering the interactions with other components. Here, we study the evolution of the cell-to-cell variability of gene expression, the so-called expression noise. To understand the evolution of gene-specific expression noise, we develop a model of gene network evolution with selection at the gene regulatory network level. We find that selection at the gene network level has different repercussions for individual genes based on their position in the network and that gene expression noise is more constrained in genes that are central in the network. Furthermore, the topological structure of the background network affects the propagation and evolvability of gene expression noise. These findings indicate that selection on a given system results in differential selective pressures at the level of subsystems. Our results further suggest that selection to mitigate inherent noise plays a role in network and gene evolution.

Introduction

Living beings are complex systems constituted of many genes that interact with each other and the environment to create an organism. From prokaryotes with a few hundred essential genes, to eukaryotes with possibly several thousands, cells require many gene products to work together to perform housekeeping functions and to replicate. Fine-tuned molecular processes, generally referred to as *gene expression*, ensure how, where and when these products are generated. However, gene expression is an inherently noisy process [1, 2], which involves many steps where molecules participating in the expression machinery diffuse and bind to target molecules. Additionally, these molecules are often present in small copy numbers, increasing the susceptibility of gene expression to stochastic events. Consequently, there is a variation in gene expression levels among cells, even if they are isogenic and grown in a homogeneous environment, and this inevitable variation has been termed *gene expression noise*. Organisms have to express hundreds of genes, each one of which is noisy—raising the question of how they evolved to cope with this inevitable noise.

The expression noise level of a particular gene may be decomposed into two components, called *extrinsic* and *intrinsic*. Extrinsic noise affects all genes equally and results from the sharing of key molecules, such as RNA polymerases and ribosomes, by all genes in the expression process, as well as, for instance, differences in cell size and phase in the cell cycle. Intrinsic noise is gene-specific and results from different chromatin states, cis-regulatory elements and kinetic parameters of transcription and translation of each gene [3]. Minor sequence mutations can have a significant effect on the level of expression noise. For example, a small number of single-nucleotide changes in a transcription factor binding site were reported to have a large effect on the expression noise level [4]. Since (i) there is variation in the level of intrinsic noise of genes, and (ii) intrinsic noise is genetically determined—and, therefore, heritable—gene expression noise can be shaped by natural selection.

Evidence of selection on expression noise was first seen in the fact that dosage-sensitive genes [5] and essential genes exhibit lower levels of expression noise [6, 7]. Intrinsic noise was also reported to correlate with the strength of selection acting on the encoded protein. Namely, proteins with a lower ratio of non-synonymous over synonymous substitution rate (Ka/Ks) have a lower level of expression noise [8]. Changes in the expression noise of a single gene may be either beneficial or deleterious, depending on how far its mean expression is from the optimal expression level [9]. Expression noise is deleterious if the mean expression level is close to the optimal, as higher variation, in this case, generates a larger number of less fit individuals,

reducing the population fitness. Conversely, expression noise can be beneficial if the mean expression level is far from the optimum, as noisy genes are more likely to generate cells with an expression level closer to the optimum. Noisy gene expression can thus be part of a bet-hedging strategy and was observed in genes involved in immune and environmental response [10–13]. The fitness cost of changes in the level of expression noise in the fitness landscapes of ≈ 30 yeast genes have been shown to be on the same order as fitness costs of changes in mean expression level [14]. Since the fitness effect of different levels of expression noise can be as detrimental as different mean expression levels, which are thought to be extensively under selection [15], it can be assumed that expression noise is extensively under selection genome-wide. Prevalent selection on expression noise has been demonstrated in naturally segregating promoter variants of *E. coli* [16].

The phenotype (and, therefore, the fitness) of an organism depends on the interaction of many genes. As a result, genes do not evolve independently, and the selective pressure acting on a gene's intrinsic noise depends on its interactions with other genes. Understanding the evolution of gene expression noise requires accounting for such gene-to-gene interactions, commonly depicted by a gene network. The propagation of noise from gene to gene in the network was established both theoretically and experimentally [17, 18]. Genes with many connections propagate their noise to a more substantial extent than genes with fewer connections and, therefore, contribute more to the global noise levels of the network. Gene networks are robust to variation in the expression level of their system components to some degree, but at a critical point the global noise of the network becomes too high and leads to network collapse. Selection against noise at the network level was, therefore, hypothesized to result in stronger constraints on the intrinsic noise of highly connected genes [8]. Moreover, the topological structure of the network has been shown to affect the pattern of noise propagation [19], suggesting that the topology of the network might impose additional selective constraints on the constituent genes.

Here, we test the hypothesis that expression noise of highly connected genes in gene networks is under stronger selective pressure than expression noise in peripheral genes using an *in silico* evolutionary experiment. We introduce a new gene regulatory network evolution model, which includes an evolvable component of stochastic gene expression, and use it to evolve thousands of network topology samples over 10,000 generations. These simulations showed that highly connected genes have a more constrained intrinsic expression noise. They further revealed that not all genes might evolve in response to network-level selection, and the probability that they do so depends on local network properties. Lastly, the average selective pressure acting on genes in a network is affected by topological features such as network diameter, centralization and average degree.

Materials and methods

We introduce a new gene regulatory network model that incorporates intrinsic expression noise. We then use this model within a forward simulation framework to simulate the evolution of populations of networks with mutable levels of intrinsic expression noise. These simulations allow us to study how the selective pressure acting on expression noise varies within the regulatory network.

A gene regulatory network model with stochastic gene expression

To investigate the evolution of stochastic gene expression in gene regulatory networks, we first extend Wagner's gene network model [20] to integrate gene-specific expression noise.

We model a network of n genes ($n = 40$ in this study) defined by a regulatory matrix $W = (w_{ij})_{1\leq i\leq n,\ 1\leq j\leq n}$, and a vector of intrinsic, gene-specific noise $\{\eta_i^{\text{int}}\}_{1\leq i\leq n}$. Each element w_{ij} of

the regulatory matrix W defines the regulatory effect of gene j on gene i. The value of w_{ij} is a real number and is referred to as regulatory strength of gene j on gene i. In case $w_{ij} > 0$, gene j is an activator of gene i and increases its expression level. Conversely, when $w_{ij} < 0$, gene j is a repressor of gene i and decreases its expression level. Lastly, if $w_{ij} = 0$, gene i is not regulated by gene j and gene j has no effect on expression level of gene i. Two genes i and j are connected by an edge in the network if at least one of w_{ij} and w_{ji} is non-null. The intrinsic noise vector $\{\eta_i^{\text{int}}\}_{1\leq i\leq n}$ defines the gene-specific expression noise of each gene in the network. The regulatory matrix and the intrinsic noise vector together constitute a unique genotype in this modeling framework (Fig 1A).

The phenotype (the expression level of each gene) in the model is represented by a state vector $\{S_i\}_{1\leq i\leq n} = \{s_1, s_2, \ldots, s_n\}$, which describes the expression level of each gene. The state vector at t_0 is set to an arbitrary basal expression level value ($\{S_i^0\}_{1\leq i\leq n} = \{S_i^{basal}\}_{1\leq i\leq n} = \{20, \ldots, 20\}$

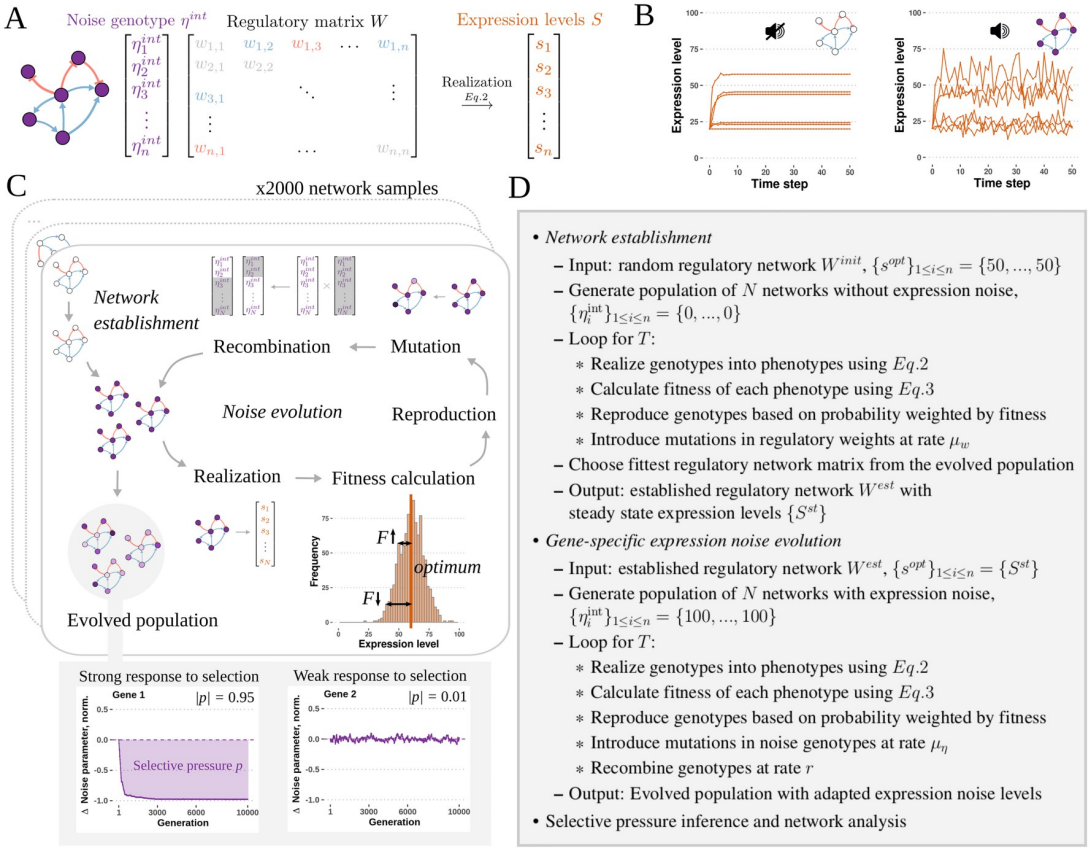

Fig 1. The evolution of gene-specific expression noise was simulated using populations of model gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions. A—Gene regulatory network model. The genotype consists of the intrinsic noise vector η^{int} and regulatory matrix W. The intrinsic noise vector defines the gene-specific expression variance of each gene in the network. The regulatory matrix defines the regulatory interactions in the network. The genotype is realized into the phenotype using the dynamical equation described in the main text. The phenotype is given by the state vector S, which represents the expression level of each gene in the network. **B**—Deterministic (left) and stochastic (right) realizations of the model. **C**—Steps of the evolutionary simulation process. Each established network configuration was used as a founding network for the network populations used in the noise evolution simulation. In every generation, genotypes are realized and phenotypes (expression levels) are sampled from the last time step. Fitness is calculated from the expression levels. If the populations are evolved under selection, fitness is calculated as the distance of the expression level of each gene from the optimal expression level. Genotypes are reproduced based on their relative fitness and mutations in the intrinsic noise vectors are introduced. Noise genotype vectors are recombined by randomly choosing individuals for recombination and shuffling their noise vectors. The process is repeated for 10,000 generations. **D**—Algorithm overview.

https://doi.org/10.1371/journal.pcbi.1010982.g001

in this study). In every time step t ($1 \leq t \leq T_r$, with $T_r = 50$ in this study), the expression level of each gene is recomputed. The cumulative effect of all transcription factors in the expression level of each gene is for simplicity considered to be additive, *i.e.* we assume there is no cooperative or competitive binding of transcription factors to transcription factor binding sites. This assumption removes the small degree of non-linearity in the response of the regulated gene to transcription factor concentrations, which is present in real transcription factor regulation dynamics. The activation rate $a_i(t)$ is defined as the sum of all effects the regulators of gene i have on its expression level at time step t:

$$a_i(t) = \sum_{j=1}^{n} w_{ij} \cdot s_j(t), \tag{1}$$

in which case the dynamic equation for the expression level of each gene in the following time step is:

$$s_i(t+1) \sim \mathcal{N}(s_i^{basal} + a_i(t), \eta_i^{int}). \tag{2}$$

In every time step the expression level of a gene is drawn from a random distribution. We implemented a simple Gaussian noise, where the mean of the normal distribution equals the sum of basal expression level (s_i^{basal}) and activation rate ($a_i(t)$), and the variance equals the gene noise genotype (η_i^{int}). If the expression level value drawn from the normal distribution is below the minimal ($s_{min} = 0$) or above the maximal expression level ($s_{max} = 100$), it is set to the minimal or maximal expression level, respectively. We note that the shape and variance of the distribution is constant in realization time in our model, but that the expression levels of each individual is the product of the trajectory of the expression levels during the realization process, during which expression levels can exhibit phenotypic switching between stable states. Consequently, there can be a non-normal expression level distribution of a certain gene in the clonal population, even though the expression levels in each time step are drawn from a normal distribution.

The expression levels of all genes are synchronously updated in each time step. The steady state expression levels are invariant to whether the expression levels of each gene are updated synchronously or asynchronously (S1 Text). Similarly, mean expression level, expression variance, CV, noise and Fano factor are invariant to the updating mode (S1 Text). The model may be realized as stochastic or deterministic, depending on the noise parameter values (Fig 1B). The deterministic realization has been used to benchmark the model and to set up the mean expression levels for the starting populations, and the stochastic realization has been used in the main bulk of the simulations, in which intrinsic noise is evolved.

Forward-in-time simulation of expression noise evolution

To investigate how gene-specific expression noise of constituent genes responds to stabilizing selection at the network level, we used the newly introduced model to perform forward-in-time evolutionary simulations in which we allow the gene-specific noise levels to mutate. An *in silico* evolutionary process consisting of rounds of mutation, selection, recombination and replication events of a population of N ($N = 1,000$ in this study) individuals was performed for T ($T = 10,000$) generations (Fig 1C).

We first generated network topologies that would serve as the founding network for the populations in our simulations. We generated 2,000 random (Erdős–Rényi model) network topologies of 40 nodes with regulatory strength values drawn from a uniform distribution $\mathcal{U}(-3, 3)$. The network density was $d = 0.05$. Only connected network graphs were used, meaning there is only one component and there are no disconnected subgraphs.

Autoregulation is not present, because it affects gene-specific noise levels and would be a confounding factor in the analysis. In order to assess the effect of the topology structure on the evolution of expression noise, we also generated an additional 1,000 scale-free (Barabási–Albert model) and 1,000 small-world (Watts–Strogatz model) network topologies with the same size and density. Both random and small-world networks are characterized by a Poisson degree distribution and short mean shortest path length, but random networks have a low clustering coefficient, while small-world networks have a high clustering coefficient. Scale-free networks are characterized by a degree distribution that follows a power law. Real-world networks exhibit degree distributions similar to power-law distributions, high clustering and short path lengths. As such, real-world networks have features of both scale-free and small-world networks [21].

In the simulation of expression noise evolution the regulatory interactions were immutable and the values of the noise genotype vectors were allowed to mutate. Stabilizing selection, the selection scenario in which individuals with extreme phenotypic values have a lower fitness, was imposed on all constituent genes by setting the value of optimal expression level as the mean equilibrium expression level of each gene. The fitness $F(s)$ of a phenotype s was calculated as in Laarits et al. [22], where fitness is defined as the distance from the optimal expression state vector $\{s_i^{opt}\}_{1\leq i\leq n}$, weighted by the fitness contribution given by $\{\rho_i\}_{1\leq i\leq n}$:

$$F(s) = e^{-\sum_{i=1}^{n}|s_i^{opt} - s_i|/n\rho_i} \tag{3}$$

The fitness contribution parameters $\{\rho_i\}_{1\leq i\leq n}$ define the contribution of each gene to the fitness of the phenotype, *i.e.* it is a scaling factor of the decrease of fitness as a function of the distance of the expression level from the optimal expression level for each gene. In this study, the strength of the imposed selective pressure is set to be identical for all constituent genes ($\forall i\ \rho_i = 1$). The assumption of all genes having identical fitness contribution is biologically unrealistic, so we have also performed simulations in which we impose unequal fitness contributions among genes in the same network. We found consistent conclusions (S5 Text), and, for simplicity, we report the results with equal fitness contributions here. Since the fitness contribution of all genes is identical, any differences in the evolutionary outcome we observe after removing the effect of drift will be due to gene differences in their network interactions. Individuals were reproduced into the next generation with a probability equal to their relative phenotype fitness. The fitness of all phenotypes in populations evolved under non-selective conditions was set to an equal constant value, regardless of gene expression levels. Mutations were introduced at a rate μ_η ($\mu_\eta = 0.01$) per gene per replication event. The values for noise genotype mutations were drawn from a normal distribution $\mathcal{N}(100, 40)$. There is no experimental evidence for the shape of the distribution of the expression noise and regulatory strength mutations. We chose a normal distribution because: 1) it defines equally frequent beneficial and deleterious mutations and 2) most mutations would have a small effect, which reflects the characteristic of many studied distributions of fitness effects in model organisms. Recombination was implemented by choosing a random offspring individual at a rate r ($r = 0.05$) and introducing a random break point in the linear genome. The genotype values in the genome segment defined by the break point were then exchanged with another randomly chosen individual from the offspring population. A constant population size N ($N = 1,000$) was maintained. To account for the effect of genetic drift, the noise evolution simulations of each founding network population were replicated 10 times under selection and 10 times under neutrality.

We found that the expression level of most genes in networks with random configurations converge to either s_{min} or s_{max} under a deterministic realization. The measurement of variance of genes that are either not expressed at all or expressed at the maximal level would be impaired since their expression range is constrained by the lower and upper expression level boundary. Since the study of expression variance is our main focus, we added a network establishment step before the noise evolution simulations, in which we subject the network regulatory matrix to mutation and selection for intermediate expression levels. During the network establishment step networks are realized deterministically, *i.e*. the intrinsic noise genotype of all genes is 0. Networks with intermediate steady state expression levels were established through the evolutionary process by imposing a target expression level $\{s_i^{opt}\}_{1\leq i\leq n}$ ($\{s_i^{opt}\}_{1\leq i\leq n} = \{\frac{s_{max}}{2}, \ldots, \frac{s_{max}}{2}\}$) for all genes and allowing the strength of regulatory interactions to mutate. Mutations were introduced at a rate μ_w ($\mu_w = 0.1$) in non-zero entries in the regulatory matrix, preserving the network topology structure (Erdős–Rényi, Barabási–Albert, or Watts–Strogatz model). The values for regulatory strength mutations were drawn from a normal distribution $\mathcal{N}(0, 2)$. Recombination was not implemented at this stage. Fitness of each individual was computed as the distance of the phenotype to the optimal expression state vector using Eq 1. Individuals were reproduced with a probability equal to the relative fitness and the population size kept constant. Network regulatory configurations in which the expression level of all genes would not converge to a fixed point and would oscillate were discarded, as in previous studies [22]. Oscillating gene expression level patterns create population-level heterogeneity generated by the system oscillations and not by stochastic gene expression. Since we are studying the evolution of gene-specific expression noise, expression noise generated by oscillations would be a confounding factor in our analysis. We note, however, that oscillatory networks can be frequent in simulations [23] and biological systems [24], and the role of expression noise in their behavior is an interesting perspective for follow-up studies. Expression level dynamics were termed oscillating if the sum of the differences between expression level in the last time step and previous τ time steps ($\tau = 10$) was higher than ϵ ($\epsilon = 10^{-6}$). A stable, *i.e*. non-oscillating, expression level dynamics satisfied the following criterion [22]:

$$\Phi(S(t)) = \frac{1}{\tau}\sum_{\theta=t-\tau}^{t} D(S(\theta), S(t)) < \epsilon \tag{4}$$

where D is the distance between two vectors $D(S^1, S^2) = \sum_{i=1}^{n} |S_i^1 - S_i^2|/n$.

The network establishment process consisting of rounds of mutation, selection and reproduction of a population of N ($N = 1,000$) individuals was performed for T ($T = 10,000$) generations, for each network topology. At the end of the network establishment process, 68% (54333/80000) of genes had intermediate expression levels (S1 Text). The reason why a minority of the genes do not reach close to optimum expression levels could be potential network configuration constraints or a non-extensive optimization/fitting algorithm. Genes that had an expression level of 0 or s_{max} were filtered out from the dataset used in the final analysis. The network regulatory configuration with the highest fitness was chosen from the evolved population and this network configuration was used to generate the starting population for the noise evolution simulations.

The gene network model and evolutionary simulations were implemented in C++ and the source code is available at https://gitlab.gwdg.de/molsysevol/supplementarydata_expressionnoise/cpp.

Analysis of simulation results: Expression noise and network centrality measures

The evolutionary outcomes (*i.e.* the change of phenotypes and genotypes) were measured as change of expression noise and selective pressure for each network, respectively. Expression noise in the first and last generation in each evolved population was measured as the variance of the population expression level states for each gene. The change of expression noise (phenotypic evolution) between the first and last generation was measured as the relative change of expression noise, calculated as the difference of expression variance between the first and last generation divided by their sum $(\sigma^2_{gen1} - \sigma^2_{gen10k})/(\sigma^2_{gen1} + \sigma^2_{gen10k})$.

The selective pressure (genotypic evolution) acting on each gene was measured as the average change of noise genotype in every second generation relative to the starting level (Fig 1C). To compare the effect of node centrality on the selective pressure acting on constituent genes, we computed node-level network centrality measures for each node in the networks. We focused our analysis on two local network centrality measures, node instrength and outstrength, but over 30 network centrality measures were analyzed (S2 Text). Instrength of node i is measure of the strength and number of in-going links, *i.e.* how strongly a gene is being regulated:

$$\text{Instrength}(i) = \sum_{j}^{n} |w_{ij}|. \tag{5}$$

Conversely, the outstrength of node j is a measure of the strength and number of outgoing links, *i.e.* how strongly a gene regulates other genes downstream:

$$\text{Outstrength}(j) = \sum_{i}^{n} |w_{ij}|. \tag{6}$$

Further, we computed global graph-level metrics, such as mean graph distance and performed a principal component analysis to reduce the dimensionality (S2 Text). The results were analysed in `R 3.6.3` [25]. Network analyses were performed using the `igraph 1.2.4.2` [26] and `statnet 2019.6` [27] packages. Principal component analysis was performed using the `ade4 1.7.15` [28] package.

Analysis of simulation results: Linear modeling

We fitted linear mixed-effects models using network centrality measures as fixed effect variables and the network topology sample as a random effect variable, allowing for control of intra-network correlation in the response variable. We tested different transformations of the response and explanatory variables in order to improve linearity, and variance structures to account for heteroskedasticity of the residuals. A model where the residual variance was an exponential function of the node absolute instrength was shown to provide the best fit according to the minimal Akaike's Information criterion and was used for all subsequent models (S3 Text). Two types of models were fitted: a logistic regression where the response variable was set to whether a gene answered to selection or not, and standard regressions that used expression variance, relative change of expression variance or selective pressure as response variables. Linear mixed-effect modelling was performed using the `nlme 3.1.144` [29] and `lme4 1.1.27.1` [30] packages. Marginal and conditional R^2 values were computed using the `MuMIn 1.43.17` [31] package. Network centrality measures used as explanatory variables in our linear models were correlated (Pearson's $r = -0.17$, p-value $< 2.2 \times 10^{-16}$, S2 Text), so we computed the variance inflation factor (VIF) using the `car 3.0.11` [32]

package. The VIF of all linear models was less than 3; therefore, colinearity was considered to have negligible impact on the inferred statistical significance [33]. To improve homoskedasticity of the residuals in the linear models, we also performed each model fit on two filtered datasets: one in which genes with zero values of instrength or outstrength were removed, and one in which only genes with zero values of instrength or outstrength were kept. The same pattern of effects and significance is observed in the filtered as in the main dataset, so we included the results of the complete dataset in the main text and reported the results of the reduced dataset in the supplementary information (S6 Text).

Finally, since in some cases variable transformation, heterogeneous variance modeling and data filtering did not ensure normality and independence of the residuals, we assessed the amount of resulting bias in the estimation of p-values using a randomization test, in which we fitted a selected model on 10,000 permuted datasets. We chose the model of relative noise change (S3 Text), as the corresponding residuals were significantly departing normality (Shapiro-Wilk test, p-value $< 2.2 \times 10^{-16}$) and independence (Box-Ljung test, p-value $= 8.9 \times 10^{-7}$). For each permutation, we shuffled the values of the response variable (relative change of variance) within each network topology, which removes the effect of network metrics on the change of noise, but preserves the distributions of each metric per network, as well as putative colinearity between explanatory variables. Using $\alpha = 0.05$ as a significance cutoff value, we found a false discovery rate (FDR) of 6.0% for the effect of instrength and and 6.7% for the effect of outstrength. While these values are above the expected 5%, the FDR inflation was found to be relatively low and we concluded that the non-normality of residuals did not affect our conclusions.

Analysis of simulation results: Information-based metrics

Generalized linear mixed-effects models make several assumptions that might be violated by the data in some cases. Namely, they assume a normal distribution and homoskedasticity of Pearson's residuals, and a normal distribution of random effects. To further validate our conclusions, we computed the mutual information (MI) between variables, which does not have any prior assumptions. We calculated mutual information between the expression noise and centrality metrics using the `infotheo 1.2.0` [34] package. Monte Carlo permutation tests with 10,000 permutations were used to compute p-values for the significance of the mutual information between each pair of tested variables.

Results

We investigate how selection at the gene network level may lead to the evolution of differential gene-specific expression noise, as observed in biological systems. To do so, we introduce a new gene regulatory model with stochastic gene expression, which extends Wagner's model [20] by adding node-specific intrinsic noise parameters (Fig 1A and 1B). In this framework, the phenotype is represented by the expression level of each gene, and is the realization of a random distribution determined by the genotype. The fitness of an individual is further determined by its distance to an optimal phenotype, therefore, stabilizing selection is implemented as acting on the expression level. We used this model to simulate the evolution of populations of gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions (Fig 1C and 1D), and assessed how node properties affect the evolution of intrinsic noise.

Expression noise propagates along the regulatory network

We first investigated how noise propagated in the model gene regulatory networks. It was shown that noise is additive in biological networks and, therefore, propagates from regulators

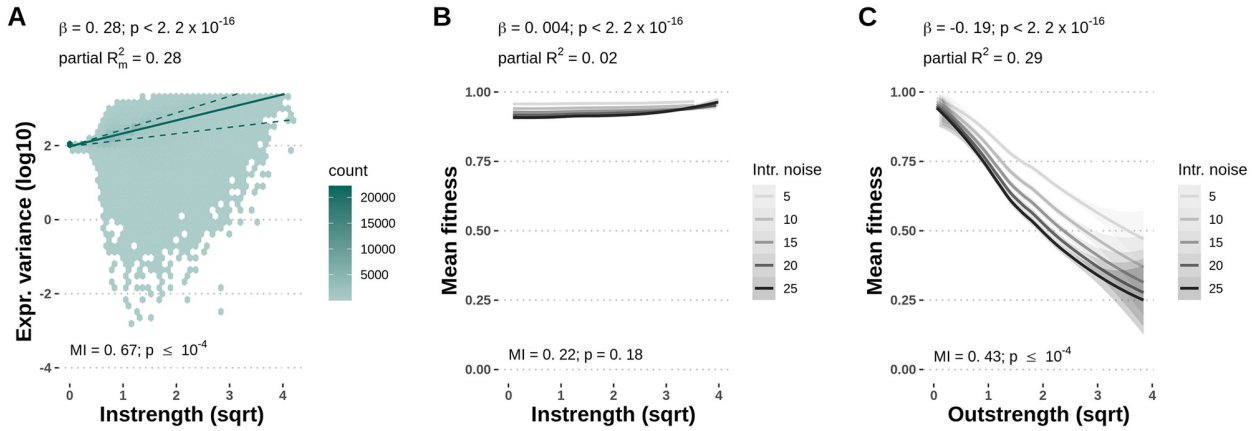

Fig 2. Noise propagation is captured by the gene regulatory network model. A—Gene-specific expression variance increases with the absolute instrength of the node, indicating noise propagation is reflected in the gene regulatory network model. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **B, C**—Gene-specific expression variance decreases fitness in gene networks under stabilizing selection on gene expression level. Increasing the level of gene-specific expression noise reduces the mean fitness of the clonal population. The mean fitness of the population is significantly, but marginally, increased by noise in genes with higher node instrength (B), and significantly decreased by noise in genes with higher node outstrength (C). Lines represent the smoothed conditional means and grey bands represent the 95% confidence interval bands. Coefficients, p-values and partial marginal R^2 measures are estimated using linear mixed-effects models with expression variance or mean fitness as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed with a permutation test with 10,000 permutations.

https://doi.org/10.1371/journal.pcbi.1010982.g002

to regulated genes [17, 18]. To assess whether our model successfully captured this property, we generated a dataset of 2,000 realized random network topologies, and tested whether gene expression variance increased with the number of ingoing regulatory links. As expected, we found that the absolute instrength of a gene had a significant positive effect on gene expression variance (linear mixed-effects model with coefficient $\beta = 0.28$, p-value $< 2.2 \times 10^{-16}$) (Fig 2A), indicating that noise propagation was captured in our model. Furthermore, the mutual information between gene expression variance and absolute instrength was significant (MI = 0.67, p-value $\leq 10^{-4}$, permutation test). High node instrength increases expression noise, in line with the experimental evidence that the noisiness of promoters increases with the number of regulatory inputs [35].

We then looked at fitness costs associated with high expression noise in regulators and regulated genes. In a dataset of 1,000 random network topologies, we assessed the mean fitness of the clonal populations of 1,000 individuals under stabilizing selection on the expression level. Each gene was imposed 5 different levels of intrinsic noise, while the intrinsic noise of the rest of the network was kept at 0. We found that increasing the level of expression noise of a single gene decreased the mean fitness of the network (linear mixed-effects model with coefficient $\beta = -0.002$, p-value $< 2.2 \times 10^{-16}$), as expected. However, the strength of this effect depended on the gene centrality. The reduction of fitness due to gene-specific expression noise was significantly, but marginally, affected by instrength (linear model with coefficient $\beta = 0.004$, p-value $< 2.2 \times 10^{-16}$, Fig 2B). The mutual information between mean fitness of the population and absolute instrength was not significant (MI = 0.22, p-value = 0.18, permutation test). However, the mean fitness significantly decreased with node outstrength (linear model with coefficient $\beta = -0.19$, p-value $< 2.2 \times 10^{-16}$, Fig 2C). The mutual information between mean fitness of the population and absolute outstrength was significant (MI = 0.43, p-value $\leq 10^{-4}$, permutation test). Higher fitness cost of expression noise in gene with high outstrength suggests there is a differential selective pressure acting

on genes based on their centrality in the gene regulatory network, which we explore in the next section using an *in silico* evolutionary experiment.

Gene expression noise is reduced under a stabilizing selection regime

To investigate how gene-specific expression noise responds to stabilizing selection at the network-level, we simulated the evolution of 2,000 random network topologies with and without selection on the gene expression level. We observed that gene expression variance decreased throughout evolution under selective conditions (Fig 3A), and the distribution of intrinsic noise parameters in the population shifted towards lower noise genotype values (Fig 3B), indicating that low-noise alleles conferred a fitness increase to the network. Conversely, gene expression variance remained constant throughout evolution under neutral conditions, and the distribution of noise genotypes reflected only the distribution of random mutations. Replicating the simulations for each network topology sample yielded similar reduction of gene expression variance (Fig 3C) and median noise parameter in the population (Fig 3D). As the initial networks were at their optimal expression level, the mean expression level did not change during evolution and was highly correlated between the first and last generations (Pearson's r = 0.99, p-value $< 2.2 \times 10^{-16}$, S1 Text), confirming that selection acted only on the gene expression variance. Population size had a positive effect on the selective pressure acting on genes, as expected, selection being more efficient in large populations (S1 Text). A population size of 1,000 individuals was chosen for the main simulations as the optimal population size in the trade-off between selecting mutations with small effects and reducing computational speed.

Next, we investigated how individual nodes within a network respond to selection, based on their centrality properties.

Evolutionary change in phenotypes: Regulators reduce their expression noise to a higher degree

We first analysed the phenotype change, *i.e.* the relative change in gene-specific expression variance after evolution. The variance of gene expression depends both on the intrinsic noise

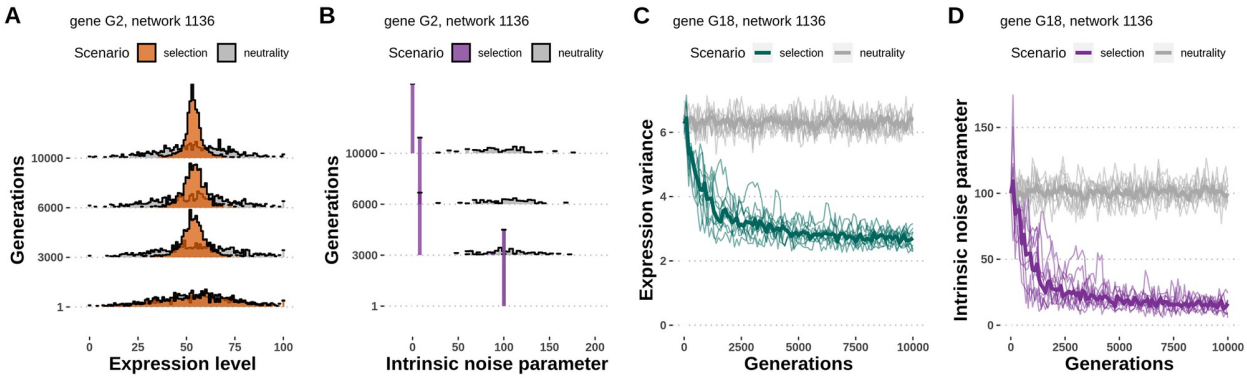

Fig 3. Gene-specific expression noise evolves in a model with selection. A—The distribution of expression levels of an example gene throughout evolution in populations evolved under stabilizing selection on gene expression level and under neutrality. The variance of gene expression level is reduced under selection, but not under neutrality. **B**—The distribution of intrinsic noise parameters of an example gene throughout evolution in populations evolved under selection and under neutrality. The median intrinsic noise parameter skews to lower values under stabilizing selection, but not under neutrality. **C, D**—Replicates of the simulations with the same input network and parameters. Replicates have different dynamics, but reach similar outcomes in terms of expression variance (C) and median intrinsic noise parameter (D) in the evolved populations. The evolution of each network topology sample was replicated 10 times under selection and 10 times under neutrality.

https://doi.org/10.1371/journal.pcbi.1010982.g003

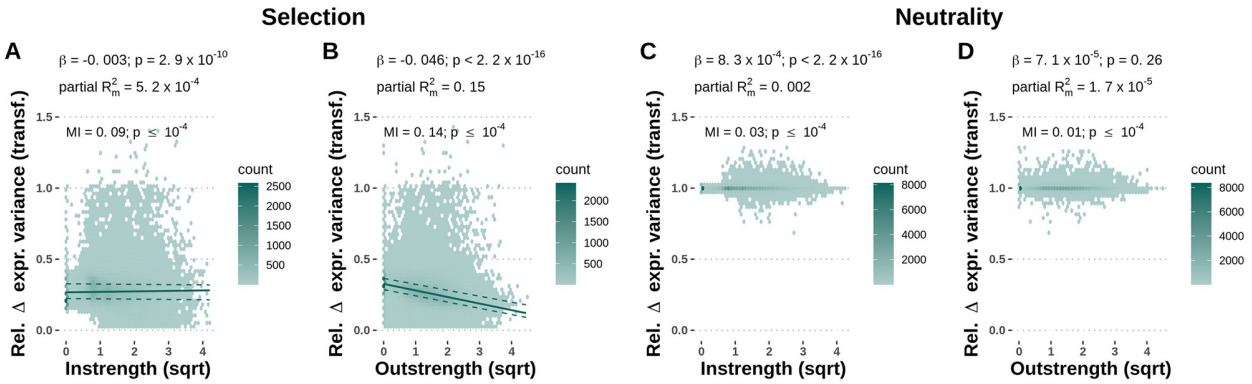

Fig 4. Node-level network centrality measures affect the relative change of gene-specific expression variance under network-level selection. For each gene, the relative change of expression variance before and after evolution (Rel. Δ expr. variance) was averaged over all replicates. **A, B**—Absolute instrength (A) and absolute outstrength (B) have a significant negative effect on the relative change in gene expression variance in populations evolved under selection. A lower value of relative change of expression variance indicates a bigger reduction in expression variance between the first and last generation and a stronger response to selection. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **C, D**—Absolute instrength (C) and absolute outstrength (D) have a significant, but negligible, negative effect on the relative change in gene expression variance in the populations evolved under neutrality. The dataset consists of 74,443 genes from 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. Coefficients, p-values and partial marginal R^2 measures were estimated using linear mixed-effects models with relative change of gene-specific variance as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed using 10,000 permutations.

https://doi.org/10.1371/journal.pcbi.1010982.g004

of the genes (that is, its genotype in our model) and the number and noise of the genes it is connected with.

We fitted linear models to assess the impact of the absolute instrength and outstrength measures on the relative change in expression variance for each node in each network. Under selection, both absolute instrength and absolute outstrength had a significant negative effect (linear mixed-effects model with coefficients $\beta_{\text{instrength}}$ = -0.003, p-value = 2.9×10^{-10}, Fig 4A; $\beta_{\text{outstrength}}$ = -0.046, p-value < 2.2×10^{-16}, Fig 4B), meaning that genes with more and stronger connections reduced their expression variance to a larger extent than less connected genes. The effect was notably stronger for outstrength (marginal R^2 = 0.15) than for instrength (marginal $R^2 = 5.2 \times 10^{-4}$). Similarly, the mutual information was significant between the relative change in gene expression variance under selection and absolute instrength (MI = 0.09, p-value $\leq 10^{-4}$, permutation test) and absolute outstrength (MI = 0.14, p-value $\leq 10^{-4}$, permutation test). Genes with high outstrength are strong regulators and their reduction of expression variance to a larger extent indicates that high expression noise is more detrimental in regulators than in regulated genes. Under neutrality, absolute instrength had a significantly positive effect (linear mixed-effects model with coefficient $\beta = 8.3 \times 10^{-4}$, p-value < 2.2×10^{-16}, Fig 4C) and absolute outstrength did not have a significant effect on the relative change in gene expression variance (linear mixed-effects model with coefficient $\beta = 7.1 \times 10^{-5}$, p-value = 0.26, Fig 4D). The mutual information was significant between the relative change in gene expression variance under neutrality and absolute instrength (MI = 0.03, p-value $\leq 10^{-4}$, permutation test) and absolute outstrength (MI = 0.01, p-value $\leq 10^{-4}$, permutation test). These effects are much smaller and of opposite direction than the ones measured in selective conditions, indicating that genetic drift did not cause the effect of centrality measures on expression variance observed in selected populations.

Evolutionary change in genotypes: Regulators are more likely to respond—And display a stronger response—To selection

To investigate differential selective pressure acting on gene-specific expression noise, we analysed the change of intrinsic noise parameters in populations of gene regulatory networks evolved with or without stabilizing selection on the expression level. We measured the selective pressure acting on individual genes as the average reduction in the intrinsic noise parameter relative to the beginning of the evolutionary simulation (see Methods). The selective pressure on genes was found to be close to 0 in neutrally evolving populations, as expected (Fig 5B). In the presence of selection, however, the distribution of selective pressures was found to be bimodal (Fig 5A). Therefore, we binned genes in two categories according to whether they responded to selection (selective pressure > 0.5) or not (selective pressure ≤ 0.5). We then separately analysed the probability to respond to selection and the strength of the response.

Absolute instrength had a significant and strongly negative effect (logistic regression with coefficient β = -1.87, p-value $< 2.2 \times 10^{-16}$, Fig 5C) on the probability of a gene to respond to selection, that is, genes with more and stronger incoming links are less likely to respond to selection. Absolute outstrength also had a significant effect on the probability of a gene to respond to selection (logistic regression with coefficient β = -0.08, p-value = 6.7×10^{-7}, Fig 5D). However, this effect was small and was lost when the interaction terms between instrength and outstrength were included in the model (SI).

For a qualitative analysis of the effect of network centrality on the selective pressure acting on individual genes, we fitted linear-mixed effects models on the set of genes that responded to selection, with selective pressure as the response variable. In the genes that responded to selection from the selected populations, absolute instrength had a significant negative effect (linear mixed-effects model with coefficient β = -0.04, p-value $< 2.2 \times 10^{-16}$, Fig 5E). Conversely, absolute outstrength had a significant positive effect (linear mixed-effects model with coefficient $\beta = 0.03$, p-value $< 2.2 \times 10^{-16}$, Fig 5F) on the selective pressure. In the selected populations, the mutual information was significant between the selective pressure and absolute instrength (MI = 0.19, p-value $\leq 10^{-4}$, permutation test) and absolute outstrength (MI = 0.31, p-value $\leq 10^{-4}$, permutation test). In the neutral populations, neither absolute instrength nor absolute outstrength had a significant effect (linear mixed-effects model with coefficient $\beta_{\text{instrength}} = 2.4 \times 10^{-8}$, p-value = 0.99, Fig 5G; $\beta_{\text{outstrength}} = -1.2 \times 10^{-5}$, p-value = 0.49, Fig 5H) on the selective pressure. Similarly, the mutual information was not significant between the selective pressure and absolute instrength (MI = 0.005, p-value = 0.34, permutation test), nor absolute outstrength (MI = 0.005, p-value = 0.45, permutation test).

The increased selective pressure in genes with high outstrength (strong regulators) can be explained by noise propagation to downstream elements. Namely, expression noise in regulators propagates to the genes they regulate, increasing the overall expression noise in the gene regulatory network. If gene expression levels in the network are under stabilizing selection, expression noise is deleterious. Therefore, regulator genes experience a comparatively higher selective pressure to reduce expression noise than regulated genes. In a genome-wide expression noise screen in *Drosophila melanogaster*, transcription factors were found to have lower expression variation [36]. Suppression of expression noise can be attained through negative autoregulation [37–39], whereby a regulator acts as its own repressor. Incidentally, 40% of transcription factors in *E. coli* [40] and many eukaryotic transcription factors [41] have negative autoregulation, indicating a wide-spread control of expression noise in natural regulatory networks.

In contrast to regulator genes, we found that regulated genes, *i.e.* genes with high node instrength, are less likely to respond to selection and the selective pressure decreases with node instrength. Since the expression noise of genes is a sum of their intrinsic noise and noise

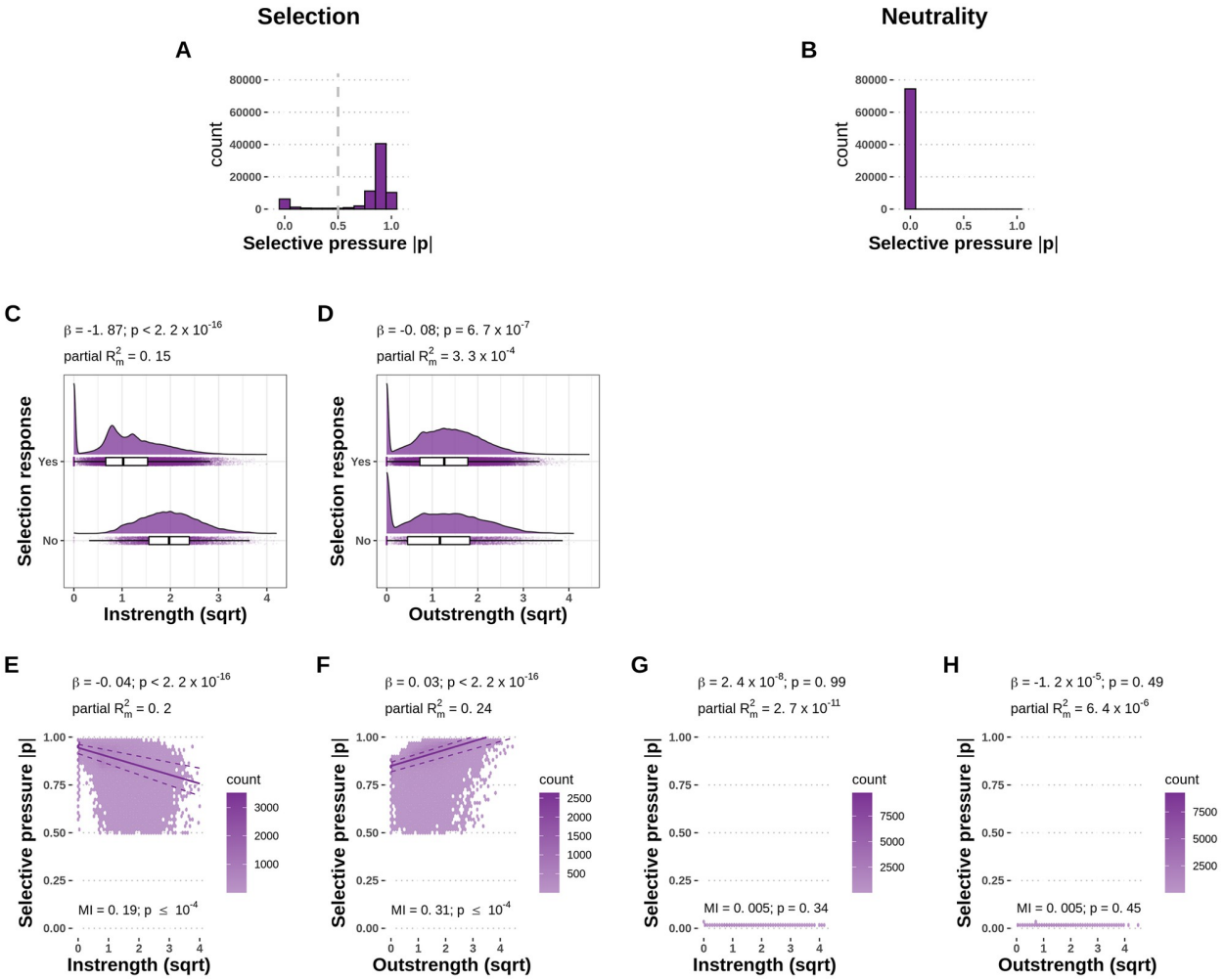

Fig 5. Differential selective pressure is acting on genes based on their centrality. A, B—Distributions of the measured selective pressure in selected (A) and neutral (B) populations. Genes with a selective pressure above 0.5 were categorized as responsive to selection. **C, D**—High instrength genes are less likely to respond to selection. Absolute instrength (C) has a strong significant negative effect on the probability of selection response. Absolute outstrength (D) has a weak significant negative effect on the probability of selection response. **E, F**—In the subset of genes that responded to selection, high instrength (E) decreases the selective pressure, while high outstrength (F) increases the selective pressure acting on individual genes. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **G, H**—Absolute instrength (G) and outstrength (H) have no significant effect on the selective pressure in the non-selected populations. The dataset consists of 74,443 genes from 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Coefficients, p-values and partial marginal R^2 measures are estimated using logistic regression and linear mixed-effects models with selection responsiveness or selective pressure as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were using 10,000 permutations.

https://doi.org/10.1371/journal.pcbi.1010982.g005

propagated from upstream elements, the contribution of intrinsic noise to the total noise of the gene will be comparatively smaller in strongly regulated genes. The network can thus respond to selection either by reducing the intrinsic noise of the focal gene, or by reducing the intrinsic noise of any of the upstream elements, which would reduce propagated noise. As a result, there is a relaxation of selective pressure in regulated genes, which is distributed on upstream genes. On the other hand, the same mechanism increases the selective pressure on upstream genes, *i.e.* regulators.

To check the robustness of our results, we performed the node-level network centrality analysis on two additional datasets with different topology structures: scale-free (Barabási–Albert) and small-world (Watts–Strogatz) topology models. We find consistent effects (direction and significance) of local network centrality metrics on the selective pressure acting on gene-specific noise across topology models, showing that our findings are robust to the topology model used (S4 Text). However, the effect size of network centrality metrics differed between the topology models, pointing at an effect of the topology model on noise propagation and the evolution of gene-specific expression noise, which we investigate in the next section.

Global network properties affect the evolvability of expression noise and selective pressure on constituent genes

Lastly, we analysed how topological structures and graph-level network properties affect the expression noise response of constituent genes to selection on a joint dataset of random (Erdős–Rényi), scale-free (Barabási–Albert) and small-world (Watts–Strogatz) network topologies. Jointly analysing genes from all three topology types with linear models, we observed statistically significant interactions between instrength and outstrength and network topology types on both the probability to respond to selection and the selective pressure acting on gene-specific expression noise (Table 1). We found that genes in scale-free networks have a significantly higher probability of responding to selection than genes in random networks. These results are in agreement with previous studies reporting a higher evolvability of scale-free networks [42, 43]. Conversely, genes in small-world networks have a significantly lower probability of responding to selection than genes in random networks. Furthermore, there are significant effects of interactions between instrength and outstrength with the topology type on the selective pressure on constituent genes.

To investigate which global topological features of the three network models affect expression noise evolution, we performed a principal component analysis (PCA) on 12 graph-level measures. The first two dimensions of the PCA expressed 85.4% of the total dataset inertia (S2 Text), so we used the first two principal components (PCs) as synthetic explanatory variables in linear mixed-effects models. The loading of the first synthetic variable (PC1) is dominated by negative loadings of diameter and mean path distance, and the centralization measures, namely positive loadings of outdegree and closeness centralization and negative loadings of indegree and betweenness centralization. The diameter of a network is defined as the longest shortest path between any two nodes. Centralization is a measure of the extent to which a network is centered around a single node and can be computed from different centrality metrics. The loading of the second synthetic variable (PC2) is dominated by the negative loading of the average degree, average indegree and average outdegree measures (S2 Text). For a more intuitive interpretation, the signs of both PCs have been switched in the statistical analysis. Therefore, PC1 shown in the results is dominated by positive loadings of diameter, mean path distance, indegree centralization and negative loadings of outdegree centralization, and PC2 is dominated by positive loadings of average degree. We refer to PC1 and PC2 as synthetic network diameter and centralization and synthetic average degree, respectively.

The average expression variance per network is significantly negatively affected by synthetic network diameter and centralization (linear model with synthetic network diameter and centralization coefficient β = -6.19, p-value $< 2.2 \times 10^{-16}$) and significantly positively affected by the synthetic average degree (linear model with synthetic average degree coefficient β = 13.26, p-value $< 2.2 \times 10^{-16}$). The mutual information was significant between the average expression variance per network and synthetic network diameter and centralization (MI = 0.21, p-value $\leq 10^{-4}$, permutation test) and synthetic average degree (MI = 0.21, p-value $\leq 10^{-4}$,

Table 1. Network topology type affects the probability of responding to selection and selective pressure on gene-specific expression noise under stabilizing selection on gene expression level.

Response	Explanatory variable	Beta	SE	p-value[1]	
Probability of responding to selection	Instrength	-1.9270	0.0284	$< 2.2 \times 10^{-16}$	****
	Outstrength	-0.0829	0.0226	$< 2.6 \times 10^{-4}$	***
	Scale-free (BA) topology[2]	0.9209	0.1075	$< 2.2 \times 10^{-16}$	****
	Small-world (WS) topology[3]	-0.2684	0.0945	0.0045	**
	Instrength:BA[4]	0.0120	0.0516	0.8159	n.s.
	Instrength:WS	0.0006	0.0401	0.9873	n.s.
	Outstrength:BA	-0.2947	0.0252	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0728	0.0333	0.0287	*
Gene-specific selective pressure	Instrength	-0.0377	0.0004	$< 2.2 \times 10^{-16}$	****
	Outstrength	0.0347	0.0003	$< 2.2 \times 10^{-16}$	****
	Scale-free (BA) topology	0.0019	0.0012	0.1404	n.s.
	Small-world (WS) topology	0.0222	0.0013	$< 2.2 \times 10^{-16}$	****
	Instrength:BA	0.0143	0.0007	$< 2.2 \times 10^{-16}$	****
	Instrength:WS	-0.0055	0.0006	$< 2.2 \times 10^{-16}$	****
	Outstrength:BA	-0.0151	0.0003	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0075	0.0005	$< 2.2 \times 10^{-16}$	****

[1] Coefficients and their significance were computed using linear mixed-effects models (see Methods). The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene was calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Genes were termed responsive to selection if their selective pressure was above 0.5. Asterisks indicate statistical significance: n.s.—p-value > 0.05;

*—p-value ≤ 0.05;

**—p-value ≤ 0.01;

***—p-value ≤ 0.001;

****—p-value ≤ 0.0001.

[2] Barabási–Albert network model.

[3] Watts–Strogatz network model.

[4] Colons (':') indicate variable interactions.

https://doi.org/10.1371/journal.pcbi.1010982.t001

permutation test). This finding means that global network properties affect the amplification of noise through noise propagation between the genes. Specifically, networks with a lower diameter, mean path distance, indegree centralization, and higher outdegree centralization and average degree, had higher average gene expression variance. In the selected populations, the average selective pressure per network was significantly negatively affected by both synthetic network diameter and centralization and the synthetic average degree (linear model with synthetic network diameter and centralization coefficient β = -0.003, p-value = 4.9×10^{-11}, Fig 6A; synthetic average degree coefficient β = -0.009, p-value $< 2.2 \times 10^{-16}$, Fig 6B). The mutual information was significant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.27, p-value $\leq 10^{-4}$, permutation test) and synthetic average degree (MI = 0.26, p-value $\leq 10^{-4}$, permutation test). This result shows that the average selective pressure acting on gene-specific expression noise in networks decreases with an increase of network diameter, mean path distance, indegree centralization and average degree per network. Conversely, the average selective pressure increases with an increase of outdegree centralization (Fig 6A and 6B). In the populations evolved under neutrality, neither synthetic network diameter and centralization, nor synthetic average degree,

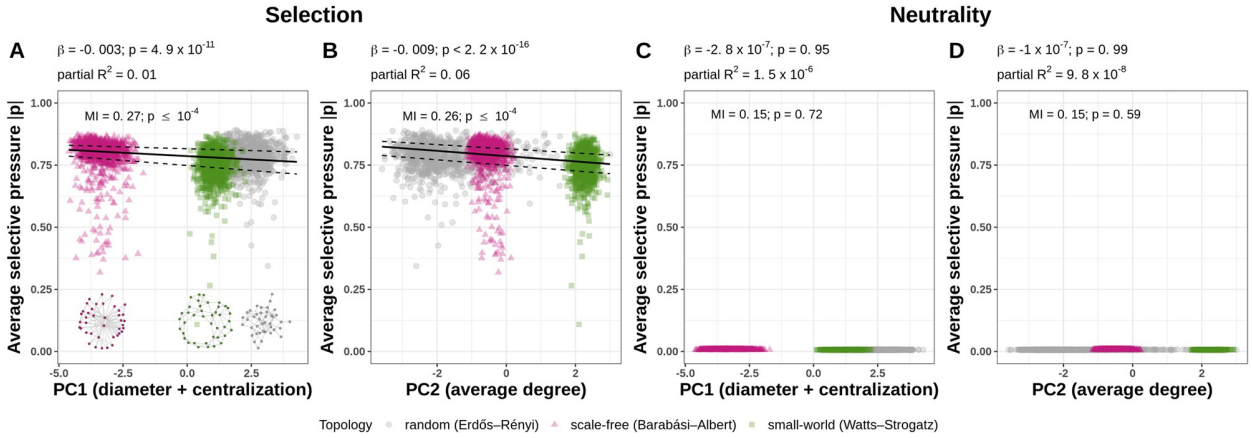

Fig 6. Global network properties affect the average selective pressure acting on gene expression noise under stabilizing selection on gene expression level. A, B—Principal component variables consisting of the diameter and network centralization (A) and average degree (B) have a significant negative effect on the average selective pressure per network. The two synthetic variables were constructed by performing a principal component analysis on 12 graph-level network metrics. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized over all replicates in each scenario. Coefficients and p-values are estimated using a linear model with average selective pressure as the response variable, and PC1 and PC2 as explanatory variables. Mutual information (MI) p-values were computed with permutation test with 10,000 permutations.

https://doi.org/10.1371/journal.pcbi.1010982.g006

had a significant effect on the average selective pressure per network (linear model with synthetic network diameter and centralization coefficient $\beta = -2.8 \times 10^{-7}$, p-value = 0.95; synthetic average degree coefficient $\beta = -1 \times 10^{-7}$, p-value = 0.99, Fig 6C and 6D). Similarly, the mutual information was insignificant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.15, p-value = 0.72, permutation test) and synthetic average degree (MI = 0.15, p-value = 0.59, permutation test).

Discussion

In this work, we aimed at understanding how natural selection shaped the distribution of expression noise levels between genes in the genome. We hypothesized that selection for low noise at the network level translates into differential selective pressures at the gene level. To test this hypothesis, we developed a new gene regulatory network evolution model that incorporates stochastic gene expression, where the gene expression mean and variance are both heritable and, therefore, potentially subject to natural selection. We simulated the evolution of gene-specific expression noise in populations of model gene regulatory networks under selective and non-selective conditions. In agreement with our hypothesis, we observed that individual genes respond differently to the global selective pressure and that this response depends on the local and global network properties. In particular, we found that genes of high centrality exhibit a stronger selective pressure to reduce gene-specific expression noise under stabilizing selection on the expression level and that the genetic network structure affects the propagation and evolvability of gene-specific expression noise. In the following, we further discuss the implications of differential selective pressure acting on constituent genes in gene networks.

Mechanisms of intrinsic noise reduction

In this study we abstracted and summarized the many determinants of intrinsic expression noise into a single parameter, which can be viewed as a modifier locus that can directly change

the intrinsic noise of a given gene. This simplification permitted us to investigate the evolution of expression noise in gene networks with computationally feasible evolutionary simulations. In reality, multiple factors that affect gene expression variance in biological systems have been reported. These include epigenetic factors, such as chromatic dynamics [44] and presence of chromatin remodelling complexes [45]. Other factors affect transcription directly and can, therefore, control expression noise: the promoter shape [36], presence of a TATA box [45], presence and number [4] of TF binding sites, TF binding dynamics [46], presence of TF decoy binding sites [47], and transcription rate. Factors affecting translation have also been shown to play a role in controlling noise: miRNA targetting [48], mRNA lifetime, translation rate, and post-translational modifications such as the protein degradation rate. Compartmentalization of proteins by phase separation has also been shown to reduce noise [49]. Lastly, gene expression costs can also affect the gene expression level distributions, and thereby expression level noise [50]. We have demonstrated the existence of a general selective pressure acting on gene expression noise. Biological organisms may differ in the mechanisms used to respond to this selective pressure, calling for further, data-driven, investigations.

Global network structure impacts noise propagation and evolution

By simulating thousands of networks with distinct structures, we were further able to assess the impact of global network characteristics on gene-specific selective pressure. Given that there is a trade-off between the fitness advantage of reducing gene-specific expression noise at the gene level and its mechanistic cost (for instance, in terms of mRNA processing [51]), evolving the global network structure may offer an alternative way to reduce network-level noise. Several motifs recurrently found in regulatory networks have an impact on expression noise, such as negative [37–39] and positive autoregulation [41], feed-forward loops [41, 52, 53] and interlinked feed-forward loops [54].

It is important, however, to distinguish two aspects when considering the effect of the network structure on the expression dynamics of constituent genes: the network structure, *i.e.* the topology of the graph, and the strength of each of the regulatory interactions, both of which impact expression noise. The same network topology, but with different regulatory interactions strengths, can give rise to markedly different network behaviours. In the *gap* gene system, for example, it was shown that multiple subcircuits share the same regulatory structure, but yield different expression patterns because of their differences in active components and strength of regulatory interactions [55]. It results that network models of gene expression noise must incorporate both graph topology and interaction strength between all constituent genes. The Wagner model constitutes a simple framework that fulfills these two conditions. However, it has its limitations. Namely, it is not fine-grained enough to capture the complex dynamics of real regulatory networks. Models that incorporate higher molecular detail, such as large systems of differential equations, are necessary to precisely capture in fine detail the expression dynamics of a real biological network, but they come with a cost in terms of high computation time (preventing their use in evolutionary simulations), low tractability and, often, the inability to model noise.

Implications of selection on expression noise on the evolution of genomes and gene regulatory networks

One mechanism by which networks and genomes evolve is gene duplication. Gene duplications are a major source of new genes and thought to be a primary source of evolutionary novelties. It has been long proposed that new functionality arises from duplicated genes by allowing the other gene copy to acquire new functions (neofunctionalization) or improve existing functions (subfunctionalization) by relaxing the selective pressure acting on a single

gene through an additional redundant copy [56]. However, most of the time the redundant copy is lost before new functionality can arise [57], either by genetic drift alone or because having the extra copy is deleterious. The redundant copy has a chance to evolve a new function or improve an existing one while it is evolving neutrally or reaches fixation in the population, or alternatively, if there is some fitness benefit of the additional copy that increases its frequency in the population. Some benefits of having additional gene copies have been shown, such as increased expression level for genes whose pre-duplication expression level was far from the optimum [58]. Moreover, duplicating a gene reduces its expression noise [59, 60], averaging the stochastic events over the two gene copies. The reduction of expression noise may, therefore, constitute another benefit of a gene duplication, increasing its chance of fixation in the population. As the gene number increases in bacterial genomes, the number of regulatory genes increases 4-fold [61], indicating a gene duplication is more likely to stay if the gene is a regulatory gene. We hypothesize that selection on expression noise, particularly on regulatory genes, could, therefore, be one of the forces driving the maintenance of duplicated genes.

Applications of the model framework to study complex systems

In this study, we developed a new regulatory and evolutionary model to study expression noise in gene regulatory networks. The model represents key features of evolving gene regulatory networks, namely the non-independence of gene expression levels and fitness determined by the expression level of many or all genes in the network. Our results revealed that differential selective pressure acts on intrinsic expression noise of constituent genes and that network-level topological properties affect noise propagation within the network.

Although our study focused on gene regulatory networks, our conclusions potentially apply to a broader range of systems. In particular, we posit that any system that fulfills two essential properties will exhibit a similar behavior: (i) the amount of product of each system component (here called "expression level") is not independent and (ii) the performance (here termed "fitness") is determined by the product level of one or several of the components of the system. There are many other complex systems that fulfill these criteria, such as biological metabolic networks, ecological food webs, neural networks, economies, transportation and other infrastructure networks, and social networks. We expect that the same constraints act on noise in elements of these systems, and that some of the conclusions from gene regulatory networks could be carefully applied to other complex systems.

Conclusion

Our results show that selection for low expression noise acting on a system (the gene network) resulted in differential selective pressures on its individual components (the genes). We demonstrated that the position of the gene in the network and the global network structure act as important drivers of the evolution of intrinsic expression noise. Investigating how gene networks evolve to cope with expression noise will reveal mechanisms of how complex biological systems adapt to function with an inevitable molecular noise in their components. A better comprehension of these mechanisms is a prerequisite to understand the evolution of complexity in biological systems, from the first self-replicating RNA systems to modern eukaryotic cells expressing tens of thousands of genes.

Supporting information

S1 Text. Gene regulatory network model and evolutionary model.
(PDF)

S2 Text. Network centrality metrics.
(PDF)

S3 Text. Diagnostics of statistical models.
(PDF)

S4 Text. Robustness of results in different topology structures.
(PDF)

S5 Text. Robustness of results to unequal fitness contribution of genes.
(PDF)

S6 Text. Filtered datasets.
(PDF)

Acknowledgments

The authors would like to thank Arnaud Le Rouzic for the beneficial discussions, Andreas Wagner and the Wagner lab for the valuable input during the model development, Arne Traulsen and Tal Dagan for the helpful suggestions throughout the project, Nikhil Sharma for his study on additional aspects of the model, and Andrea Bours, Artemis Efstratiou and Carolina Peralta for the careful reading of the manuscript.

Author Contributions

Conceptualization: Nataša Puzović, Julien Y. Dutheil.

Formal analysis: Nataša Puzović, Tanvi Madaan.

Investigation: Nataša Puzović, Julien Y. Dutheil.

Methodology: Nataša Puzović, Julien Y. Dutheil.

Software: Nataša Puzović.

Supervision: Julien Y. Dutheil.

Validation: Nataša Puzović.

Visualization: Nataša Puzović, Tanvi Madaan.

Writing – original draft: Nataša Puzović.

Writing – review & editing: Nataša Puzović, Julien Y. Dutheil.

References

1. Elowitz MB. Stochastic Gene Expression in a Single Cell. Science. 2002; 297(5584):1183–1186. https://doi.org/10.1126/science.1070919 PMID: 12183631
2. Raser JM, O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. Science (New York, NY). 2005; 309(5743):2010–2013. https://doi.org/10.1126/science.1105891 PMID: 16179466
3. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between gene expression noise and regulatory network architecture. Trends in Genetics. 2012; 28(5):221–232. https://doi.org/10.1016/j.tig.2012.01.006 PMID: 22365642
4. Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z, et al. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. Genome Research. 2014; 24 (10):1698–1706. https://doi.org/10.1101/gr.168773.113 PMID: 25030889
5. Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. Molecular Systems Biology. 2008; 4(1):170. https://doi.org/10.1038/msb.2008.11 PMID: 18319722

6. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise Minimization in Eukaryotic Gene Expression. PLoS Biology. 2004; 2(6):e137. https://doi.org/10.1371/journal.pbio.0020137 PMID: 15124029
7. Wang Z, Zhang J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proceedings of the National Academy of Sciences. 2011; 108(16):E67–E76. https://doi.org/10.1073/pnas.1100059108 PMID: 21464323
8. Barroso GV, Puzovic N, Dutheil JY. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. Genetics. 2018; 208(1):173–189. https://doi.org/10.1534/genetics.117.300467 PMID: 29097405
9. Duveau F, Hodgins-Davis A, Metzger BP, Yang B, Tryban S, Walker EA, et al. Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. eLife. 2018; 7:e37272. https://doi.org/10.7554/eLife.37272 PMID: 30124429
10. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. Nature. 2009; 462(7269):90–93. https://doi.org/10.1038/nature08504 PMID: 19890329
11. Bódi Z, Farkas Z, Nevozhay D, Kalapis D, Lázár V, Csörgő B, et al. Phenotypic heterogeneity promotes adaptive evolution. PLOS Biology. 2017; 15(5):e2000644. https://doi.org/10.1371/journal.pbio.2000644 PMID: 28486496
12. Farquhar KS, Charlebois DA, Szenk M, Cohen J, Nevozhay D, Balázsi G. Role of network-mediated stochasticity in mammalian drug resistance. Nature Communications. 2019; 10(1):2766. https://doi.org/10.1038/s41467-019-10330-w PMID: 31235692
13. Nevozhay D, Adams RM, Itallie EV, Bennett MR, Balázsi G. Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit. PLOS Computational Biology. 2012; 8(4):e1002480. https://doi.org/10.1371/journal.pcbi.1002480 PMID: 22511863
14. Schmiedel JM, Carey LB, Lehner B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. Nature Communications. 2019; 10(1):3180. https://doi.org/10.1038/s41467-019-11116-w PMID: 31320634
15. Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression. Trends in Genetics. 2006; 22(8):456–461. https://doi.org/10.1016/j.tig.2006.06.002 PMID: 16806568
16. Vlková M, Silander OK. Gene regulation in Escherichia coli is commonly selected for both high plasticity and low noise. Nature Ecology & Evolution. 2022; p. 1–15. PMID: 35726087
17. Pedraza JM. Noise Propagation in Gene Networks. Science. 2005; 307(5717):1965–1969. https://doi.org/10.1126/science.1109090 PMID: 15790857
18. Blake WJ, KÆrn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003; 422(6932):633–637. https://doi.org/10.1038/nature01546 PMID: 12687005
19. Hens C, Harush U, Haber S, Cohen R, Barzel B. Spatiotemporal signal propagation in complex networks. Nature Physics. 2019; 15(4):403–412. https://doi.org/10.1038/s41567-018-0409-0
20. Wagner A. Does Evolutionary Plasticity Evolve? Evolution. 1996; 50(3):1008–1023. https://doi.org/10.1111/j.1558-5646.1996.tb02342.x PMID: 28565284
21. Newman M. Networks: An Introduction. Oxford: Oxford University Press; 2010. Available from: https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650.
22. Laarits T, Bordalo P, Lemos B. Genes under weaker stabilizing selection increase network evolvability and rapid regulatory adaptation to an environmental shift. Journal of Evolutionary Biology. 2016; 29(8):1602–1616. https://doi.org/10.1111/jeb.12897 PMID: 27213992
23. Pinho R, Borenstein E, Feldman MW. Most Networks in Wagner's Model Are Cycling. PLoS ONE. 2012; 7(4):e34285. https://doi.org/10.1371/journal.pone.0034285 PMID: 22511935
24. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. Proceedings of the National Academy of Sciences. 2014; 111(45):16219–16224. https://doi.org/10.1073/pnas.1408886111 PMID: 25349387
25. Team RC. R: A Language and Environment for Statistical Computing; 2021. Available from: https://www.R-project.org/.
26. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, complex systems. 2006; 1695(5):1–9.
27. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Journal of statistical software. 2008; 24(3):nihpa54860. https://doi.org/10.18637/jss.v024.i03 PMID: 19756229
28. Dray S, Dufour AB. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software. 2007; 22:1–20. https://doi.org/10.18637/jss.v022.i04

29. Pinheiro J, Bates D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models; 2022. Available from: https://CRAN.R-project.org/package=nlme.
30. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015; 67:1–48. https://doi.org/10.18637/jss.v067.i01
31. Bartoń K. MuMIn: Multi-Model Inference; 2020. Available from: https://CRAN.R-project.org/package=MuMIn.
32. Fox J, Weisberg S. An R Companion to Applied Regression. 3rd ed. Thousand Oaks CA: Sage; 2019. Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
33. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics. New York, NY: Springer US; 2013. Available from: https://link.springer.com/10.1007/978-1-0716-1418-1.
34. Meyer PE. infotheo: Information-Theoretic Measures; 2014. Available from: https://cran.r-project.org/package=infotheo.
35. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, Nimwegen Ev. Genome-wide gene expression noise in Escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network. PLOS Biology. 2021; 19(12):e3001491. https://doi.org/10.1371/journal.pbio.3001491 PMID: 34919538
36. Sigalova OM, Shaeiri A, Forneris M, Furlong EE, Zaugg JB. Predictive features of gene expression variation reveal mechanistic link with differential expression. Molecular Systems Biology. 2020; 16(8). https://doi.org/10.15252/msb.20209539 PMID: 32767663
37. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature. 2000; 405 (6786):590–593. https://doi.org/10.1038/35014651 PMID: 10850721
38. Dublanche Y, Michalodimitrakis K, Kümmerer N, Foglierini M, Serrano L. Noise in transcription negative feedback loops: simulation and experimental analysis. Molecular Systems Biology. 2006; 2(1):41. https://doi.org/10.1038/msb4100081 PMID: 16883354
39. Grönlund A, Lötstedt P, Elf J. Transcription factor binding kinetics constrain noise suppression via negative feedback. Nature Communications. 2013; 4(1):1864. https://doi.org/10.1038/ncomms2867 PMID: 23673649
40. Rosenfeld N, Elowitz MB, Alon U. Negative Autoregulation Speeds the Response Times of Transcription Networks. Journal of Molecular Biology. 2002; 323(5):785–793. https://doi.org/10.1016/S0022-2836(02)00994-4 PMID: 12417193
41. Alon U. Network motifs: theory and experimental approaches. Nature Reviews Genetics. 2007; 8 (6):450–461. https://doi.org/10.1038/nrg2102 PMID: 17510665
42. Oikonomou P, Cluzel P. Effects of topology on network evolution. Nature Physics. 2006; 2(8):532–536. https://doi.org/10.1038/nphys359
43. Greenbury SF, Johnston IG, Smith MA, Doye JPK, Louis AA. The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. Journal of Theoretical Biology. 2010; 267 (1):48–61. https://doi.org/10.1016/j.jtbi.2010.08.006 PMID: 20696172
44. Sun M, Zhang J. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. PLOS Genetics. 2019; 15(9):e1008389. https://doi.org/10.1371/journal.pgen.1008389 PMID: 31525198
45. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006; 441(7095):840–846. https://doi.org/10.1038/nature04785 PMID: 16699522
46. Azpeitia E, Wagner A. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System. Frontiers in Molecular Biosciences. 2020; 7:67. https://doi.org/10.3389/fmolb.2020.00067 PMID: 32411721
47. Dey S, Soltani M, Singh A. Enhancement of gene expression noise from transcription factor binding to genomic decoy sites. Scientific Reports. 2020; 10(1):9126. https://doi.org/10.1038/s41598-020-65750-2 PMID: 32499583
48. Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, et al. MicroRNA control of protein expression noise. Science. 2015; 348(6230):128–132. https://doi.org/10.1126/science.aaa1738 PMID: 25838385
49. Klosin A, Oltsch F, Harmon T, Honigmann A, Jülicher F, Hyman AA, et al. Phase separation provides a mechanism to reduce noise in cells. 2020; p. 6.
50. Charlebois DA. Effect and evolution of gene expression noise on the fitness landscape. Physical Review E. 2015; 92(2):022713. https://doi.org/10.1103/PhysRevE.92.022713 PMID: 26382438

51. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. Nature Communications. 2019; 10(1):68. https://doi.org/10.1038/s41467-018-07391-8 PMID: 30622246

52. Charlebois DA, Balázsi G, Kærn M. Coherent feedforward transcriptional regulatory motifs enhance drug resistance. Physical Review E. 2014; 89(5):052708. https://doi.org/10.1103/PhysRevE.89.052708 PMID: 25353830

53. Camellato B, Roney IJ, Azizi A, Charlebois D, Kaern M. Engineered gene networks enable non-genetic drug resistance and enhanced cellular robustness. Engineering Biology. 2019; 3(4):72–79. https://doi.org/10.1049/enb.2019.0009

54. Chepyala SR, Chen YC, Yan CCS, Lu CYD, Wu YC, Hsu CP. Noise propagation with interlinked feed-forward pathways. Scientific Reports. 2016; 6(1):23607. https://doi.org/10.1038/srep23607 PMID: 27029397

55. Verd B, Monk NA, Jaeger J. Modularity, criticality, and evolvability of a developmental gene regulatory network. eLife. 2019; 8:e42832. https://doi.org/10.7554/eLife.42832 PMID: 31169494

56. Ohno S. Evolution by Gene Duplication. Springer Berlin, Heidelberg; 1970.

57. Lynch M, Conery JS. The Evolutionary Fate and Consequences of Duplicate Genes. Science. 2000; 290(5494):1151–1155. https://doi.org/10.1126/science.290.5494.1151 PMID: 11073452

58. Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in Escherichia coli. Proceedings of the National Academy of Sciences. 2001; 98(2):525–530. https://doi.org/10.1073/pnas.021448998 PMID: 11149947

59. Rodrigo G, Fares MA. Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach. eLife. 2018; 7:e29739. https://doi.org/10.7554/eLife.29739 PMID: 29303479

60. Chapal M, Mintzer S, Brodsky S, Carmi M, Barkai N. Resolving noise–control conflict by gene duplication. PLOS Biology. 2019; 17(11):e3000289. https://doi.org/10.1371/journal.pbio.3000289 PMID: 31756183

61. Molina N, van Nimwegen E. The evolution of domain-content in bacterial genomes. Biology Direct. 2008; 3:51. https://doi.org/10.1186/1745-6150-3-51 PMID: 19077245

Appendix C

CV

Nataša Puzović

Email: npuzovic@protonmail.com
Personal webpage
GoogleScholar ResearchGate Github Orcid

Formal education

- MSc Molecular Biology and Evolution, Christian-Albrechts-Universität zu Kiel, Germany (2017 – 2019)
- BSc Molecular Biology and Physiology, Faculty of Biology, University of Belgrade, Serbia (2012 – 2017)

Publications

- **Puzović, N.**, Dutheil, J. Y. (2023). The evolution of gene expression mean and noise in changing environments is constrained by the gene position in the gene regulatory network. *In prep.*
- Ernst, E. , Nisioti, E., Clark, C., Kunal Das, K., Friedenberg, N., Gates, E., Lambros, M., Lazurko, A., **Puzovic, N.** Salas, I. (2023). Resilience - A complex-systems perspective. Submitted to Frontiers in Complex Systems.
- **Puzović, N.**, Madaan, T. Dutheil, J. Y. (2023). Being noisy in a crowd: Differential selective pressure on gene expression noise in model gene regulatory networks. PLOS Computational Biology, doi:10.1101/2022.08.01.502352
- V. Barroso, G., **Puzović, N.** Dutheil, J. Y. (2019). Inference of recombination maps from a single pair of genomes and its application to ancient samples. PLOS Genetics, 15(11), 1–21. doi:10.1371/journal.pgen.1008449
- Barroso, G. V., **Puzovic, N.** Dutheil, J. Y. (2018). The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. Genetics, 208(1), 173–189. doi:10.1534/genetics.117.300467

Selected conferences during the PhD

- 21st International Conference on Systems Biology, poster presentation (see poster), Berlin, Germany, 08-12 Oct 2022
- Evolutionary Systems Biology, virtual poster presentation, Wellcome Genome Campus, UK, 09-11 Feb 2022
- 55th annual Population Genetics Group Meeting, oral presentation (see talk recording), Norwich, UK, 05-07 Jan 2022
- Society for Molecular Biology & Evolution Annual Meeting (SMBE), virtual poster presentation, Society for Molecular Biology & Evolution, 03-08 Jul 2021
- CSHL Probabilistic Modeling in Genomics, virtual poster presentation, Cold Spring Harbor Laboratory, New York, USA, 14-16 Apr 2021
- CSHL Network Biology Meeting, virtual poster presentation, Cold Spring Harbor Laboratory, New York, USA, 16-19 Mar 2021

Supervision Experience

- Barbara D'Albis, Undergraduate Internship, Apr – May 2022
- Nikhil Sharma, IMPRS PhD rotation, Nov – Dec 2020
- Tanvi Madaan, Undergraduate Internship, Jun – Aug 2020