BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

The population genetics of cooperative gene regulation

Alexander J Stewart[1,2], Robert M Seymour[2,3], Andrew Pomiankowski[2,4*] and Joshua B Plotkin[1*]

Abstract

Background: Changes in gene regulatory networks drive the evolution of phenotypic diversity both within and between species. Rewiring of transcriptional networks is achieved either by changes to transcription factor binding sites or by changes to the physical interactions among transcription factor proteins. It has been suggested that the evolution of cooperative binding among factors can facilitate the adaptive rewiring of a regulatory network.

Results: We use a population-genetic model to explore when cooperative binding of transcription factors is favored by evolution, and what effects cooperativity then has on the adaptive re-writing of regulatory networks. We consider a pair of transcription factors that regulate multiple targets and overlap in the sets of target genes they regulate. We show that, under stabilising selection, cooperative binding between the transcription factors is favoured provided the amount of overlap between their target genes exceeds a threshold. The value of this threshold depends on several population-genetic factors: strength of selection on binding sites, cost of pleiotropy associated with protein-protein interactions, rates of mutation and population size. Once it is established, we find that cooperative binding of transcription factors significantly accelerates the adaptive rewiring of transcriptional networks under positive selection. We compare our qualitative predictions to systematic data on *Saccharomyces cerevisiae* transcription factors, their binding sites, and their protein-protein interactions.

Conclusions: Our study reveals a rich set of evolutionary dynamics driven by a tradeoff between the beneficial effects of cooperative binding at targets shared by a pair of factors, and the detrimental effects of cooperative binding for non-shared targets. We find that cooperative regulation will evolve when transcription factors share a sufficient proportion of their target genes. These findings help to explain empirical pattens in datasets of transcription factors in *Saccharomyces cerevisiae* and, they suggest that changes to physical interactions between transcription factors can play a critical role in the evolution of gene regulatory networks.

Background

It is often difficult for a population to acquire an adaptive phenotype that requires simultaneous changes in the co-expression of multiple genes [1-10]. If selection favours a change in the way a group of genes are regulated, then each of the target genes must independently gain novel binding sites and/or lose existing ones [8,9,11,12]. This has led to the proposal that adaptive rewiring of a regulatory network can be accelerated if pairs of transcription factors bind their targets cooperatively, through a physical interaction between the transcription factor proteins themselves [8,9,11].

Given the potential adaptive benefit of cooperative regulation, it makes sense to ask, when will cooperative binding between a pair of transcription factors be able to invade a population that lacks such cooperativity? To answer this we must understand the following tradeoff: although cooperative binding between a pair of factors may result in improved regulation at the target genes shared by both factors, any mutation that results in a physical interaction between the transcription factors will effect *all of their targets* (Figure 1). Thus the advantageous fitness effects of improved binding at some, shared targets must outweigh any deleterious effects of misregulation at other, non-shared targets in order for cooperative binding to be favored by evolution.

A number of previous studies have explored the mechanistic details of cooperative transcription factor binding

*Correspondence: a.pomiankowski@ucl.ac.uk; jplotkin@sas.upenn.edu
[2]CoMPLEX, University College London, London, UK
[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
Full list of author information is available at the end of the article

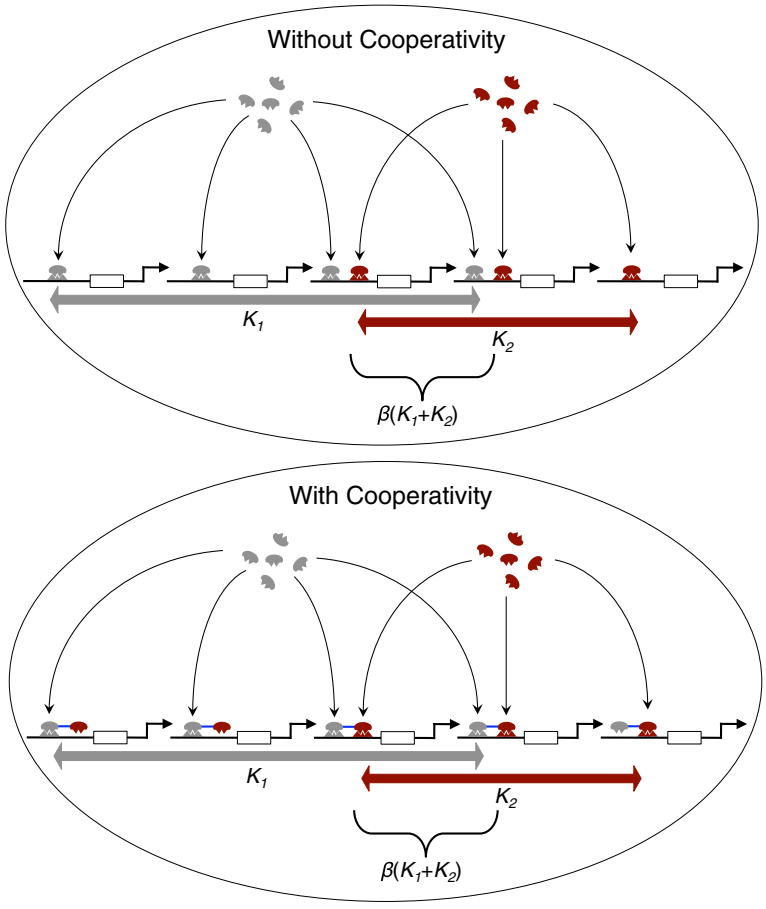

Figure 1 Schematic of the population-genetic model. A schematic cartoon of our population-genetic model. (top) When cooperativity is absent different transcription factors (gray and red) must bind to sites at each of their targets independently. Each factor has a number of targets, K_1 and K_2, and a number $\beta(K1 + K2)$ of shared targets (bottom). When cooperativity is present, a physical interaction between transcription factors (blue line) can mitigate the need to bind independently at shared targets, but may cause misregulation at targets that are not shared, by causing the factor with which it interacts cooperatievly to misbind. Cooperatively is therefore advantageous between transcription factors that share many targets, but it may be deleterious at targets that are not shared.

at a given target gene [13-15]. Such biophysical studies focus on transcription factor binding at a single target gene and are able, with remarkable accuracy, to account for a number of the physical properties of binding sites [14,16,17]. However, the evolution of cooperative binding occurs through mutations at transcription factor proteins, and such mutations can alter transcription factor binding at every binding site across the genome. To understand the fitness effects of such a mutation therefore requires that we understand the evolution of the whole ensemble of binding sites for a transcription factor. The population genetics of such an ensemble cannot be understood in a simple way just by focusing on the details of a single member of the ensemble. They depend critically on the population-genetic parameters of the ensemble, such as number of target genes, overall mutation rates and selection coefficients, and population size. Therefore in this paper, we do not focus on the details of a cooperative binding at a single target gene. Instead our analysis is in terms of these population-genetic parameters, and whilst we estimate selective coefficients from biophysical studies, we do not specify the mechanistic details of protein-DNA interactions that give rise to them.

We use a mathematical model to study the conditions under which cooperative binding between pairs of transcription factors is favoured. We first determine the evolutionary conditions that favour cooperative binding under stabilising selection, in terms of the basic evolutionary parameters of the population: the strength of selection on binding sites, the rate of mutation, and the population size. We then study the influence of cooperative regulation on the capacity for a transcriptional circuit to adapt under positive selection. We calculate the time required for a target gene to gain a new, adaptive transcription

factor binding site, in the presence or absence of cooperative interactions among its regulators. We confirm our analytical results on the evolution of cooperative regulation by comparison to Monte-Carlo simulations of the Wright-Fisher process associated with our system, and we compare our qualitative conclusions to systematic empirical data.

Our population-genetic model describes a pair of transcription factors, each with its own set of target genes, with some degree of overlap between these sets (Figure 1). According to our model, which is specified in detail below, a target gene that is regulated by both factors has two corresponding binding sites, while a target gene that is regulated by only one of the factors has a single binding site. We assume that mutations that result in loss of function can occur at any binding site, and that non-functional binding sites can also undergo gain of function mutations. When there is no cooperative regulation between the two transcription factors, binding to each of their targets is determined solely by their binding sites. If a binding site is not functional, this results in reduced fitness. When cooperative binding is present, two conflicting effects occur: On the one hand, cooperative binding partially compensates for the deleterious effects of loss of function mutations to the binding sites at shared targets. On the other hand, cooperative binding results in some degree of mis-regulation at each of the targets that are not shared, and this has a deleterious impact on fitness. By constructing our model in terms of these fitness benefits and costs we are able to study the evolutionary dynamics of the system, and determine the effects of varying different population-genetic parameters on the evolution of cooperative gene regulation. This approach therefore complements the detailed mechanistic models of gene regulation that have been studied elsewhere [13-15].

Results and discussion

Stabilising selection without cooperative binding

We consider a pair of transcription factors, labelled 1 and 2, that have K_1 and K_2 targets, respectively. A fraction β of the binding sites are at shared target genes, so that the number of binding sites at genes that are co-regulated by the pair is $\beta(K_1 + K_2)$, as illustrated in Figure 1. Loss of function mutations occur at binding sites at a rate u_l, and back mutations, which result in a functional binding site being gained at a target, occur at rate u_g. An individual incurs a fitness penalty s, where $0 \leq s < 1$, for each non-functional binding site, and fitness is assumed to be multiplicative across loci. Therefore the fitness of an individual that lacks $i \leq K_1 + K_2$ of its required binding sites is $w_i = (1 - s)^i$. The fitness landscape associated with our model thus has a single peak at $i = 0$; and for each transcription factor binding site that is lost, fitness is reduced by an additional factor $(1-s)$. Empirical estimates of the strength of selection on transcription factor binding sites suggest that typically $Ns \sim 10$ [18], suggesting that s is small. We assume that s is the same for all binding sites, an assumption which we relax in the Methods section.

We consider a population of N asexual individuals. The evolution of the population can be described by keeping track of the relative abundances of each "hamming class" [19-21]. Hamming class i corresponds to those individuals who currently lack i transcription factor binding sites. We denote the frequency of individuals in hamming class i by x_i. In an infinitely large population, the evolution of hamming class i is then described by the differential equations [20,21]

$$\dot{x}_i = \sum_{j=0}^{K_1+K_2} \frac{w_i}{\bar{w}} z_i P_{ij}, \tag{1}$$

where $\bar{w} = \sum_{i=0}^{K_1+K_2} w_i x_i$, and P_{ij} is the probability a genotype lacking j functional binding sites mutates to a genotype lacking i functional binding sites (see Methods). Previous work [19-21] has shown that at equilibrium, when rates of forward and back mutations are identical ($u_l = u_g$), the solution to Equation 1 is a binomial distribution. In the more general case of a finite population, with $u_l \neq u_g$, we find that the equilibrium continues to be well approximated by a binomial distribution, with mean $(K_1 + K_2)a_s$. The term a_s is the probability that a binding site will be non-functional in a randomly chosen individual at equilibrium. The probability a_s depends on the strength of selection against non-functional binding sites, s, population size, N, and the rates of forward and back mutation, u_l and u_g (see Methods and [20,21]).

The equilibrium distribution above describes how stabilizing selection determines the frequencies of functional binding sites in a population. The associated mean fitness for a pair of transcription factors that do not bind cooperatively is $\bar{w} = (1 - a_s s)^{K_1+K_2}$ (see Methods), and the mean fitness contribution of each binding site is $1 - a_s s$. We are typically concerned with the case in which $u_l, u_g \ll s$. In this case, when $2Ns > 1$, a_s can be approximated by

$$a_s \approx \frac{1}{2Ns} \frac{u_l}{u_l + u_g} + \frac{u_l}{s}$$

and otherwise by

$$a_s \approx \frac{u_l}{u_l + u_g} \tag{2}$$

(see Methods). These equations have an intuitive interpretation: When $2Ns > 1$ the first term describes the effect of genetic drift which tends to push the system towards its neutral equilibrium, $a_0 = u_l/(u_l + u_g)$, and the second term describes the effect of selection. In the limit $N \to \infty$,

a_s equals u_l/s, which is the standard result for the frequency of a deleterious allele in an infinite population under mutation-selection balance. When $2Ns < 1$, evolution is nearly neutral and drift dominates, so the system is close to the neutral equilibrium a_0.

Stabilising selection with cooperative binding

Here we modify our model to account for cooperative regulation by a pair of factors. This allows us to ask when cooperative regulation is favored by evolution. A mutation that results in cooperative binding between a pair of transcription factors has two effects on the fitness of a transcriptional circuit. For a target that is regulated by both transcription factors, we assume that cooperative binding mitigates the effects of deleterious mutations at transcription factor binding sites [7-9]. This results in a reduced fitness penalty for a mutation at the $\beta(K_1 + K_2)$ shared targets, so that $(1 - s)$ is replaced by $(1 - hs)$ for some constant $0 \leq h \leq 1$. Nonetheless, there are also $(1 - \beta)(K_1 + K_2)$ targets that are regulated by only one or the other of the transcription factors. We assume that the cooperative binding of the transcription factors causes pleiotropic mis-regulation at these targets (since the other transcription factor, which does not have a binding site at such sites, now binds to the first transcription factor through a physical interaction). This results in a fitness penalty t at each of the $(1-\beta)(K_1+K_2)$ targets that are not co-regulated. Fitness is again assumed to be multiplicative, so that the cost of pleiotropy associated with cooperative binding is $(1 - t)^{(1-\beta)(K_1+K_2)}$.

Provided $u_l, u_g \ll 1$, genes that are co-regulated and genes that are not co-regulated have equilibrium distributions described by independent binomial distributions with means a_{hs} and a_s respectively, which are approximated by Equation 2 (substituting hs for s appropriately, see Methods). We can now specify the conditions for the invasion of cooperative gene regulation. A mutation resulting in cooperative binding between a pair of factors will be favoured if the expected fitness of the mutant is greater than the equilibrium mean fitness. Using the expressions for mean fitness given above, this occurs when $(1 - a_s s)^{\beta(K_1+K_2)} < (1 - t)^{(1-\beta)(K_1+K_2)}(1 - a_s hs)^{\beta(K_1+K_2)}$. Assuming $t, s \ll 1$, this expression can be simplified to give $\beta > \frac{t}{t+sa_s(1-h)}$. This means that, when the fraction of binding sites at shared targets, β, is greater than a threshold depending on s, h, t and a_s, a mutation that results in cooperative binding can invade a population at equilibrium.

Similarly, a mutation that results in the loss of cooperative binding in a population where it is present will be favoured when $(1-a_{hs}s)^{\beta(K_1+K_2)} < (1-t)^{(1-\beta)(K_1+K_2)}(1-a_{hs}hs)^{\beta(K_1+K_2)}$. Again assuming $t, s \ll 1$, this expression can be simplified to give $\beta < \frac{t}{t+sa_{hs}(1-h)}$ so that, when the fraction of binding sites at shared targets, β, is less than a threshold depending on s, h, t and a_{hs}, a mutation that results in loss of cooperative binding can invade a population at equilibrium.

Since the first expression in Equation 2 is monotonically decreasing in s, and the second expression is independent of s, it is always true that $a_{hs} \leq a_s$, i.e populations that have cooperative binding accumulate more deleterious mutations, that result in weaker transcription factor binding sites, than populations that lack it. As a result there is a range of β for which *both* a population that lacks cooperative binding, *and* a population that has cooperative binding are not invadable by mutations that gain or remove cooperative binding respectively. In this range, the evolutionary dynamics of the system are bi-stable. In this range, we expect to find some genes that are regulated by pairs of transcription factors that act cooperatively and some that don't.

Using the expression for a_s given in Equation 2, and recalling that $a_0 = u_l/(u_l+u_g)$ is the neutral equilibrium in a system dominated by drift, the threshold value of β above which selection favours a mutation causing cooperative binding in a population that lacks it, is given by

$$\beta > \begin{cases} \frac{2Nt}{2Nt+a_0(1-h)} & \text{if } 2Ns > 1 \\ \frac{t}{t+sa_0(1-h)} & \text{otherwise} \end{cases}. \quad (3)$$

Similarly, the threshold value of β below which selection favours a mutation resulting in loss of cooperative binding in a population that has it, is given by

$$\beta < \begin{cases} \frac{2Nth}{2Nth+a_0(1-h)} & \text{if } 2Nhs > 1 \\ \frac{t}{t+sa_0(1-h)} & \text{otherwise} \end{cases}. \quad (4)$$

These equations allow us to make a number of observations about the evolution of cooperative gene regulation (Figure 2, and see Methods). Beginning with Equation 3 for a population lacking cooperative binding, we see that when N and/or s is large, so that $2Ns > 1$, the threshold number of shared targets β above which cooperative binding becomes advantageous is independent of the strength of selection s (Figure 2a). However the threshold decreases as the mutation-buffering effect of cooperative binding increases (i.e. as h decreases, Figure 2b). As population size N increases, selection becomes more efficient and the threshold value of β increases (Figure 2c). Finally, the threshold also increases with the cost of pleiotropy t (Figure 2d). In contrast, when N and/or s is small, so that $2Ns < 1$, drift dominates and the threshold number of shared targets β is independent of population size N (Figure 2c). However the threshold decreases with the strength of selection s (Figure 2a), because when drift dominates the number of deleterious mutations is at the neutral equilibrium, and increasing s increases the impact of each mutation on overall fitness.

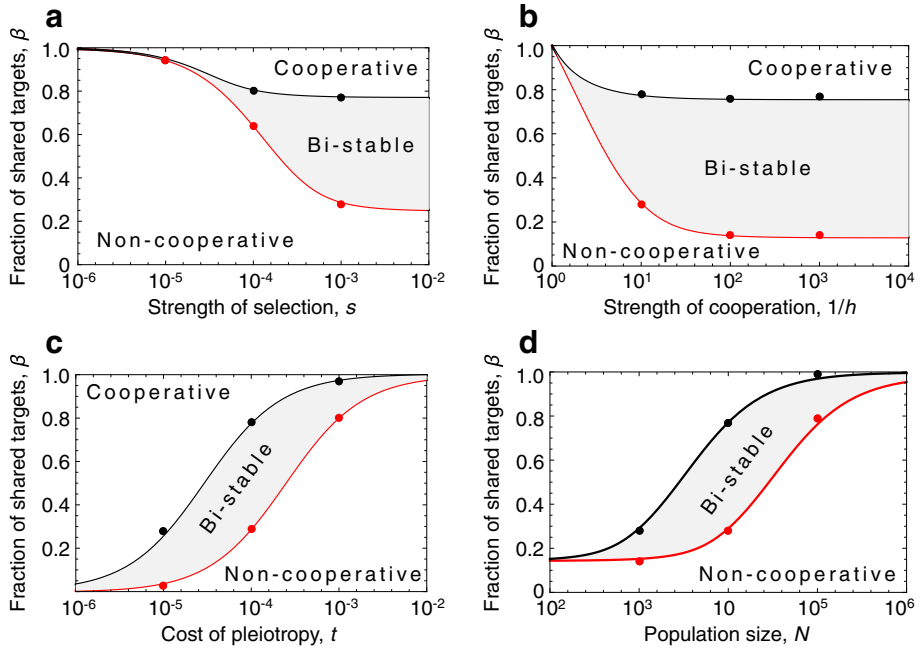

Figure 2 Evolutionary parameters that permit cooperative regulation. Evolutionary parameters that permit the evolution of gene regulation by cooperative transcription factors. Threshold number of shared targets for gain (black) and loss (red) of cooperative binding to be advantageous in a population at equilibrium under stabilising selection. The black line shows the value of β above which a new mutation that results in cooperative binding will invade in a population that lacks cooperative binding. The red line shows the value of β below which a mutation resulting in loss of cooperative binding will invade, in a population that has cooperative binding. For values of β that lie in the gray region, the dynamics are bistable: a population with cooperative binding will preserve it, and one without binding will not gain binding. The threshold fraction of shared targets varies with (top left) strength of selection, s, (top right) strength of cooperativity in reducing the effects of deleterious mutations $1/h$, (bottom left) the cost of pleiotropy t and (bottom right) the population size, N. Lines show our analytic equations (Equations 2 and 3), and points show the results of 10^5 replicate Monte-Carlo simulations. Parameter values (unless stated otherwise) are $u_l = 2 \times 10^{-7}$, $u_g = 10^{-7}$, $K_1 + K_2 = 100$, $s = 10^{-3}$, $h = 10^{-1}$, $t = 10^{-4}$ and $N = 10^4$.

Similarly, from Equation 4 for a population with cooperative binding, we see that when N and/or hs is large, so that $2Nhs > 1$, the threshold number of shared targets β below which cooperative binding becomes disadvantageous is independent of the strength of selection s (Figure 2a). As before, the threshold decreases as the mutation buffering effect of cooperative binding increases (i.e. as h decreases, Figure 2b) and the threshold increases with population size N (Figure 2c), and the cost of pleiotropy t (Figure 2d). In contrast, when N and/or hs is small, so that $2Nhs < 1$, drift dominates and the threshold number of shared targets β is independent of population size N (Figure 2c), but decreases with the strength of selection s (Figure 2a). The size of the bistable region is largest when s is large and h is small, and for intermediate values of N and t, as shown in Figure 2. As this analysis demonstrates, there is a broad range of possible evolutionary outcomes and, crucially, cooperative binding can evolve under a wide range of circumstances despite the deleterious pleiotropic effects associated with physical interactions among transcription factors.

Adaptation of transcriptional circuits under positive selection

When cooperative binding is present, under stabilising selection, transcription factor binding sites at co-regulated genes are better able to tolerate mutations (i.e $a_{hs} > a_s$). Under positive selection for a novel expression phenotype, this may speed adaptation, since greater mutational robustness generates greater genetic diversity and can help speed adaptation (Figure 3a) [22]. This may occur, for example, when adaptation involves change in the transcription factor that regulates a target gene [7-9,11], through turnover of transcription factor binding sites [23-25]. We use our model to quantify the extent to which cooperative binding among transcription factors accelerates the adaptive rewiring of transcriptional circuits under positive selection.

We study adaptive change that involves replacement of an existing transcription factor by a new one that confers higher fitness. We assume that the target gene must first suffer an initially deleterious mutation at its existing binding site before a newly adaptive binding site can be acquired (Figure 3) [8,9,11]. The newly adaptive binding site is produced from binding sites that have already

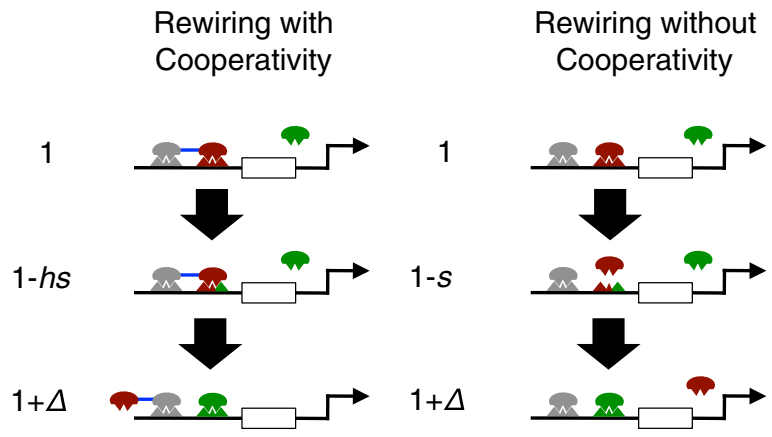

Figure 3 A schematic cartoon of rewiring. A schematic cartoon of rewiring with (left) and without (right) cooperative binding. Selection favours a change in the regulation of target genes from the red TF to the green TF. Rewiring requires an initially deleterious mutation at the red binding site before a green binding site can be acquired. The fitness of the different states is shown on the left hand side for each case. The reduced fitness of the intermediate state is less when cooperative binding is present than when it is absent.

mutated at a rate u_r. The expected waiting time for such a gene to produce a newly adaptive binding site therefore depends on the number of binding sites in the population that harbor a deleterious mutation, which is proportional to a_s when cooperativity is absent and a_{hs} when it is present. Since $a_{hs} > a_s$, this number is greater when cooperative binding is present than when it is absent.

The ratio of waiting times before a newly adaptive binding site arises, t_r^*/t_r (for populations without, t_r^*, or with, t_r, cooperative binding), quantifies the degree to which cooperative binding of transcription factors accelerates adaptation under positive selection. This ratio is given by a_{hs}/a_s (Figure 4, see Methods). As Figure 4 shows, provided $Ns > 1$ (i.e. provided deleterious mutations at binding sites are not nearly neutral), rewiring of transcriptional circuits is significantly accelerated by cooperative binding among transcription factors. Thus, a population that has cooperative binding among transcription factors under stabilizing selection, can also experience an accelerated rate of adaptation.

Cooperative binding and the fraction of shared targets in yeast

Our model predicts that, under stabilising selection, cooperative binding will be favoured when the fraction of targets shared by a pair of transcription factors exceeds a certain threshold. In order to test this prediction, and to get some idea of the degree of overlap that is required for cooperative binding to arise in natural systems, we inspected pairs of transcription factors in *Saccharomyces cerevisiae*. A total of 186 pairs are reported as participating in cooperative binding [26], based on a combination of ChIP-chip data, transcription factor knockout data, and direct experimental evidence. Using the set of genes regulated through a transcription factor binding site for a total of 204 yeast transcription factors [27,28], we determined the fraction of overlapping targets, β, for all pairs of transcription factors (Figure 5). It is important to note that, typically, studies that systematically look for cooperative gene interactions take into account the number of targets shared by a gene pair. Therefore, to minimise the risk of circularity in our analysis, we have used separate datasets to determine cooperative gene interactions, and to determine regulatory targets. The mean fraction of overlapping targets for genes identified as participating in cooperative binding was 10-fold greater (0.21) than the mean fraction of overlapping targets at genes that do not bind cooperatively (0.02) which is highly statistically significant ($p < 2 \times 10^{-16}$, Wilcoxon test). This supports the prediction of our population-genetic analysis, and it suggests that a sizeable overlap in targets is required before cooperative binding becomes advantageous.

Cooperative binding in the yeast sex determination network

The ability of cooperative transcription factors to facilitate adaptation also has empirical support, from observations in the sex determination networks of different yeast species [7-9]. The acquisition of a protein-protein interaction between the mating factor MATα2 and Mcm1 was able to buffer the deleterious effects of mutations that strengthened Mcm1 binding sites [7]. Prior to the emergence of a protein-protein interaction, sex determining genes were activated only in the presence of Mcm1 and MATa2 together [7]. The buffering effects of the protein-protein interaction allowed Mcm1 binding sites to acquire strengthening mutations such that sex determining genes became activated by Mcm1 alone. As a

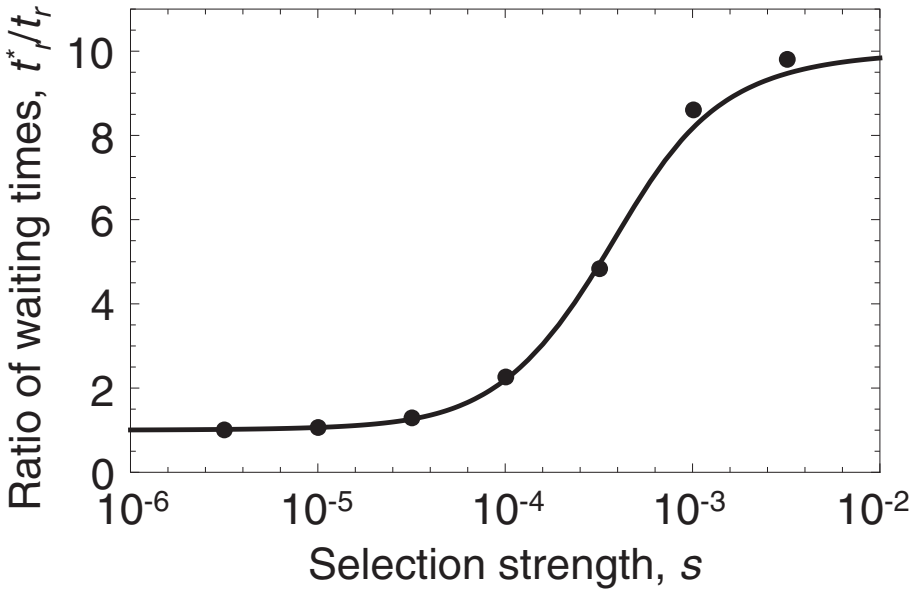

Figure 4 Cooperative binding accelerates adaptation. Cooperative binding accelerates adaptation under positive selection. The ratio of waiting times before the arrival of novel adaptive binding sites for populations without (t^*_r) and with (t_r) cooperative binding. Provided $Ns > 1$, cooperative binding reduces the adaptation time up to 10-fold, compared to populations that lack cooperative binding. The line shows our analytic expression, and points show the result of 10^5 replicate Monte-Carlo simulations. Parameter values $u_l = 2 \times 10^{-7}$, $u_g = 10^{-7}$, $K_1 + K_2 = 100$, $h = 10^{-1}$, $t = 10^{-4}$, $N = 10^4$, $u_r = 10^{-7}$.

result, MATa2 became redundant and was lost [7]. The result was a significant upstream reorganization of the yeast sex determination network without the need for any parallel changes to the downstream output of the network. Similar patterns, in which acquisition of cooperative binding between transcription factors is followed by changes to the regulation of their shared targets, are observed across the yeast transcriptome [8], and support the prediction of our analysis of positive selection on transcriptional networks.

Conclusions

We have shown that cooperative binding between a pair of transcription factors is favoured under stabilising selection, provided the overlap between their targets is sufficiently large. The threshold fraction of shared targets depends upon the strength of selection on binding sites, the cost of pleiotropy associated with protein-protein interactions, and the rates of mutations. It also depends on the population size. Just as in models that consider the evolution of redundancy [20,29], we find that

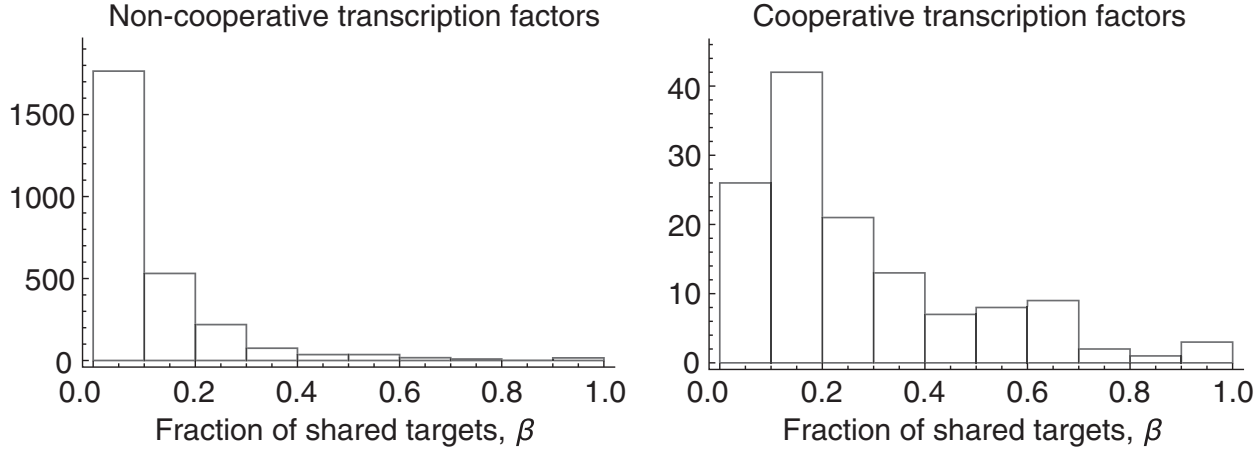

Figure 5 Number of shared targets. Fraction of targets that are shared between pairs transcription factors in *S. cerevisiae* [26-28]. (left) The fraction of targets that are shared among paris of transcription factors that lack cooperative binding and (right) the fraction of targets that are shared among transcription factors that bind cooperatively. The fraction of targets that are shared is larger among cooperative factors ($p < 2 \times 10^{-16}$, Wilcoxon test).

greater redundancy (i.e. cooperative regulation) is more strongly favoured in smaller populations; and that for intermediate population sizes the evolutionary dynamics are bistable, such that cooperative binding is maintained if it is already present, but cannot evolve if it is absent. Finally, we found that cooperative binding facilitates the rewiring of transcriptional circuits under positive selection.

This study shows that, even when the deleterious effects of pleiotropy are taken into account, mutations that change transcription factor function can play an important role in the evolution of gene expression. Taking account of mutations both to regulatory binding sites and to the transcription factors themselves reveals a rich set of evolutionary dynamics that helps explain how complex transcriptional networks can rapidly rewire large sets of genes in order to adapt to new environments.

Methods

Equilibrium distribution

To find the equilibrium relative abundances of the hamming classes x_i that give the solution to Equation 1, we follow [20,21] and look for a solution of the form $x_i = \binom{K_1+K_2}{i} a_s^i (1-a_s)^{K_1+K_2-i}$. Given this assumed form, the mean fitness of the population at equilibrium is $\bar{\omega} = \sum_i (1-s)^i x_i$. Since $\sum_i x_i = 1$ it is easy to show that $\sum_i \binom{K_1+K_2}{i} p^i = (1+p)^n$. Taking $p = a(1-s)/(1-a)$, this gives a mean fitness of $\bar{\omega} = (1-a_s s)^{(K_1+K_2)}$, which is the form given in the main text.

To compute a_s we follow [20] and write down the generating function $\Pi_V(z)$ of a random variable V defined by the Hamming class after mutation of an individual chosen from the population according to its relative fitness, where z is a formal variable. The function $\Pi_V(z)$ may be thought of as a probability generating function, where the probability distribution associated with it gives the distribution of Hamming classes in the population at equilibrium [20]. In the case of non-identical forward and backward mutation rates, u_l and u_g, this is given by $\Pi_V(z) = \sum_i w_i x_i (u_l + (1-u_l)z)^i ((1-u_g)+u_g z)^{K_1+K_2-i}$. Following [20], we analyse the eigensystem problem associated with the population dynamics to determine the equilibrium distribution of Hamming classes (i.e the distribution of genotypes in the population under mutation selection balance). In equilibrium we have $\Pi_V(z) = \lambda \sum_i x_i z^i$, where λ is the eigenvalue associated with the system. Using our assumed form of x_i results in the infinite population equilibrium distribution:

$$a_s = \frac{1}{2}\left[1 + \frac{u_l+u_g}{s} - u_g - \sqrt{\left(1+\frac{u_l+u_g}{s}-u_g\right)^2 - \frac{4u_l}{s}}\right] \tag{5}$$

When cooperative binding is present a subset $\beta(K_1+K_2) = K_{hs}$ of the target genes have selective coefficient hs and the remaining $(1-\beta)(K_1+K_2) = K_s$ have selective coefficient s. The hamming class of an individual now has two indices i and j such that x_{ij} refers to an individual with i mutations at shared targets and j mutations at unshared targets. In this case we look for solutions of the form $x_{ij} = \binom{K_{hs}}{i} a_{hs}^i (1-a_{hs})^{K_{hs}-i} \binom{K_s}{j} a_s^j (1-a_s)^{K_s-j}$. The generating function of V is now given by

$$\begin{aligned}\Pi_V(z_{hs}, z_s) = \sum_i \sum_j & w_{ij} x_{ij} (u_l + (1-u_l) z_{hs})^i \\ & \times ((1-u_g) + u_g z_{hs})^{K_{hs}-i} (u_l + (1-u_l) z_s)^j \\ & \times ((1-u_g) + u_g z_s)^{K_s - j}\end{aligned}$$

and at equilibrium $\Pi_V(z_{hs}, z_s) = \lambda \sum_i \sum_j x_{ij} z_{hs}^i z_s^j$. Because we are assuming that w_{ij} is just the product of the two independent fitness landscapes associated with the different selective coefficients, i.e $w_{ij} = (1-hs)^i (1-s)^j$ using our assumed form of x_{ij} results in values of a_s and a_{hs} as given by Equation 5 for the independent distributions with the appropriate selective coefficients.

The finite N approximation of Equation 5, can be obtained from the moment equations of Woodcock and Higgs [21], assuming $u_l, u_g, s, N^{-1} \ll 1$. This gives

$$a_s = \frac{1}{2}\left[1 + \frac{1+2(u_l+u_g)}{2Ns} - \sqrt{\left(1+\frac{1+2(u_l+u_g)}{2Ns}\right)^2 - \frac{4u_l}{u_l+u_g}\frac{1+2(u_l+u_g)}{2Ns}}\right]. \tag{6}$$

Assuming $u_l, u_g \ll s$, we obtain the Taylor expansion of a_s to first order, in terms of $1/(2Ns)$ (which is relevant when $2Ns > 1$) and in terms of $2Ns$ (relevant when $2Ns < 1$) to obtain Equation 2.

Using the above distributions, the equilibrium mean fitness w_{ind}, in the absence of cooperative binding is $w_{ind} = (1-a_s s)^{K_1+K_2}$, and in the presence of cooperative binding, w_{coop} is $w_{coop} = (1-t)^{(1-\beta)(K_1+K_2)}(1-a_s s)^{(1-\beta)(K_1+K_2)}(1-a_{hs} hs)^{\beta(K_1+K_2)}$ which can be Taylor expanded to first order and, combined with Equation 2, give Equations 3 and 4. We can also find the conditions for the equilibrium mean fitness of a population with cooeprative binding to be greater than that for a population that lacks it, i.e $w_{coop} > w_{ind}$. When $t, s \ll 1$ this can be expressed as

$$\beta > \begin{cases} \frac{2Nt}{2Nt+a_0(1-2Nhs)} & \text{if } 2Ns > 1 \text{ and } 2Nhs < 1 \\ 1 & \text{otherwise} \end{cases} . \quad (7)$$

According to this inequality, cooperative binding is advantageous only when the fraction of targets shared by the pair of transcription factors is greater than a threshold. Since by definition $\beta \leq 1$, Equation 7 says that cooperative binding can only increase population mean fitness when $2Nhs < 1$ and $2Ns > 1$, i.e if the benefit of cooperativity, h, is sufficient to make mutations at transcription factor binding sites that are deleterious in the absence of cooperativity nearly neutral when it is present.

Rewiring time

For a given binding site the waiting time T for the arrival of the first adaptively rewired mutant to arise is given by the distribution

$$\mathbb{P}\{T > t\} = \mathbb{E}\left[e^{-\zeta Yt)}\right],$$

where t is time, ζ is the rate at which the rewiring mutations occur and Y is fraction of the population at equilibrium that is able to undergo rewiring mutations. We assume rewiring mutations can occur only following an initially deleterious mutation. In the absence of cooperativity, the fraction of the population with a mutation at a given site is a_s, therefore $Y \propto Na_s$, and we are able to write

$$\mathbb{P}\{T > t\} = \mathbb{E}\left[e^{-u_r N a_s t}\right],$$

where u_r gives the rate at which rewiring mutations occur at sites that have already undergone an initially deleterious mutation. The excepted waiting time for a single gene is thus

$$T_s = \frac{1}{u_r N a_s}$$

If the gene to be rewired is coregulated by a pair of transcription factors that bind cooperatively, we similarly have

$$T_{hs} = \frac{1}{u_r N a_{hs}}$$

and the ratio of waiting times T_s/T_{hs} is therefore simply a_{hs}/a_s. Finally, if k genes must be rewired before adaptation occurs, the waiting time for the first event is T/k (where T is the waiting time for 1 gene to rewire), the waiting time for the second event is $T/(k-1)$ and so on. Therefore the total expected waiting time, $T_s(k)$ is

$$T_s(k) = \frac{1}{u_r N a_s} \sum_{i=1}^{k} \frac{1}{i}$$

Therefore the ratio of expected waiting times with and without cooperative binding is independent of the number of genes to be rewired, and depends only on the ratio a_{hs}/a_s.

Variation in selection strength across sites

Up to this point we have assumed that the selective coefficients, s and h, are constant across binding sites. However it is obviously possible that these parameters may vary between binding sites. Such a generalization of our model represents a significant complication, and a full treatment is outside the scope of this paper. However, it can be analyzed in the simple case that the coefficients associated with each binding site i satisfy $s_i \ll 1$, such that the fitness landscape is approximately additive.

We assume that there are a finite set of selective coefficients, s^α and h^γ, where the super-scripts α and γ index the different sets, and that the number of binding sites with a given coefficient s^α or h^γ are distributed according to some function $F(s)$ and $G(h)$. We also assume that the coefficients s and h are distributed independently of one another. Each binding site i has a value s_i associated with it, drawn according to $F(s)$. In the quasi-species regime the probability that binding site i has a mutation is simply given by a_{s_i}, as given by Equation 5. In this case the distribution of hamming classes, rather than being binomial, is poisson binomial, paramaterized by a_{s_i}. Similarly, when cooperative binding is present, the distribution is poisson binomial with the modification that shared targets have a mutation with probability $a_{s_i h_i}$ where h_i is drawn independently from the distribution $G(h)$. The system is easiest to analyse if we separate binding sites into sub-classes, α, of binding sites with the same selective coefficients, where the size of each sub-class, n_α is given by $n_\alpha = F(s^\alpha)(K_1+K_2)$. The number of mutations in each subclass is then given by a binomial distribution.

When the fitness landscape is close to additive, the method of [19] can be applied independently to each subclass to determine the expected number of mutations in the sub-class. This is only true in an additive fitness landscape, or in a multiplicative fitness landscape in which cross terms between subclasses are sufficiently small that they can be neglected. When this condition holds, the value of a_{s^α} associated with each sub-class is given by Equation 6.

The expected number of mutations in each subclass is simply $a_{s^\alpha} n_{s^\alpha}$ and the expected fitness of each sub-class is $(1 - a_{s^\alpha} s^\alpha)^{n_\alpha}$. The expected mean fitness of the population is then $\bar{\omega} = \prod_\alpha (1 - a_{s^\alpha} s^\alpha)^{n_\alpha}$. Using our almost additive assumption, this can be approximated by $\bar{\omega} \approx 1 - (K_1 + K_2) \sum_\alpha F(s^\alpha) a_{s^\alpha} s^\alpha$. This is the fitness of the population when cooperative binding is absent. Similarly, when cooperative binding is present we have,

$$\begin{aligned} w_{coop} \approx 1 &- (1-\beta)(K_1+K_2)t - (1-\beta)(K_1+K_2) \\ &\times \sum_\alpha F(s^\alpha) a_{s^\alpha} s^\alpha + \beta(K_1+K_2) \\ &\times \sum_\alpha \sum_\gamma F(s^\alpha) G(h^\gamma) a_{h^\gamma s^\alpha} h^\gamma s^\alpha \end{aligned}$$

From these the invasion probabilities and threshold values of β can be calculated in the same way as in the case of constant s and h above. The only difference in the case with variable selective coefficients is that the invisibility criteria for a mutation resulting in gain or loss of cooperative biding dependants on the average of $a_s s$ across the distribution $F(s)$, and on the average $a_{hs} hs$ across the joint distribution $F(s)G(h)$. Investigating different forms of the functions $F(s)$ and $G(h)$ represents an interesting avenue for further work.

Authors' contributions

AJS, AP, RMS, and JBP designed the research. AJS and RMS performed the analysis. AJS, AP, and JBP wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

JBP and AJS acknowledge funding from the Burroughs Wellcome Fund, the David and Lucile Packard Foundation, the James S. McDonnell Foundation, the Alfred P. Sloan Foundation, grant #D12AP00025 from the U.S. Department of the Interior and Defense Advanced Research Projects Agency, and grant RFP-12-16 from the Foundational Questions in Evolutionary Biology Fund. AP acknowledges grants from the Natural Environment Research Council (NE/G00563X/1) and the Engineering and Physical Sciences Research Council (EP/F500351/1, EP/I017909/1). AJS also acknowledges an EPSRC PhD Plus fellowship.

Author details

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. [2]CoMPLEX, University College London, London, UK. [3]Department of Mathematics, University College London, London, UK. [4]Department of Genetics, Evolution and Environment, University College London, London, UK.

Received: 13 April 2012 Accepted: 31 August 2012
Published: 6 September 2012

References

1. Carroll SB: **Evolution at two levels: on genes and form.** *PLoS Biol* 2005, **3**(7):e245.
2. Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, Barkai N: **Rewiring of the yeast transcriptional network through the evolution of motif usage.** *Science* 2005, **309**(5736):938–940.
3. Lemos B, Araripe LO, Fontanillas P, Hartl DL: **Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression.** *Proc Natl Acad Sci U S A* 2008, **105**(38):14471–14476.
4. Prud'homme B, Gompel N, Carroll SB: **Emerging principles of regulatory evolution.** *Proc Natl Acad Sci U S A* 2007, **104**(Suppl 1):8605–8612.
5. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB: **Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene.** *Nature* 2006, **440**(7087):1050–1053.
6. Stern DL: **Evolutionary developmental biology and the problem of variation.** *Evolution* 2000, **54**(4):1079–1091.
7. Tsong AE, Tuch BB, Li H, Johnson AD: **Evolution of alternative transcriptional circuits with identical logic.** *Nature* 2006, **443**(7110):415–420.
8. Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD: **The Evolution of Combinatorial Gene Regulation in Fungi.** *PLoS Biol* 2008, **6**(2):e38.
9. Tuch BB, Li H, Johnson AD: **Evolution of eukaryotic transcription circuits.** *Science* 2008, **319**(5871):1797–1799.
10. Wray GA: **The evolutionary significance of cis-regulatory mutations.** *Nat Rev Genet* 2007, **8**(3):206–216.
11. Lynch VJ, Wagner GP: **Resurrecting the role of transcription factor change in developmental evolution.** *Evolution* 2008, **62**(9):2131–2154.
12. Wagner GP, Lynch VJ: **The gene regulatory logic of transcription factor evolution.** *Trends Ecol Evol* 2008, **23**(7):377–385.
13. Chu D, Zabet NR, Mitavskiy B: **Models of transcription factor binding: sensitivity of activation functions to model assumptions.** *J Theor Biol* 2009, **257**(3):419–429.
14. Berg J, Willmann S, Lässig M: **Adaptive evolution of transcription factor binding sites.** *BMC Evol Biol* 2004, **4:**42.
15. He X, Samee MAH, Blatti C, Sinha S: **Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression.** *PLoS Comput Biol* 2010, **6**(9).
16. Gerland U, Moroz JD, Hwa T: **Physical constraints and functional characteristics of transcription factor-DNA interaction.** *Proc Natl Acad Sci U S A* 2002, **99**(19):12015–12020.
17. Moses AM, Chiang DY, Kellis M, Lander ES, Eisen MB: **Position specific variation in the rate of evolution in transcription factor binding sites.** *BMC Evol Biol* 2003, **3:**19.
18. Hahn MW, Stajich JE, Wray GA: **The effects of selection against spurious transcription factor binding sites.** *Mol Biol Evol* 2003, **20**(6):901–906.
19. Higgs PG, Woodcock G: **The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection.** *J Math Biol* 1995, **33**(7):677–702.
20. Krakauer DC, Plotkin JB: **Redundancy, antiredundancy, and the robustness of genomes.** *Proc Natl Acad Sci U S A* 2002, **99**(3):1405–1409.
21. Woodcock G, Higgs PG: **Population evolution on a multiplicative single-peak fitness landscape.** *J Theor Biol* 1996, **179:**61–73.
22. Draghi J, Parsons T, Wagner G, Plotkin J: **Mutational robustness can facilitate adaptation.** *Nature* 2010, **436:**353–355.
23. Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, Biggin MD, Eisen MB: **Large-Scale Turnover of Functional Transcription Factor Binding Sites in Drosophila.** *PLoS Comput Biol* 2006, **2**(10):e130.
24. Stewart AJ, Seymour RM, Pomiankowski A: **Degree dependence in rates of transcription factor evolution explains the unusual structure of transcription networks.** *Proc R Soc B Biol Sci* 2009, **276**(1666):2493–2501. [doi:10.1098/rspb.2009.0210]
25. Venkataram S, Fay JC: **Is transcription factor binding site turnover a sufficient explanation for cis-regulatory sequence divergence?** *Genome Biol Evol* 2010, **2:**851–858.
26. Yang Y, Zhang Z, Li Y, Zhu XG, Liu Q: **Identifying cooperative transcription factors by combining ChIP-chip data and knockout data.** *Cell Res* 2010, **20**(11):1276–1278.
27. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**(7004):99–104.
28. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA: **Transcriptional regulatory networks in Saccharomyces cerevisiae.** *Science* 2002, **298**(5594):799–804.
29. Rajon E, Masel J: **Evolution of molecular error rates and the consequences for evolvability.** *Proc Natl Acad Sci U S A* 2011, **108**(3):1082–1087.

doi:10.1186/1471-2148-12-173
Cite this article as: Stewart *et al.*: The population genetics of cooperative gene regulation. *BMC Evolutionary Biology* 2012 **12**:173.