Under-Dominance Constrains the Evolution of Negative Autoregulation in Diploids

Alexander J. Stewart[1,2]*, Robert M. Seymour[2,3], Andrew Pomiankowski[2,4], Max Reuter[4]

1 Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 2 CoMPLEX, University College London, London, United Kingdom, 3 Department of Mathematics, University College London, London, United Kingdom, 4 Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

Abstract

Regulatory networks have evolved to allow gene expression to rapidly track changes in the environment as well as to buffer perturbations and maintain cellular homeostasis in the absence of change. Theoretical work and empirical investigation in *Escherichia coli* have shown that negative autoregulation confers both rapid response times and reduced intrinsic noise, which is reflected in the fact that almost half of *Escherichia coli* transcription factors are negatively autoregulated. However, negative autoregulation is rare amongst the transcription factors of *Saccharomyces cerevisiae*. This difference is surprising because *E. coli* and *S. cerevisiae* otherwise have similar profiles of network motifs. In this study we investigate regulatory interactions amongst the transcription factors of *Drosophila melanogaster* and humans, and show that they have a similar dearth of negative autoregulation to that seen in *S. cerevisiae*. We then present a model demonstrating that this stiking difference in the noise reduction strategies used amongst species can be explained by constraints on the evolution of negative autoregulation in diploids. We show that regulatory interactions between pairs of homologous genes within the same cell can lead to under-dominance — mutations which result in stronger autoregulation, and decrease noise in homozygotes, paradoxically can cause increased noise in heterozygotes. This severely limits a diploid's ability to evolve negative autoregulation as a noise reduction mechanism. Our work offers a simple and general explanation for a previously unexplained difference between the regulatory architectures of *E. coli* and yeast, *Drosophila* and humans. It also demonstrates that the effects of diploidy in gene networks can have counter-intuitive consequences that may profoundly influence the course of evolution.

Introduction

Negative autoregulation is a network motif in which a transcription factor inhibits its own expression. Theoretical work has shown that this type of regulation reduces intrinsic noise and quickens the response time to environmental perturbations [1–3] and experiments using artificial gene regulatory circuits in *E. coli* have confirmed these predictions [2]. Negative autoregulation therefore represents a simple yet powerful mechanism to maintain cellular homeostasis in the face of environmental and metabolic perturbations and reduce the often substantial fitness costs that noise can incur [4]. Different organisms, however, vary a great deal in their use of the motif. In *E. coli*, close to 50% of transcription factors (82 out of 182) [5–8] have been shown to negatively autoregulate. In contrast, negative autoregulation is almost entirely absent amongst the transcription factors that have been studied in *S. cerevisiae* (3 out of 169) [6,8–11].

How can we account for this discrepancy? In order to answer this, we looked at the extent to which negative autoregulation is used in other species. We interrogated systematic datasets on the regulatory interactions amongst the known transcription factors of

D. melanogaster and humans and found a similar pattern to that observed in yeast: in *D. melanogaster* 3 out of 87 [12–14] and in humans 5 out of 301 [13–15] transcription factors negatively autoregulate (see SI, Table S1, S2, S3). Currently, there is no obvious way to account for this striking discrepancy between these organisms, despite widespread interest in the strategies they employ to tackle noise [1–4,16–18]. Here we develop a model, founded in biophysics, for the evolution of negative autoregulation in diploid species. We use it to support the hypothesis that a dearth of negatively autoregulating genes in yeast, flies and humans can be explained by constraints on the evolution of negative autoregulation that arise due to diploidy.

Results

Gene expression under negative autoregulation

Previous theoretical work on the dynamics of gene expression under negative autoregulation has considered single genes and so is implicitly haploid [1–3,18]. Such models exclude the more complex interactions that occur due to cross-regulation between homologous gene copies within a diploid cell (Fig. 1). Here we

Author Summary

All genes have to deal with intrinsic noise, and a variety of mechanisms have evolved to reduce it. One important mechanism of noise reduction for transcription factors is negative autoregulation, in which a gene product represses its own rate of transcription. Negative auotregulation occurs frequently in *E. coli* but, we find, occurs much more rarely in *S. cerevisiae*, *D. melanogaster* and humans. Whilst there are a great many important differences in the genetic architectures of these organisms, they tend to share, with the exception of negative autoregulation, similar profiles of network motifs. This makes the discrepancy in the degree of negative autoregulation all the more striking, as it lacks any obvious explanation. Our study presents a potential explanation, by comparing the evolvability of negative autoregulation as a noise reduction mechanism in haploids and diploids. We show that, in diploids, mutations that increase the strength of negative autoregulation at one gene copy often increase overall noise in gene expression. This results in under-dominance, in which heterozygotes are less fit than homozygotes. The result is that the evolution of negative autoregulation in diploids is significantly constrained. We verify our results using a combination of detailed molecular simulations and evolutionary simulations

characterise the expression dynamics and regulatory evolution of homologous pairs of negatively autoregulating genes, taking into account the cross-talk between alleles.

We model negative autoregulation in a diploid using a set of ordinary differential equations that track changes in the mRNA and protein concentrations for each of a pair of alleles (labelled with subscripts 1 and 1), r_1, r_2, p_1 and p_2. The total concentration of mRNA and protein in the diploid cell are given by the summed output of the two alleles $r = r_1 + r_2$ and $p = p_1 + p_2$. Changes in mRNA and protein concentrations for the pair of alleles over time are given by

$$
\begin{aligned}
\frac{dr_1}{dt} &= k_l + \phi_1(p) - \gamma_r r_1, \\
\frac{dr_2}{dt} &= k_l + \phi_2(p) - \gamma_r r_2, \\
\frac{dp_1}{dt} &= r_1 k_p - \gamma_p p_1, \\
\frac{dp_2}{dt} &= r_2 k_p - \gamma_p p_2.
\end{aligned}
\tag{1}
$$

According to these equations, mRNA is transcribed at a (usually low) constant background rate k_l, plus a rate $\phi(p)$ due to negative autoregulation, that decreases as the total cellular protein level $p = p_1 + p_2$ increases. Protein is produced from mRNA at the rate of translation k_p, whilst protein and mRNA degrade with rates γ_p and γ_r, respectively.

As in previous work [1,2], we model the repression function $\phi(p)$ in Eqs. 1 as a Hill function

$$
\phi_i(p) = \frac{k_0}{1 + \left(\frac{p}{K_i}\right)^n}
$$

where K_i is the dissociation constant associated with the autoregulating transcription factor binding site. Smaller values of K (lower rates of dissociation) indicate stronger regulation. The Hill coefficient n governs the steepness of the function at the inflection point and hence determines how step-like regulation will be. In systems where transcription is regulated by a single binding site, $\phi(p)$ has a Michaelis-Menten-like form, corresponding to a Hill coefficient of $n = 1$ [2,19,20]. A single binding site is the simplest, and perhaps the most relevant case for evolving negative autoregulation, and it is the one we focus on here. We analyse the more general case of arbitrary Hill coefficient in the Methods and in the SI we show that our results also hold for different values of n.
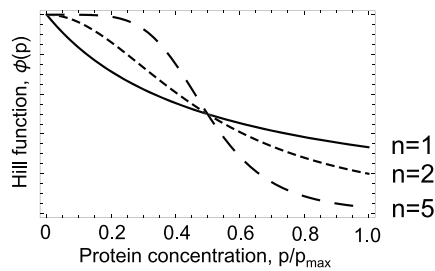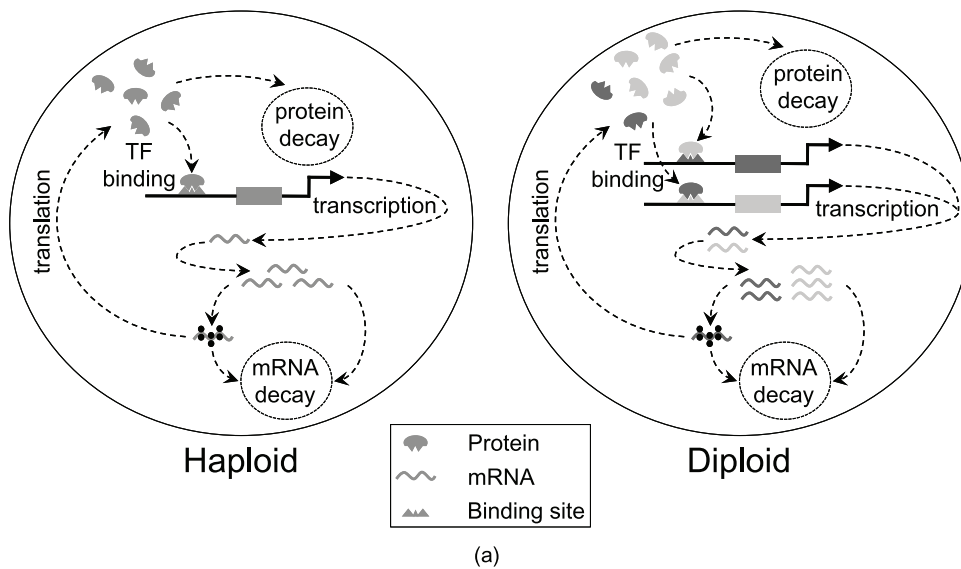
In the absence of negative autoregulation (i.e., $\phi(p) = k_0$), mRNA is produced at the maximum rate of transcription $k_l + k_0$. In this case, concentrations of mRNA and protein reach equilibrium values of $r_{max} = \frac{2(k_0 + k_l)}{\gamma_r}$ and $p_{max} = r_{max}\frac{k_p}{\gamma_p}$. Starting from these values, equilibrium mRNA and protein levels decrease with increasing autoregulatory binding strength (decreasing K). The minimum mRNA and protein levels are reached when negative autoregulation is strongest (i.e. $\phi(p) \to 0$ as $K \to 0$). The resulting minimum equilibrium concentrations are $r_{min} = \frac{2k_l}{\gamma_r}$ and $p_{min} = r_{min}\frac{k_p}{\gamma_p}$.

Evolution of negative autoregulation for homeostasis and faster response times

In order to analyse the evolution of autoregulatory binding sites we consider two separate but related functions of negative autoregulation: faster response times and maintaining mRNA and protein homeostasis. First, to study the evolution of negative autoregulation for faster response times, we simply equate the fitness of a system with its response time (i.e the time taken to return to equilibrium following a perturbation). We use Eqs. 1 to infer selection pressures on the strength of autoregulation, i.e., the dissociation constant K, by analysing how quickly genotypes with different autoregulatory binding strength return to equilibrium following a perturbation in protein level. To do this we calculate a genotype's "response time": the time taken for cellular protein concentration to return to equilibrium following a perturbation. We model perturbations as a reduction of the protein level to a fraction α of the equilibrium level. The value of α varies continuously between 0 and 1 to encompass both small perturbations, for example those resulting from intrinsic noise in transcription and translation ($\alpha \approx 1$), and larger perturbations, for example those resulting from resource deprivation in the environment or following cell division [17]. We present results derived from numerical analysis of Eqs. 1 that are applicable to perturbations of any size. These are complemented with an analytical treatment of the response time of the system to small perturbations, based on its maximal eigenvalue (see Methods), which allows us to develop an intuition for how autoregulating genes in diploids respond to perturbations.

To study the evolution of negative autoregulation for homeostasis, we turn to stochastic simulations of negatively autoregulating genes, which allow us to assess the amount of intrinsic noise associated with gene expression. Previous work has shown that negative autoregulation can help maintain homeostasis in gene expression by reducing the amount of intrinsic noise in negatively autoregulating genes, compared to other genes [3]. In fact, reducing the response time of a gene to very small perturbations away from equilibrium, also decreases the intrinsic noise in gene expression. Therefore, the two functions of negative autoregulation we consider (producing faster response times and reduced intrinsic noise) are

Figure 1. Cross-talk in diploid autoregulators. (a) Schematic representation of negative autoregulation when one (left) and two (right) copies of a gene are present in a cell. In the haploid the amount of negative autoregulation the gene experiences depends on on its own expression level. In the diploid, two gene copies are present (shown as light gray and dark gray), and the amount of negative autoregulation experienced by each gene depends on the expression level of *both* genes combined. If the two gene copies differ from one another in the strength of their transcription factor binding sites, complex dynamics can arise that are not observed in haploids. (b) Illlustration of variation in the repression function, $\phi(p)$, with protein concentration for different Hill coefficients, $n=1$ (sold line), $n=2$ (small dashes) and $n=5$ (large dashes).
doi:10.1371/journal.pcbi.1002992.g001

highly inter-related. To study the evolution of negative autoregulation for lower intrinsic noise, we equate the fitness of the system with the amount of intrinsic noise it displays (i.e the ratio of the variance in gene expression to the mean gene expression level). We infer selection pressures on the strength of autoregulation, i.e., the dissociation constant K, by the intrinsic noise of genotypes with different autoregulatory binding strengths. These are determined by performing Monte Carlo simulations for a full, molecular model of transcription, translation and autoregulation (see Methods).
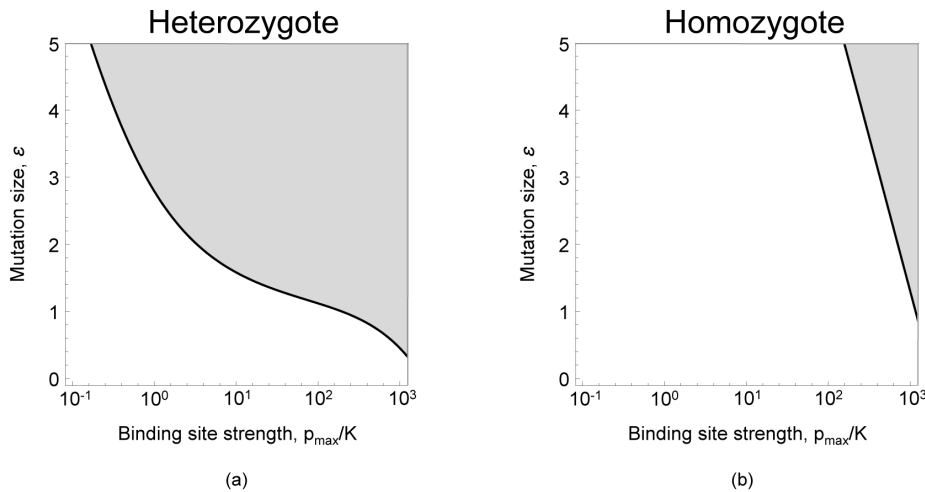
Response time in homozygotes

We first compare the response times of two homogozytes whose alleles are identical in every respect except for the dissociation constant. One homozygote carries two copies of a resident allele with dissociation constant K_1, the other carries two mutant alleles that have a decreased dissociation constant $K_2 = K_1 \exp[-\varepsilon]$ (with $\varepsilon > 0$) and hence stronger autoregulatory binding. Numerical analysis of the system shows that homozygotes for the more strongly autoregulating allele (with K_2) respond more quickly than homozygotes for the more weakly autoregulating allele (with K_1, Fig. 2a). This is true up to a value of $K \approx K_{opt}$, which provides the fastest response time attainable by the system and hence provides the optimal binding strength. Further increases in regulation beyond this value are not favoured and

lead to overshooting the optimal binding strength. These results for diploid homozygotes mirror those obtained for haploids [2] (see Methods) and show that regulatory interactions between pairs of identical alleles do not, in themselves, diminish the beneficial effects of negative autoregulation. Negative autoregulation can therefore, in principle, function as a mechanism to produce faster response times in diploids just as it does in haploids.

Response time in heterozygotes

The results above depend on comparing homozygotes for alleles with different dissociation constants, K_1 and K_2. The evolution of negative autoregulation, however, must occur through the stepwise accumulation of new mutations that are initially rare and found only in heterozygotes. In order to assess whether autoregulation can evolve in diploids, we therefore need to determine whether a mutant allele with a stronger binding site (K_2) will confer a selective advantage to a heterozygote that also carries a resident allele with a weaker binding site (K_1). A mutation will be favoured and increase in frequency if a heterozygote is able to respond more quickly to perturbations than a homozygote carrying two copies of the more weakly binding resident allele.

Numerical analysis of Eqs. 1 reveals that heterozygotes often have greater response times than homozygotes with the more weakly binding resident allele. Fig. 2b shows that heterozygotes

Figure 2. Invasibility of autoregulatory binding sites. The response time of mutant (a) homozygotes and (b) heterozygotes are shown. Different values of the binding strength of the resident allele, in units of p_{max}/K (x-axis), are plotted against mutations to binding site strength ε of different size (y-axis). Thus the graphs compare a resident allele, K_1 with a mutant allele, $K_2 = K_1 \exp[-\varepsilon]$. Mutations falling into white region result in decreased response time in the carrier compared to resident genotype and are favoured by selection; mutations falling into the gray region result in increased response time and are not favoured by selection. Weak binding occurs when $p_{max}/K \gtrsim 10^0$ [1,2]. Response times were calculated by numerically integrating Eq. 1 from zero protein concentration to 90% of the equilibrium. The optimal binding strength in these graphs is $p_{max}/K = 1250$, corresponding to a background transcription rate $k_l/\gamma_p = 10^{-3}$.
doi:10.1371/journal.pcbi.1002992.g002

only have improved response times when the resident allele binding strength is weak ($K \gtrsim p_{max}$ [1,2]) or if the effect of a mutation that increases binding strength is small (ε is small). As the resident allele binding strength increases (i.e. p_{max}/K increases) an ever larger range of mutation sizes result in increased heterozygote response times (Fig. 2b), resulting in under-dominance (i.e. heterozygote disadvantage). Typicaly mutation sizes ε for transcription factor binding sites are in the range $1 < \varepsilon < 3$, [20–22]. In this range regulatory mutations are subject to under-dominance even when the resident allele has relatively weak binding strength, and increasingly so as the binding strength of the resident allele increases. As a consequence, the maximum binding strength that can evolve is likely to be significantly lower than in haploids (Fig. 2). Based on these results, we expect under-dominance to pose a significant barrier to the evolution of negative autoregulation in diploids.

To better understand why under-dominance arises in this system, we calculated the eigenvalues associated with Eqs. 1. These provide a measure of the rate at which the system returns to equilibrium following a small perturbation, and allow us to elucidate the relative contributions of the different alleles to the response dynamics of the gene pair. The maximal eigenvalue of Eqs. 1 for a heterozygote, $|\lambda_{het}|$, can be expressed as

$$|\lambda_{het}| = |\lambda_{hom}| - \frac{V}{p_{het}^*}, \qquad (2)$$

(see Methods) where V is the squared difference of the mean steady state expression levels of the two alleles in the heterozygote and $|\lambda_{hom}|$ is the maximal eigenvalue of a homozygote with protein concentration equal to that of the heterozygote at equilibrium, p_{het}^* (see Methods). Eq. 2 says that, even if increasing autoregulatory binding strength leads to a faster response time in a homozygote, this advantage is offset in the heterozygote by an amount V/p_{het}^*, which measures how different the expression levels of the two alleles are (it is analogous to the Fano factor, a measure of the

spread in a probability distribution [3]). As the difference in the expression of the alleles increases, V/p_{het}^* increases from 0 to a maximum $p_{het}^*/2$.

We can understand why increasing the difference in allelic expression results in increased response time by considering the contribution of the individual alleles to the response time of the gene pair (Fig. 3). The level of negative autoregulation at each allele depends on the strength of its binding site and the amount of protein product present in the cell. In a heterozygote, the allele with the stronger binding site is more strongly suppressed (compared to the same allele in a homozygote), since there is more protein available to bind to it. At the same time, the allele with the weaker binding site is less strongly suppressed compared to the same allele in a homozygote. As a result, the allele with the stronger binding site has a faster response time than in a homozygote, whilst the allele with the weaker binding site has a slower response time than in a homozygote. However, the overall effect tends to be to increase the response time of the heterozygote, because the dynamics of protein expression in the heterozygote are dominated by the allele with the weaker binding site (Fig. 3).

Evolution of faster response times

Under-dominance for response time occurs across a wide range of parameter values, but can be avoided if mutations have small effects on binding site strength (Fig. 2b). To determine whether a series of mutations with small effect could offer a feasible way for genes to evolve strong negative autoregulation in diploids, we carried out simulations of binding site evolution that incorporated established properties of real binding sites.

Transcription factor binding sites in eukaryotes vary between 5 and ~ 30 nucleotides in length, with an average of 10 nucleotides [23]. They have a small number of optimal sequences that bind the transcription factor with maximum affinity [20–22,24,25]. The binding strength of a site can be expressed as a function of the total binding energy E of its sequence, so $K = \exp[-E]$. This total binding energy is generated by the additive contributions of individual nucleotides to overall binding, $E = \sum_i \varepsilon_i$. Individual
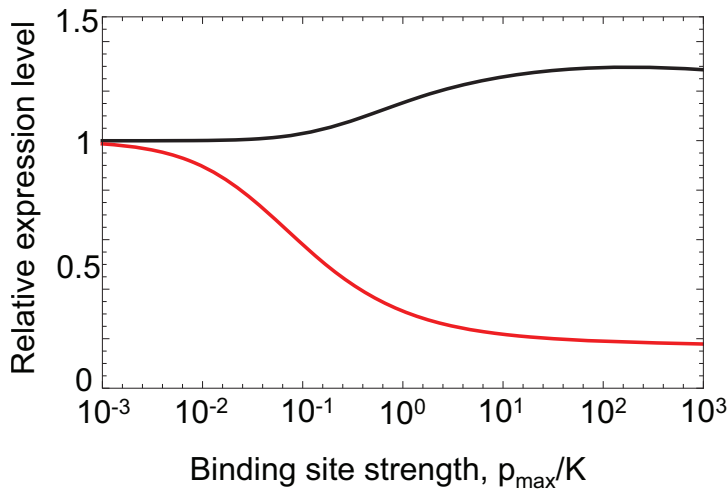
(a)
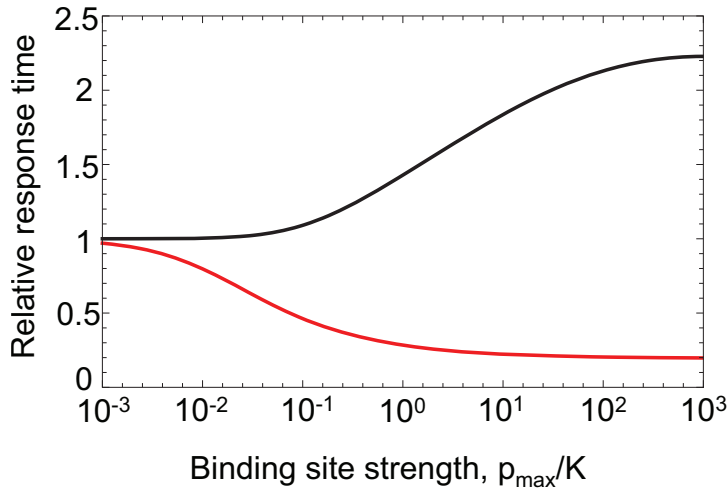
(b)

**Figure 3. Response times and allele expression.** This figure shows quantitative results for the contributions of different alleles to expression and to response time. (a) Expression level of the resident allele (black line) and the mutant allele (red line) in the heterozygote relative to the resident allele in the homozygote. As binding strength increases the resident allele is over-expressed. (b) Response times for individual alleles (time to return to 90% of the equilibrium expression level) in the heterozygote. The response time of the resident allele (black line) and the mutant allele (red line) in the heterozygote are shown relative to the response time of the resident allele in the homozygote. The resident allele in the heterozygote shows an increased response time with increasing binding strength. Mutant alleles in these graphs have dissociation constant $K \exp[-2]$, and the optimal binding strength in these graphs is $p_{max}/K = 1250$, corresponding to a background transcription rate $k_l/\gamma_p = 10^{-3}$.
doi:10.1371/journal.pcbi.1002992.g003

contributions are set to $\varepsilon_i = 0$ for nucleotides that do not match the optimal sequence and $\varepsilon_i > 0$ for matched nucleotides [20–22].

Based on these properties, we performed simulations of the evolution of an autoregulatory binding site under selection for decreased response time. These took into account the empirical distribution of binding site length in model eukaryotes and the variation in contributions to binding strength ε_i across the binding site sequence (see Methods). The values of ε_i were drawn from a uniform distribution in the interval (0,3). This sampling covers the empirically estimated range $1 < \varepsilon_i < 3$ [20–22]. It also ensures that mutations of small effect ($\varepsilon < 1$) occur frequently and so allows for the possibility that autoregulation could evolve via the accumulation of mutations with small effect. Evolution was started from a state of minimum affinity (all nucleotides non-optimal) and

proceeded through a series of single nucleotide substitutions. A mutant was assumed to go to fixation if it resulted in a response time less than or equal to that of the resident. Simulations were carried out for both haploids and diploids (for which the response time of mutants was evaluated in the heterozygote state).

The results (Fig. 4) confirm that under-dominance strongly constrains the evolution of negative autoregulation in diploids. Haploids readily evolved binding sites with dissociation constants close to K_{opt}. In contrast, the average binding strength in diploids was around 100 times weaker than K_{opt} and only a small proportion of sites reached binding strengths comparable to those of haploids. This shows that under realistic conditions, diploids will rarely be able to evolve the level of autoregulation observed in haploids.
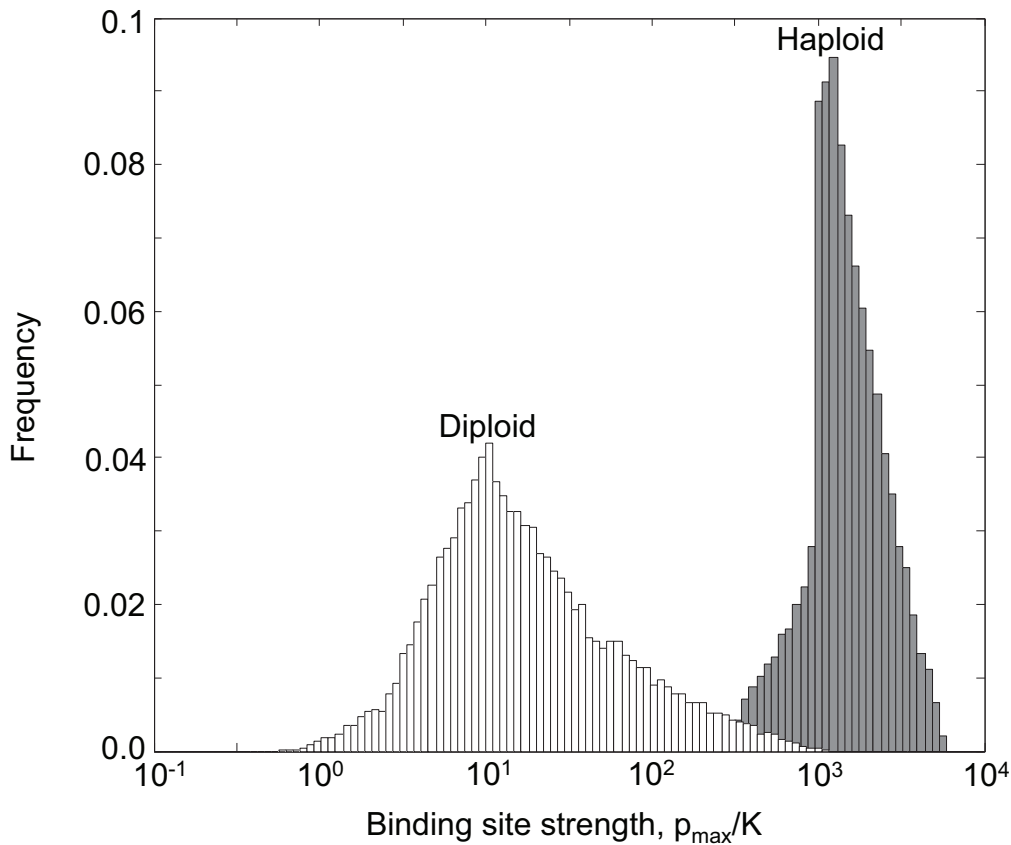
Figure 4. Evolution of autoregulatory binding sites. Distribution of binding site strength achieved in evolutionary simulations for haploids (gray) and diploids (white). Hapoids are able to evolve stronger binding than diploids. The histograms shows results of 10^5 replicate simulations for each ploidy level. The simulation procedure is described in the main text and the Materials and Methods. The optimal binding strength used was $p_{max}/K = 1250$, corresponding to a a background transcription rate $k_l/\gamma_p = 10^{-3}$.
doi:10.1371/journal.pcbi.1002992.g004
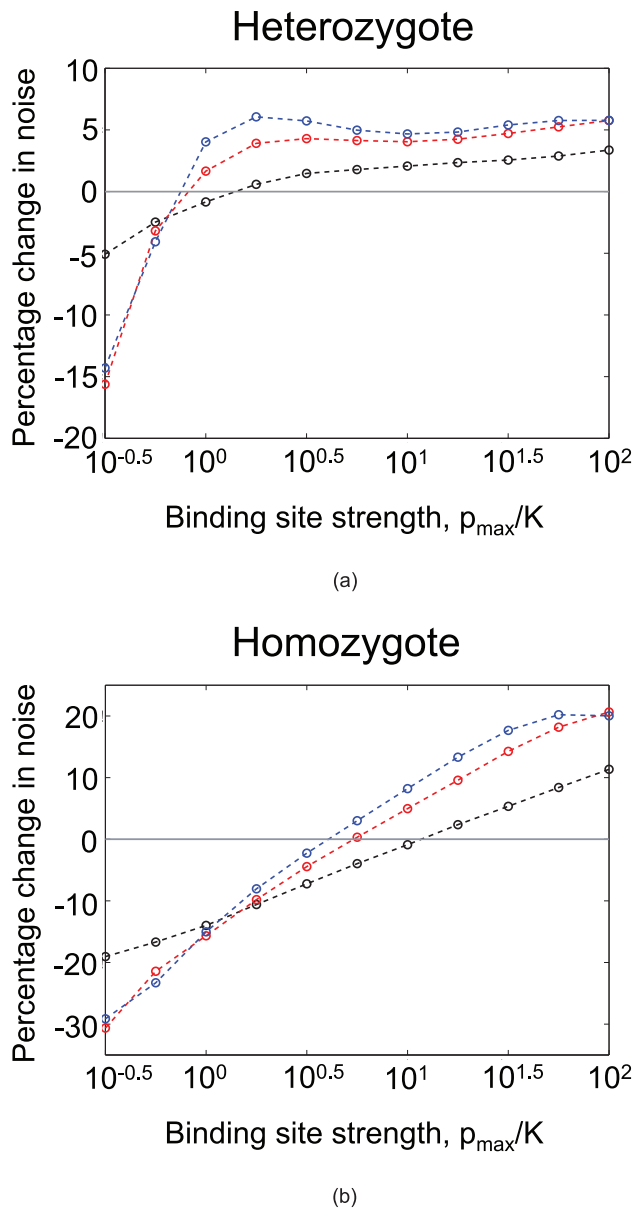
Intrinsic noise in diploids

In order to investigate the evolution of negative autoregulation as a mecahnism to reduce intrinsic noise in diploids, we turned to stochastic simulations. Intrinsic noise in gene expression occurs because transcription and translation are inherently noisy processes: all genes experience constant fluctuations in their mRNA and protein levels. The greater intrinsic noise associated with a particular gene, the higher the variance in its expression level relative to the mean. Therefore, a natural way to characterise the amount of intrinsic noise associated with a gene is to measure the ratio of the variance to the mean expression level at equilibrium (known as the Fano factor) [3]. We performed molecular simulations that capture transcription, translation and degradation in the presence of negative autoregulation (see Materials and Methods). Just as in our analysis of response times, we compared a resident allele with dissociation constant K_1, to a mutant allele with dissociation constant $K_2 = K_1 \exp[-\varepsilon]$. We compared the intrinsic noise (as measured by the Fano factor) in the resident homozygote to that of the heterozygte and the mutant homozygote, and thus determined whether under-dominance occurs in the evolution of negative autoregulation as a mechanism to reduce intrinsic noise. The results are shown in Fig. 5. We find once again that under-dominance occurs. Whereas the optimal binding strength for a single negatively autoregulating binding site is found to be $p_{max}/K \sim 10$, the maximum evolvable binding strength (i.e that which can evolve without encountering under-dominance) is found to be $p_{max}/K \sim 1$, an order of magnitude

weaker. A similar pattern occurs when steeper Hill coefficients are considered (Fig. 5). Therefore we conclude that under-dominance poses a barrier to the evolution of strong negative autoregulation both as a mechanism to speed response times and to reduce intrinsic noise.

The effects of mutations to other parameters

To test the generality of our findings, we also considered variation in other parameters (see SI Fig. S1, S2, S3, S4, S5 and Text S1). We first relaxed our assumption of a single binding site and explored the case of Hill coefficients $n > 1$, implying regulation through multiple, cooperatively acting binding sites. In line with the effect of increasing binding strength through changes in K, we find that mutations increasing the Hill coefficient are subject to under-dominance (see SI Fig. S1, S2 and Text S1). Therefore, a mutation that increases the strength of negative autoregulation is subject to the same evolutionary constraints, independent of whether they increase regulation by changing the dissociation constant K or the Hill coefficient n.

We also considered variation in the rates of mRNA and protein degradation (γ_r and γ_p) to see whether they provide conditions in which the effects of under-dominance on autoregulatory binding strength can be avoided (see SI Fig. S4 and Text S1). Variation in the rate of mRNA or protein degradation did not remove the tendency for mutations that increase autoregulatory binding strength to be subject to under-dominance. However, as has been pointed out elsewhere [17,26], faster rates of protein degradation

Heterozygote

(a)

Homozygote

(b)

Figure 5. Intrinsic noise in gene expression. The figure shows quantitative results for the intrinsic noise of autoregulating genes, as measured by the ratio of the variance to mean expression in protein concentration at equilibrium. (a) Percentage change in the noise of a heterozygote compared to the resident homozygote. These are shown for different Hill coefficients, $n=1$ (black), $n=2$ (red) and $n=3$ (blue). Mutations become deleterious in the heterozygote when $p_{max}/K>1$. (b) Percentage change in the noise of a mutant homozygote compared to the resident homozygote. Mutations become deleterious in the mutant homozygote when p_{max}/K is about 10. The graphs show the results of stochastic simulations (see Materials and Methods) for parameter values typical for transcription factors [3], $k_r=0.01s^{-1}$, $k_p=0.17s^{-1}$, $k_l=0.001s^{-1}$, $\gamma_r=\frac{1}{120}s^{-1}$ and $\gamma_p=\frac{1}{3600}s^{-1}$. The resident homozygote has binding strength p_{max}/K (as indicated by the x-axis), mutations are of size $\varepsilon=2$.
doi:10.1371/journal.pcbi.1002992.g005

result in faster response times, and regulation of protein degradation can reduce noise. As might be expected, the constraints we describe on the evolution of response times through stronger negative autoregulation do not preclude the evolution of

response times through other mechanisms, such as changes in protein degradation rates.

Discussion

Negative autoregulation is found to occur in 46% of *E. coli* transcription factors [1–3,18], but is rare in other species for which systematic data on transcriptional regulation is available, occurring in <2% of the known transcription factors of yeast, *Drosophila* and humans (see SI, Table S1, S2, S3). We have put forward the hypothesis that this difference can, at least in part, be explained by considering the different evolutionary dynamics of autoregulating genes in haploids and diploids: selection for genes to have a decreased response time to perturbations favours negative autoregulation in haploids, but under-dominance tends to prevent the evolution of stronger autoregulatory binding sites for this purpose in diploids. This constraint on the evolution of negative autoregulation in diploids is compelling because it offers a simple and general explanation for the apparent dearth of the motif in yeast, humans and flies. Furthermore, it is important to note that under-dominance is not built into our model but arises as an emergent property of our analysis of regulatory evolution – an analysis that simply extends to diploids previous models that have been shown to provide a good description of regulatory behaviour in haploids [2,3].

The empirical patterns we present are striking, however it is important to ask weather they can be explained by other means than those proposed in this paper. In particular we asked whether negative autoregulation is truly under-represented in the yeast, human and *Drosophila* data sets, as compared to *E. coli*, or whether the apparent reduction in the number of negative autoregulators is due to under-representation of genes with repressive function generally. To address this we interrogated each dataset to find the number of transcription factors with documented repressor activity. These account for 58 factors in humans, 37 in *Drosophila*, 54 in yeast and 82 in *E. coli*. If we include only transcription factors with known repressor function in our analysis, we find that 5 out of 58 (8.6%) genes negatively autoregulate in humans, 3 out of 37 (8.1%) in *Drosophila*, 3 out of 54 (5.6%) in yeast and 82 out of 138 (59%) in *E. coli*. Thus, the relative rarity of negative autoregulation in eukaryotes is not due to a general underrepresentation of repressive transcription factor effects among the genetic interactions described for these species. Instead, they appear to be a true property of their regulatory networks. This interpretation is based on our current knowledge of these networks. *E. coli* has been more intensively studied, so we look forward to more complete data on regulatory interactions in yeast, human and *Drosophila*, which will provide a more rigorous test of our hypothesis by enabling us to better establish the extent of negative autoregulation in eukaryotes.

It is also possible to conceive of experimental work to directly test our hypothesis that under-dominance constrains the evolution of negative autoregulation in diploids. This could exploit synthetic negative autoregulatory loops [1], comparing their regulation in haploid and duplicated copies. For example, a duplicated version of the tetracycline repressor-GFP system could be constructed in *E. coli* and expression dynamics monitored in cells that carry different combinations of wildtype and mutant promoters. Similar tests would then need to be performed with haploid and diploid circuits in eukaryotes such as budding or fission yeast, in order to show generality.

Another approach would be to examine haploid genes in diploid species and duplicate genes in haploid species. Haploid genes in a diploid organism should escape the evolutionary constraint on negative autoregulation. Unfortunately, the data on

genetic regulation are too sparse to test this prediction with any degree of rigor. The only candidate for a haploid gene in our dataset is the human Y-linked transcription factor *Sry* (see SI Table. S3). However, its mode of regulation (positive or negative) is unknown. Duplicate genes in haploids offer a better prospect as they are far more common [8]. Our model implies that in haploids, divergence in the expression levels of negatively autoregulating duplicates will tend to slow the response time of the pair. This is because expression divergence will tend to increase the response time in exactly the same way as we have described for heterozygotes in diploid cells. So negative autoregulating, multi-copy genes in haploids may be subject to evolutionary constraints similar to those we have described for diploids. The evidence for this is inconclusive. Negative autoregulating duplicates in *E. coli* are not more common than duplicates of other genes [8]. This is despite the prediction that they should be more common as they suffer less from the deleterious effects of increased dosage following duplication [8]. However, this test is not particularly strong as the evolutionary dynamics of duplication and divergence are complex [27], so simple predictions are not without alternative explanations.

An alternative hypothesis to the one analysed here is that eukaryotes experience different types of noise, and accordingly have different mechanisms for dealing with it, making negative autoregulation unnecessary. There are several points worth noting. The use of response time as a measure of fitness makes our model quite general, because all cells have to deal with large perturbations, such as occur across the cell cycle. The speed with which the concentration of a transcription factor returns to equilibrium, and the regulatory dynamics allowing it to do so, are important across all levels of biological complexity. Although our model captures the response time to perturbations and the amount of intrinsic noise associated with a gene [3], it does not capture other, extrinsic sources of noise. In particular, eukaryotes tend to be affected by "input noise" that results, for example, from the stochastic ON-OFF switching occurring in eukaryotic cells [17,28]. Previous work shows that this is best dealt with by positive autoregulation, not negative autoregulation [17,28,29]. However, positive autoregulation does not feature any more prominently than negative autoregulation within the regulatory networks of the three eukaryotes we analysed, with 9 instances in yeast, 16 in humans and 11 in *Drosophila*. These figures are not comparable to the frequency of negative autoregulation in *E. coli*, indicating that we are not simply observing a shift in the importance of different types of perturbations.

It is possible that eukaryotes deal differently with the kind of perturbations that require negative autoregulation in prokaryotes. Eukaryotes may be able to achieve negative autoregulation through multiple, weak autoregulatory binding sites, along with cooperation (see Figs. S1, S2). Our work shows that the evolution of strong cooperative autoregulation is subject to under-dominance (Fig. S1), but we find that the evolution of multiple, weak autoregulatory binding sites (Fig. S2) is less constrained. Since weak binding sites would likely be under-represented or absent from systematic datasets, it is possible that diploids achieve negative autoregulation in this way, and a study based on human sequence conservation suggest that autoregulatory binding sites are quite widespread [30]. Eukaryotes may also achieve negative autoregulation through mechanisms other than direct transcription regulation, for example, through changes in local chromatin structure or covalent changes in the protein structure of transcription factors. As these regulatory mechanisms are less likely to generate cross-regulation that occurs in diploid transcription regulation, they may not to be subject to under-dominance.

Finally, it is important to reiterate that our study is only concerned with the evolution of negative autoregulation for noise reduction and faster response times. Genes can achieve noise reduction through other means than autoregulation, and autoregulation can be used for other purposes than noise reduction [31–33]. We do not suggest that eukaryotes are exempt from the problem of noise. We do suggest that diploid gene networks, in contrast to those of haploids, must seek a different solution to the same problem.

Conclusion

We have put forward the hypothesis that regulatory interactions between homologous genes can generate deleterious effects that constrain the evolution of negative autoregulation. The predictions of our model show that the high incidence of autoregulation in *E. coli* and the dearth of negatively autoregulating genes in yeast, flies and humans can be reconciled by taking into account a simple biological attribute—ploidy. Importantly, the difference between haploid and diploid regulation dos not appear to be a mere correlate of the prokaryote-eukaryote divide. This was already suggested by the finding that the genetic networks of *E. coli* and yeast are—with the exception of their use of autoregulation— very similar [11].

More generally, our work demonstrates that regulatory evolution can be considerably complicated by the presence of multiple copies of a gene in a cell, as is typically the case for eukaryotes. By explicitly considering the evolution of regulatory interactions, we have highlighted constraints that would not be evident from an analysis of the functional properties of an existing regulatory interaction in isolation—strong negative autoregulation quickens the response of genes to perturbation, but it is hard to evolve for this purpose due to under-dominance. This evolutionary perspective needs to be absorbed into attempts at unravelling the function of regulatory networks in higher organisms, a key problem for systems biology.

Methods

Monte-Carlo simulations

We used simulations of the molecular dynamics within a cell to determine the amout of intrinsic noise of autoregulating genes in diploids. A model that tracks the number of mRNA and protein molecules for a negatively autoregulating gene within a haploid cell is described in [3]. We generalised this to account for diploidy. The state of the system is described by the number of mRNA molecules r_i, and the number of protein molecules p_i produced from the two alleles $i \in \{1,2\}$. The probability of a state $\{r_1, r_2, p_1, p_2\}$ is specified by the joint probability distribution $n_{r_1, r_2, p_1, p_2}(t)$. The transition probabilities for the system to move between states due to changes in r_1 and p_1 (and, analogously, due to changes in r_2 and p_2) are given by

$$\{r_1, r_2, p_1, p_2\} \xrightarrow{k_l + \phi_1(p)} \{r_1 + 1, r_2, p_1, p_2\},$$

$$\{r_1, r_2, p_1, p_2\} \xrightarrow{r_1 k_p} \{r_1, r_2, p_1 + 1, p_2\},$$

$$\{r_1, r_2, p_1, p_2\} \xrightarrow{r_1 \gamma_r} \{r_1 - 1, r_2, p_1, p_2\},$$

$$\{r_1, r_2, p_1, p_2\} \xrightarrow{p_1 \gamma_p} \{r_1, r_2, p_1 - 1, p_2\},$$

where $p = p_1 + p_2$, $k_l + \phi_1(p)$ is the rate at which mRNA molecules are transcribed from allele 1, γ_r is the rate of mRNA degradation, k_p is the rate at which mRNA is translated into protein and γ_p is the rate of protein degradation. As in the ODE model, $\phi_1(p)$ is a function of the number of proteins p present in the cell, such that

$$\phi_1(p) = \frac{k_0}{1 + \frac{p}{K_1}},$$

where k_0 is the maximum rate of mRNA transcription, and K_1 is the dissociation constant of the binding site of allele 1.

To calculate response times we first determined the equilibrium expression level of the system from the average of 10^5 replicate Monte-Carlo simulations. We then reduced mRNA and protein levels to a fraction α of the equilibrium level. The time for each replicate to return to equilibrium was measured and the average across the ensemble used as an estimate of the response time of the system. In order to determine how response times vary with the level of perturbation, simulations were run for values of α between 0 and 1 in steps of 0.01.

Simulations of binding site evolution

Binding site evolution was modelled by generating a transcription factor binding motif with a length n nucleotides and an optimal base associated with each nucleotide. As in other models of TF-DNA binding, when a given nucleotide i was matched for for the optimal base it contributed an amount ε_i to binding energy, otherwise it contributed 0 [20–22].

Binding site lengths were drawn from an empirical distribution generated from the binding motifs of 454 eukaryotic transcription factors contained in the JASPAR CORE database [23]. The value of ε_i for each nucleotide was drawn from a uniform distribution in the interval (0,3). The optimal binding strength K_{opt} was determined numerically (see Methods), using the values for the system parameters that are given in the legend of Fig. 4 4. We excluded from our analysis any binding sites for which the total binding strength of the optimal sequence was too low to achieve the fastest response time (i.e., those sequences for which $K = \exp[-\sum_i \varepsilon_i] > K_{opt}$). Evolution started from a state of minimum affinity (all nucleotides non-optimal) and proceeded through a series of single nucleotide substitutions. At each time step, a random mutation was introduced into the binding site sequence, switching one nucleotide from the non-optimal to the optimal state. If the mutation resulted in a response time less than or equal response time of the resident, the mutant sequence was assumed to go to fixation in the population. Deleterious mutations that increased response times were assumed to be lost. The simulation was ended when no further advantageous mutations were available. Simulations were carried out for both haploids and diploids (for which response time of mutants was evaluated in the heterozygote state).

Derivation of response times in haploids

Here we derive results for the response time of a haploid autoregulating gene. We derive results for the general case in which autoregulation is described by a Hill function with arbitrary coefficient n (the analyses in the main text assumes $n = 1$).

The set of ODEs describing transcription and translation of mRNA and protein at a single autoregulating gene are analogous to those given for one allele in Eqs. 1 for a pair of autoregulating genes in a diploid. In order to simplify the analysis of the system we make the change of variables

$$s = \frac{r}{r_{max} - r_{min}},$$

$$q = \frac{p}{p_{max} - p_{min}},$$

$$\tau = \gamma_p t,$$

with $L = \frac{K}{p_{max} - p_{min}}$ and $\gamma = \frac{\gamma_p}{\gamma_r}$. The dynamics of the system can then be rewritten as

$$\gamma \frac{ds}{d\tau} = \beta + k_s(q) - s,$$
$$\frac{dq}{d\tau} = s - q,$$
(3)

where $\beta = \frac{r_{min}}{r_{max} - r_{min}} = \frac{p_{min}}{p_{max} - p_{min}} = \frac{k_l}{k_0}$. In general $\beta \ll 1$ since $k_l \ll k_0$ and

$$k_s(q) = \frac{1}{1 + \left(\frac{q}{L}\right)^n}, \quad (4)$$

is the rescaled form of the repression function $\phi(p)$ described in the main text. Assuming that mRNA decays much faster than protein [2,3] $\gamma_p \ll \gamma_r$, then $\gamma \ll 1$, it follows that $\gamma \frac{ds}{d\tau}$ is small relative to $\frac{dq}{d\tau}$ and we can assume that transcription output goes to equilibrium rapidly. That is, we can take $\gamma \frac{ds}{d\tau} \approx 0$ and hence that the quasi equilibrium condition $s \approx \beta + k_s(q)$ holds. Substituting into Eqs. 3, generates a 2-dimensional system that is well approximated by the 1-dimensional system

$$\frac{dq}{d\tau} = \beta + k_s(q) - q. \quad (5)$$

Small perturbations

The Lyapunov exponent associated with Eq. 5 at equilibrium gives the rate at which the system returns to equilibrium following a small perturbation. It is given by

$$\lambda = -\left(1 + \left|\frac{dk_s}{dq}\right|\right). \quad (6)$$

Eq. 6 is always negative. In what follows we will discuss only the magnitude of the Lyapunov exponent $|\lambda|$ with the understanding that this quantity is always negative and therefore describes the rate at which the system returns to equilibrium. From Eq. 6 it is clear that a mutation which increases $\left|\frac{dk_s}{dq}\right|$ will always serve to decrease the Lyapunov exponent and thus increase the rate at which the system converges to equilibrium.

Evolution of a new binding site

We compare a wild-type binding site, with dissociation constant L_1, to a mutant binding site with dissociation constant L_2 such that $L_1 > L_2$ —meaning that the mutant has a stronger binding site than the wild-type. At equilibrium, the protein concentrations

satisfy

$$q_1^* = \beta + \frac{1}{1 + \left(\frac{q_1^*}{L_1}\right)^n} \quad ,$$

$$q_2^* = \beta + \frac{1}{1 + \left(\frac{q_2^*}{L_2}\right)^n} \quad . \tag{7}$$

It is simple to show that $q_1^* > q_2^*$ by differentiating, with respect to L. Thus, strengthening the autoregulatory binding site (i.e., decreasing L) will lead to a decrease in the equilibrium protein concentration, and so with $L_1 > L_2$ we always have $q_1^* > q_2^*$. To calculate the value of L_{opt} for which $|\lambda|$ is maximum, we note that

$$\left|\frac{dk_s}{dq}\right| = \frac{n}{q} k_s(q)(1 - k_s(q)).$$

At equilibrium $q = q^* = k_s(q^*) + \beta$, and the Lyapunov exponent can be written as

$$|\lambda| = \left(1 + \frac{n}{q^*}(q^* - \beta)(1 - q^* + \beta)\right),$$

and we can find the value of q^* that results in the largest Lyapunov exponent. This is given by

$$q^* = \sqrt{\beta(1 + \beta)} \quad .$$

Translating this back into units of protein concentration, this means that the fastest response to small perturbations about equilibrium occurs when

$$p^* = \sqrt{p_{max} p_{min}} \quad . \tag{8}$$

Thus, mutations which increase the strength of negative autoregulation, (and therefore decrease K), will decrease response time provided the equilibrium protein concentration is $p^* > \sqrt{p_{max} p_{min}}$, as discussed in the main text. The optimal binding site strength K_{opt} can be determined by calculating the value of K which gives the optimal equilibrium protein concentration of Eq. 8. In the general case of arbitrary n, K_{opt} cannot be found analytically, but it can always be found numerically.

The derivation of K_{opt} presented here is based on the assumption that perturbations of the system are small, in which case the dynamics of the system are well captured by its Lyapunov exponent. The optimal binding strength under perturbations of arbitrary size can be obtained by numerical integration of the system. As might be expected, the values K_{opt} obtained in this way are similar to those calculated for small perturbations above.

Derivation of response times in diploids

For diploids we proceed in the same way as for a single gene, and obtain a 1-D system for expression of a pair of alleles (with dissociation coefficients L_i and L_j)

$$\frac{dq_{ij}}{d\tau} = 2\beta + k_{s_i}(q_{ij}) + k_{s_j}(q_{ij}) - q_{ij} \quad , \tag{9}$$

where

$$k_{s_i}(q_{ij}) = \frac{1}{1 + \left(\frac{q_{ij}}{L_i}\right)^n} \quad .$$

Evolution of a new binding site

We now consider the response time of a pair of autoregulating alleles in a diploid. When an organism is homozygous, both binding sites have the same dissociation constant, L_1 and Eq. 9 is of the same form as Eq. 5 for a haploid, and the results for response time in haploids can be applied. When an organism is heterozygous however, the results for haploids do not hold. We compare the Lyapunov exponents of a heterozygote with dissociation constants L_1 and L_2, where $L_2 < L_1$, to a resident homozygote in which both binding sites have strength L_1. At equilibrium the total protein concentrations satisfy

$$q_{11}^* = 2\beta + \frac{2}{1 + \left(\frac{q_{11}^*}{L_1}\right)^n} \quad ,$$

$$q_{12}^* = 2\beta + \frac{1}{1 + \left(\frac{q_{12}^*}{L_1}\right)^n} + \frac{1}{1 + \left(\frac{q_{12}^*}{L_2}\right)^n} \quad , \tag{10}$$

where q_{11}^* is the equilibrium protein concentration of the (resident) homozygote and q_{12}^* is the equilibrium expression of the (mutant) heterozygote. It is simple to show that $q_{11}^* > q_{12}^*$. by differentiating Eq. 10 with respect to L_2.

Small perturbations

Following a small displacement from equilibrium, under-dominance will occur if the heterozygote has a smaller Lyapunov exponent than the homozygote. The maximal Lyapunov exponent of the system is given by

$$|\lambda_{11}| = 1 + \frac{2n}{q_{11}^*}\left(\frac{q_{11}^*}{2} - \beta\right)\left(1 - \frac{q_{11}^*}{2} + \beta\right), \tag{11}$$

for the homozygote, and

$$|\lambda_{12}| = 1 + \frac{n}{q_{12}^*}\left(q_{1,12}^* - \beta\right)\left(1 - q_{1,12}^* + \beta\right) + \frac{n}{q_{12}^*}\left(q_{2,12}^* - \beta\right)\left(1 - q_{2,12}^* + \beta\right), \tag{12}$$

for the heterozygote, where $q_{i,ij}$ referes to allele i in a diploid carrying alleles i and j. We can observe that the squared difference in the mean allele expression, V, is given by $V = \left(q_{1,12}^* - \frac{q_{12}^*}{2}\right)^2 + \left(q_{2,12}^* - \frac{q_{12}^*}{2}\right)^2$, which can be expanded to give

$$V = (q_{1,12}^*)^2 + (q_{2,12}^*)^2 - \frac{q_{12}^*}{2}.$$

Substituting this expression for V in Eq. 12 we find

$$|\lambda_{11}| = 1 + n\left(1 - \frac{q_{11}^*}{2} + 2\beta - \frac{2\beta}{q_{11}^*}(1+\beta)\right), \quad (13)$$

$$|\lambda_{12}| = 1 + n\left(1 - \frac{q_{12}^*}{2} + 2\beta - \frac{2\beta}{q_{12}^*}(1+\beta) - \frac{V}{q_{12}^*}\right). \quad (14)$$

Note that Eq. 14 is of the same form as Eq. 13, with an additional term that depends on the ratio of the squared difference in allele expression, V to the total expression. We can define λ_{hom} to be the Lyapunov exponent associated with a homozygote of a given equilibrium expression and λ_{het} to be the Lyapunov exponent associated with a heterozygote of the same equilibrium expression and obtain Eq. 2 of the main text (with $n=1$).

Supporting Information

Figure S1 Steeper repression functions further limit the range of mutations that escape under-dominance. The x-axis shows the geometric mean of the binding strength across the set of resident alleles, in units of p_{max}/K, and the y-axis shows the size of mutations to binding site strength, as described in the main text. In the gray region, mutations to one of the n binding sites result in increased response time in the mutant compared to the resident alleles. In the white region mutations result in decreased response time in the mutant compared to the resident alleles; only mutations that fall within the white region can invade a population. Mutant invasibility is shown for Hill coefficients $n=2$ (left), and $n=5$ (right). Weak binding occurs when $p_{max}/K \gtrsim 10^0$. Response times are calculated by numerically integrating Eq. 1 from zero protein concentration to 90% of the equilibrium. The optimal binding strength in these graphs is $p_{max}/K = 1250$ corresponding to a background transcription rate $k_l/k_0 = 10^{-3}$.
(PDF)

Figure S2 Increasing the Hill coefficient leads to slower response times unless binding strength is weak. The x-axis shows the binding strength in the resident allele, in units of p_{max}/K, and the y-axis shows the ratio of response times for a heterozygote in which one allele has a Hill coefficient $n=1$ and the other has a Hill coefficient $n=2$, to a homozygote with Hill coefficient $n=1$. Below the gray dashed line, mutations result in increased response time in the mutant compared to the resident allele. Weak binding occurs when $p_{max}/K \gtrsim 10^0$. Response times are calculated by numerically integrating Eq. 1 from zero protein concentration to 90% of the equilibrium. The optimal binding strength in these graphs is $p_{max}/K = 1250$ corresponding to a background transcription rate $k_l/k_0 = 10^{-3}$.
(PDF)

Figure S3 Changing the background rate of transcription does not substantially alter the impact of under-dominance on mutations of size $\varepsilon > 2$. The x-axis shows the binding strength in the resident allele, in units of p_{max}/K, and the y-axis shows the size of mutations to binding site strength, as described in the main text. In the gray region, mutations result in increased response time in the mutant compared to the resident allele. In the white region mutations result in decreased response time in the mutant

compared to the resident allele; only mutations that fall within the white region can invade a population. Mutant invasibility is shown for background transcription rates $k_l/k_0 = 10^{-2}$ (left), and $k_l/k_0 = 10^{-4}$ (right). Weak binding occurs when $p_{max}/K \lesssim 10^0$. Response times are calculated by numerically integrating Eq. 1 from zero protein concentration to 90% of the equilibrium.
(PDF)

Figure S4 Changing degradation rates changes response times but does not allow autoregulation to escape under-dominance. The figure shows results for the response time of autoregulating genes, to return to 90% of their equilibrium. (left) Percentage change in the response time of a heterozygote compared to the resident homozygote. These are shown for different protein degradation rates coefficients, $\gamma_p = \frac{1}{3600}$ (black), $\gamma_p = \frac{1}{36000}$ (red) and $\gamma_p = \frac{1}{360}$ (blue). Mutations become deleterious in the heterozygote when $p_{max}/K > 1$. (right) Percentage change in the response time of a mutant homozygote compared to the resident homozygote. Mutations become deleterious in the mutant homozygote when p_{max}/K is about 10. The graphs show the results of stochastic simulations (see Materials and Methods) for parameter values typical for transcription factors, $k_r = 0.01s^{-1}$, $k_p = 0.17s^{-1}$, $k_l = 0.001s^{-1}$ and $\gamma_r = \frac{1}{120}s^{-1}$. The resident homozygote has binding strength p_{max}/K (as indicated by the x-axis), mutations are of size $\varepsilon = 2$.
(PDF)

Figure S5 Invasibility of autoregulatory binding sites. The response time of mutant (left) homozygotes and (right) heterozygotes are shown. Different values of the binding strength of the resident allele, in units of p_{max}/K (x-axis), are plotted against mutations to binding site strength ε of different size (y-axis). Thus the graphs compare a resident allele, K_1 with a mutant allele, $K_2 = K_1 \exp[-\varepsilon]$. Mutations falling into white region result in decreased response time in the carrier compared to resident genotype and are favoured by selection; mutations falling into the gray region result in increased response time and are not favoured by selection. Weak binding occurs when $p_{max}/K \gtrsim 10^0$ [1,2]. Response times were calculated by numerically integrating Eq. 1 from zero protein concentration to 99% of the equilibrium. The optimal binding strength in these graphs is $p_{max}/K = 1250$, corresponding to a background transcription rate $k_l/\gamma_p = 10^{-3}$.
(PDF)

Table S1 Autoregulation in *Saccharomyces cerevisiae* [10,13,14,34–36].
(PDF)

Table S2 Autoregulation in Humans [13–15,35,36].
(PDF)

Table S3 Autoregulation in *Drosophila* [12–14,35,36].
(PDF)

Text S1 The supporting information text describes the methods used in constructing Tables S1, S2, S3 from curated databases, and in creating Fig. S1, S2, S3, S4, S5 by relaxing the assumptions of the model outlined in the main text.
(PDF)

Acknowledgments

Author Contributions

References

1. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. Nature 405: 590–3.
2. Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. J Mol Biol 323: 785–93.
3. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98: 8614–9.
4. Wang Z, Zhang J (2011) Pnas plus: Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc Natl Acad Sci U S A 108: E67–76.
5. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of escherichia coli. Nat Genet 31: 64–8.
6. Thieffry D, Huerta AM, Pérez-Rueda E, Collado-Vides J (1998) From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in escherichia coli. Bioessays 20: 433–40.
7. Sánchez A, Kondev J (2008) Transcriptional control of noise in gene expression. Proc Natl Acad Sci U S A 105: 5081–6.
8. Warnecke T, Wang GZ, Lercher MJ, Hurst LD (2009) Does negative auto-regulation increase gene duplicability? BMC Evol Biol 9: 193.
9. Guelzim N, Bottani S, Bourgine P, Kepes F (2002) Topological and causal structure of the yeast transcriptional regulatory network. Nat Genet 31: 60–3.
10. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, et al. (2002) Transcriptional regulatory networks in saccharomyces cerevisiae. Science 298: 799–804.
11. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, et al. (2004) Superfamilies of evolved and designed networks. Science 303: 1538–42.
12. Bergman CM, Carlson JW, Celniker SE (2005) Drosophila dnase i footprint database: a systematic genome annotation of transcription factor binding sites in the fruity, drosophila melanogaster. Bioinformatics 21: 1747–9.
13. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al. (2006) Transfac and its module transcompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34: D108–10.
14. Wingender E, Dietze P, Karas H, Knüppel R (1996) Transfac: a database on transcription factors and their dna binding sites. Nucleic Acids Res 24: 238–41.
15. Bauer T, Eils R, König R (2011) Rip: the regulatory interaction predictor–a machine learningbased approach for predicting target genes of transcription factors. Bioinformatics 27: 2239–47.
16. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467: 167–73.
17. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135: 216–226.
18. Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental limits on the suppression of molecular uctuations. Nature 467: 174–8.
19. Chu D, Zabet NR, Mitavskiy B (2009) Models of transcription factor binding: sensitivity of activation functions to model assumptions. J Theor Biol 257: 419–29.
20. Gerland U, Moroz JD, Hwa T (2002) Physical constraints and functional characteristics of transcription factor-dna interaction. Proc Natl Acad Sci U S A 99: 12015–20.
21. Lässig M (2007) From biophysics to evolutionary genetics: statistical aspects of gene regulation. BMC Bioinformatics 8 Suppl 6: S7.
22. Berg J, Willmann S, Lässig M (2004) Adaptive evolution of transcription factor binding sites. BMC Evol Biol 4: 42.
23. Bryne JC, Valen E, Tang MHE, Marstrand T, Winther O, et al. (2008) Jaspar, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36: D102–6.
24. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104.
25. Gerland U, Hwa T (2009) Evolutionary selection between alternative modes of gene regulation. Proc Natl Acad Sci U S A 106: 8841–6.
26. Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, et al. (2011) Proteome half-life dynamics in living human cells. Science 331: 764–8.
27. Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11: 97–108.
28. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mrna synthesis in mammalian cells. PLoS Biology 4: e309.
29. Gregor T, Tank DW, Wieschaus EF, Bialek W (2007) Probing the limits to positional information. Cell 130: 153–64.
30. Kie lbasa SM, Vingron M (2008) Transcriptional autoregulatory loops are highly conserved in vertebrate evolution. PLoS One 3: e3210.
31. Amit I, Citri A, Shay T, Lu Y, Katz M, et al. (2007) A module of negative feedback regulators defines growth factor signaling. Nat Genet 39: 503–12.
32. Legewie S, Herzel H, Westerhoff HV, Blüthgen N (2008) Recurrent design patterns in the feedback regulation of the mammalian signalling network. Mol Syst Biol 4: 190.
33. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, et al. (2011) Global quantification of mammalian gene expression control. Nature 473: 337–42.
34. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, et al. (1997) Genetic and physical maps of saccharomyces cerevisiae. Nature 387: 67–73.
35. Kersey PJ, Lawson D, Birney E, Derwent PS, Haimel M, et al. (2010) Ensembl genomes: extending ensembl across the taxonomic space. Nucleic Acids Res 38: D563–9.
36. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, et al. (2011) Ensembl biomarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011: bar030.