The evolution of genetic networks by non-adaptive processes

Michael Lynch

Abstract | Although numerous investigators assume that the global features of genetic networks are moulded by natural selection, there has been no formal demonstration of the adaptive origin of any genetic network. This Analysis shows that many of the qualitative features of known transcriptional networks can arise readily through the non-adaptive processes of genetic drift, mutation and recombination, raising questions about whether natural selection is necessary or even sufficient for the origin of many aspects of gene-network topologies. The widespread reliance on computational procedures that are devoid of population-genetic details to generate hypotheses for the evolution of network configurations seems to be unjustified.

Although the ubiquity of genetic pathways underlying metabolic and developmental processes is beyond dispute, the mechanisms by which genetic networks become established evolutionarily are far from clear. Many physicists, engineers and computer scientists, and some cell and developmental biologists, are convinced that biological networks exhibit properties that could only be products of natural selection (for example, REFS 1–5); however, the matter has rarely been examined in the context of well-established evolutionary principles. Five popular concepts in biology today — redundancy, robustness, modularity, complexity and evolvability — invoke a vision of the cell as an electronic circuit, designed by and for adaptation. Terms like plug-ins, nucleation kernels, input and output switches, capacitance and hard wiring abound. Alon states that it is "…wondrous that the solutions found by evolution have much in common with good engineering,"[6] and Adami states that digital organisms allow "…researchers to address fundamental questions about the genetic basis of the evolution of complexity, genome organization, robustness and evolvability."[7]

Is this reduction of the field of evolution to a subdiscipline of engineering justified? For some, the answer is far from a definitive yes[8,9]. The population-level processes and cellular limitations that dictate a genome's capacity for evolution are quite distinct from the constraints that govern the human design of non-biological constructs[10,11]. In addition, although there is little question that adaptive exploitation of networks has occurred in numerous instances (for example, REF. 12), and the final products of pathways may almost always be of adaptive value, the physical mechanisms that give rise to genome architectural features are logically distinct from the adaptive processes that utilize such features as evolutionary resources[13]. Contrary to some suggestions (for example, REF. 14), there is no evidence that genetic pathways emerge *de novo* in response to a selective challenge. Rather, pathway evolution probably proceeds by a gradual augmentation, making use of mutational variation arising independently at different loci, as occurs in nearly all other evolutionary processes.

Qualitative observations suggest that the complexity of regulatory and protein-interaction networks increases from prokaryotes to unicellular eukaryotes to multicellular eukaryotes, with simple autoregulatory loops being more common and multicomponent loops being less common in microbes, although the existing data are known to be subject to numerous potential biases[15–17]. Moreover, it is an open question as to whether pathway complexity is a necessary prerequisite for the evolution of complex phenotypes, or whether the genome architectures of multicellular species are simply more conducive to the passive emergence of network connections. Given the large numbers of transcription factors in most cells and their reliance on simple binding sites that are subject to stochastic mutational turnover, there are many plausible mechanisms for the emergence of novel intracellular transactions by effectively neutral processes[13,18–20].

The following analyses show the ease with which commonly observed features of genetic pathways can emerge without any direct selection for such properties,

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.
e-mail: milynch@indiana.edu
doi:10.1038/nrg2192

Table 1 | **Glossary of mathematical notation**

Notation	Definition
c	Rate of recombination per base pair
d	Number of nucleotides separating two recombining sites
k	Number of potential transcription factors in a higher-order system
L	Total number of base pairs available for the origin of a novel TFBS
n	Number of nucleotide sites in a TFBS
N	Effective population size
p_s, p_r, p_u	Equilibrium frequencies of the self-sufficient, redundantly regulated and upstream-dependent alleles
r	Expected rate of crossing over between two sites separated by distance d
u	Mutation rate per base pair per generation
u_g	Rate of gain of a TFBS per gene per generation
u_l	Rate of loss of a TFBS per gene per generation

TFBS, transcription-factor binding site.

under conditions that are known to exist in natural populations. Such observations have significant implications for the widespread reliance on the adaptive paradigm for understanding network evolution. The demonstration that the relative power of the non-adaptive forces of evolution — genetic drift, mutation and recombination — define the trajectories that are open to evolutionary exploitation will raise questions regarding previous conclusions on network evolution derived from models that are devoid of population-genetic details. The demonstration that redundantly regulated genetic pathways can arise by very different processes in small versus large populations will raise doubts about the justification for the search for universal adaptive explanations for the evolution of genetic redundancy. Furthermore, the demonstration that observed distributions of node connectivity in microbial species can emerge passively by mutational processes will contradict the idea that such patterns reflect some sort of universal trend toward adaptive design (for example, REFS 2,21). Finally, the conclusion that convergent evolution of network architectures in distantly related microbes provides compelling evidence for 'optimal design'[22] will also be shown to be questionable.

Although contrarian in tone, the models presented here provide the seeds for the development of biologically realistic null hypotheses for the origins of pathway complexity, a tool that many feel is essential to the development of a rigorous theory for network evolution[8,9,23–25]. If we are to be confident that the architectural features of genetic networks are advanced by natural selection, it should be possible to formally reject the possibility of neutral evolution. There is room for disagreement with what constitutes an appropriate null model for network evolution but, given that evolution is a population-level process, we start with the assumption that any such model must incorporate the fundamental principles of population genetics. Such an approach is a significant departure from most previous work in this area.

The case for neutral pathway evolution

Three observations motivate the hypothesis that a considerable amount of regulatory-pathway evolution is driven by non-adaptive processes. First, one of the most puzzling aspects of many genetic pathways is their seemingly baroque structure[12,26]. In multicellular species, it is common for linear pathways to consist of a series of genes encoding products that are essential to the activation or deactivation of the next downstream member, with only the final gene product in the series yielding a direct phenotypic effect that can serve as a potential target of selection. For example, the product of gene D might be necessary to turn on gene C, the product of which is necessary to turn on gene B, the product of which finally turns on gene A. Pathways involving only inhibitory steps also exist, and these lead to genes with an alternating series of high and low expression, depending on the activation state of the first gene in the pathway. For example, gene D might produce a product that inhibits the expression of gene C, silencing of which allows gene B to be turned on, which inhibits the expression of gene A. It is often unclear whether such complexity has any advantages over the simple constitutive expression or self-regulation of the final member in the pathway.

Second, many lines of evidence support the idea that the regulatory machinery underlying complex adaptations is capable of undergoing frequent and dramatic shifts without altering the outwardly expressed phenotype[4,27–36]. Even cellular features involving extraordinarily conserved protein functions are subject to substantial evolution at the regulatory level. For example, although the core histone proteins have nearly invariant sequences across all eukaryotes, they are highly divergent with respect to regulatory mechanisms, with those from animals having many more *cis*-regulatory binding sites than those from yeast[37]. Similarly, despite their strong protein-sequence conservation, the ribosomal protein genes of yeast species have evolved dramatically different mechanisms of regulation, apparently through neutral intermediate stages of redundant regulation[38]. Dramatic changes in modes of regulation also exist among orthologous prokaryotic genes[3,39] and, although natural selection has been invoked as an explanation for such alterations[3], no direct evidence in support of this hypothesis has emerged. Such gene-regulation remodelling is entirely consistent with our knowledge of transcription factors as renewable resources. The genomes of microbial species contain up to a few hundred transcription-factor genes, whereas multicellular eukaryotes can harbour well over 1,000 (REFS 40–42). Because genes duplicate at rates of 0.1–1% per generation per haploid genome, and because redundant genes are susceptible to the evolution of novel functions[13], each transcription-factor gene will be liable to undergo thousands of functional changes at the population level on a timescale of 10^6 generations. In addition, because transcription-factor binding sites are typically small (~ 5 –10 bp[12,43]), *de novo* origins of novel sites by mutation are expected to be frequent[44,45].

Third, as will be shown below, a key determinant of the evolutionary patterning of alternative genetic pathways is the ratio of rates at which regulatory sites are lost and gained by mutational processes, $\alpha = u_l/u_g$ (see TABLE 1 for a summary of the mathematical terms used in this Analysis). Although direct estimates of α do not exist, some informed inferences can be made. The loss rate of a transcription-factor binding site involving n nucleotides is simply $u_l = nu$, where u denotes the per-nucleotide mutation rate. As it is conditional on u and n, this expression should be applicable to all species. By contrast, the rate of gain of binding sites will be a function of the total span of DNA within which such sites can originate, which is denoted by L. Because of the broad range of variation in mutational target sizes (that is, the amount of DNA that is not allocated to essential coding functions), L can vary by orders of magnitude among species. The number of potential gain sites is $L–n+1$, assuming no prior occupancy. Assuming equal frequencies of all four nucleotides, the probability that a span of n sites deviates from the requisite binding site signature by just a single nucleotide is $n{\cdot}0.75{\cdot}0.25^{n-1}$. Thus, with each one-off site mutating to the appropriate sequence at a rate of $u/3$, the expected rate of gain of regulatory sites is $u_g = (L–n+1)nu/4^n$, which implies that $\alpha \approx 4^n/(L–n+1)$. With n = 5–10, 4^n is in the range of 10^3 to 10^5. If it is assumed that transcription-factor binding sites must reside in the non-coding DNA that lies adjacent to a gene (or within its introns), L will rarely exceed 100 bp in prokaryotic species, but might be as high as 10^4–10^6 bp in multicellular eukaryotes[13]. So, for prokaryotes, α for any particular transcription factor is likely to be in the range of 10 to 10^4 whereas, for multicellular eukaryotes, the range is more on the order of 0.001 to 10, with vertebrates and land plants being near the lower limit of this range. This implies that, from a physical birth and death perspective alone, genes from multicellular species are expected to harbour many more transcription-factor binding sites than their orthologues in unicellular lineages.

Recruitment of an upstream activator

We first examine the simplest possible case — a focal gene A that can be entirely self-reliant, redundantly regulated by an upstream factor B, or entirely dependent on the upstream factor (FIG. 1). As we are concerned only with the recruitment of a new activator of gene A, self-reliance simply implies that the expression of gene A is regulated by a factor other than B (which could include or be restricted to A itself). For ease of explanation, we will assume that the gain of a mode of regulation involves the acquisition of a specific transcription-factor binding site, here assumed to occur at rate u_g per gene copy per generation, regardless of the regulating factor, whereas individual regulatory sites are mutationally inactivated at rate u_l. The only constraint on transcription factor B is that it must be present at a sufficient level to drive the expression of A in all cellular contexts that are crucial to organismal fitness. Mutations that equally influence all three possible modes of expression (for example, coding-region mutations) can be ignored, as they have no effect on relative allelic success.

For convenience, a haploid population will be assumed throughout, with the effective population size being denoted by N. However, most of the results also apply to diploid populations of size $N/2$, assuming that mutations have additive effects on fitness. Our focus will be on the long-term average frequencies of the three alternative viable genotypic states, all of which are assumed to be equivalent with respect to fitness, other than the differential vulnerability to mutations. Self-reliant and upstream-dependent alleles can be converted to null alleles by single loss-of-function mutations, whereas redundantly regulated alleles survive such mutations, in that they are simply converted to one of the two alternative functional allelic types (FIG. 1).

Most of the focus will be on situations in which the birth–death dynamics of regulatory sites are such that the simultaneous residence of more than one binding site for the upstream factor B is unlikely. This is a reasonable assumption in that, for yeast, few transcription-factor binding sites reside more than 1 kb upstream of the translation start site of a gene[43]. In the absence of selection, a span of L nucleotides is expected to reach an equilibrium number of binding sites for a particular transcription factor, equal to $\sim L/4^n$, provided that $n \leq 10$ and $L >> n$. Thus, if n = 5, a region of $L = 10^3$ nucleotides

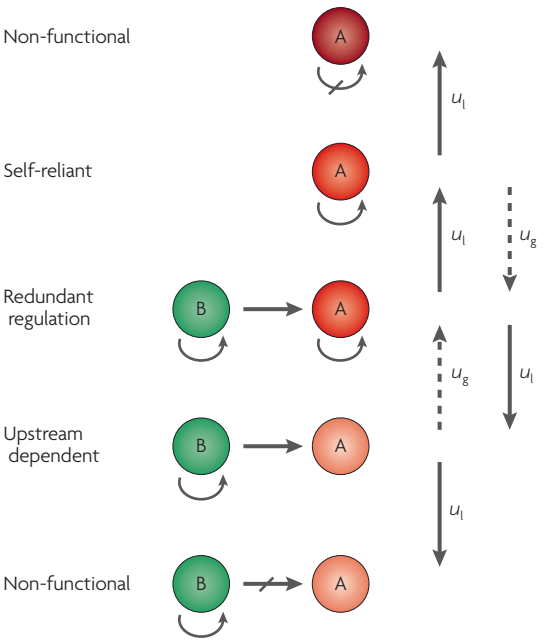

Figure 1 | **A model for the recruitment of an upstream activator B.** Gene B is assumed to be essential for reasons other than for the function for which it is recruited here, and hence it remains fixed in the population. A curved arrow denotes the ability of gene A to self-regulate, a straight arrow denotes regulation by B and hash marks denote loss of a regulatory mechanism. Gain and loss rates are defined by mutational processes that give rise (or loss) to transcription-factor binding sites; gain rates are a function of the local amount of DNA-level substrate for the origin of such sites. Non-functional alleles are assumed to be lethal and would be eliminated from the population by selection.

Effective population size
A scaled measure of the size of a natural population that is relevant to population genetics. This value is equivalent to the size of the idealized, random-mating population that gives equivalent allele-frequency dynamics, and is generally one or more orders of magnitude smaller than the actual population size.

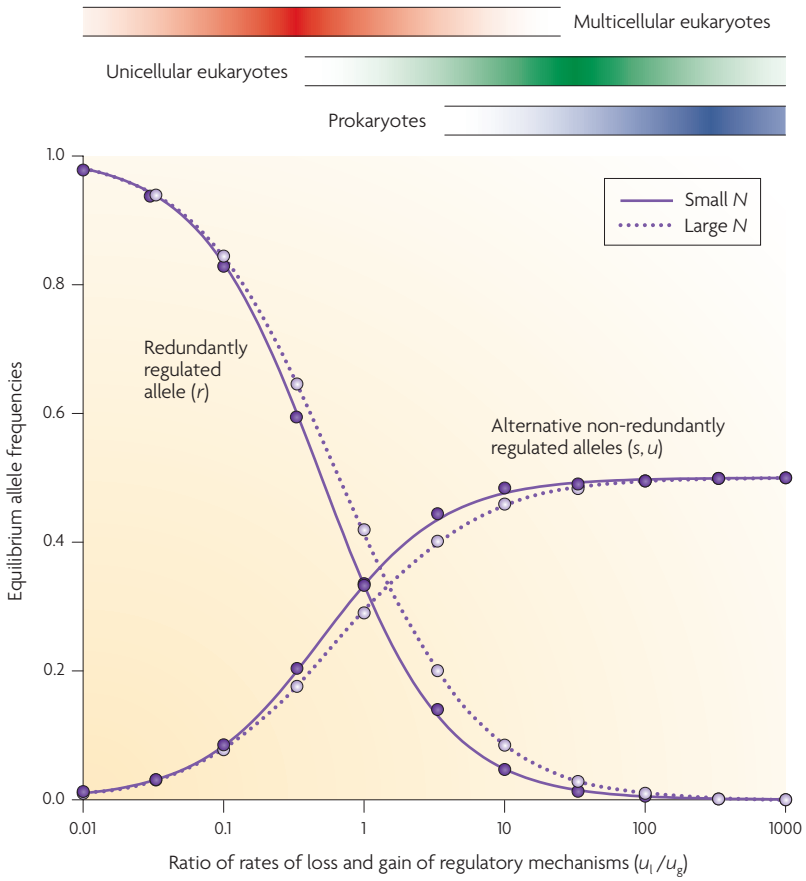

Figure 2 | **The likelihood of alternative modes of gene regulation depends strongly on the relative rates of loss and gain of transcription-factor binding sites.** The expected frequencies of the three alternative types of regulation are shown, under the assumptions of complete linkage of the DNA-level elements that are involved in the activation of a gene. If the two rates of gain of regulatory mechanisms are assumed to be equal, as they are here, the model is symmetrical and the self-reliant and upstream-dependent alleles have equal frequencies. The curves are derived from the theoretical approximations given in the text. The data points are derived from stochastic simulations, which explicitly incorporate per-generation allele-frequency changes resulting from mutation, drift and selection, averaging over many millions of generations. The approximate domains of the parameter u_l/u_g, on the basis of known genomic features of various phylogenetic groups (see text), are denoted in the upper margin.

is expected to harbour zero copies or one copy of a particular binding site by chance, whereas, if $n = 10$, regions smaller than $L = 10^6$ nucleotides are unlikely to harbour more than one such site.

The weak mutational advantage of regulatory redundancy. If the effective population size is sufficiently small that both Nu_g and Nu_l are <1, the population will generally be nearly monomorphic for one of the alternative states, as the average time to fixation of a neutral allele (~2N generations) will be less than the arrival time of mutations that are destined for fixation ($1/u_g$ and $1u_l$ for gain and loss mutations, respectively). When these conditions are met, because selection can operate only in a variable population, the advantage that is associated with the reduced inactivation rate of redundantly regulated alleles will be rendered ineffective.

The population-genetic conditions that are necessary for such a setting seem to be common. From the results given above, $Nu_l < 1$ requires that $Nun < 1$, and $Nu_g < 1$ requires that $NuLn/4^n < 1$. These relationships are useful because estimates of Nu are available for many species, averaging 0.002 for vertebrates, 0.008 for land plants, 0.013 for invertebrates, 0.029 for unicellular eukaryotes and 0.052 for prokaryotes; the last two values are likely to be downwardly biased, perhaps as much as tenfold[13]. Thus, taking $n = 8$ as the approximate size of a transcription-factor binding site, Nu_l is generally <<1 in multicellular species, but might be as high as 4.0 in unicellular species, particularly prokaryotes. There is rarely more than 200 bp of flanking DNA per prokaryotic gene, so the condition $NuLn/4^n < 1$ is certainly met for such species, and can even be considered likely for the bloated genomes of large multicellular species, in which the average amount of non-coding DNA per protein-coding gene can be as high as 10^6 bp (REF. 13). So, it will often be the case, especially for multicellular species, that the level of polymorphism in populations is too low for the mutational advantage of redundant regulation to be realized.

Under the assumed condition of monomorphism, the long-term equilibrium expectations for the frequencies of the three regulatory states are defined entirely by the ratio of mutation rates:

$$\hat{p}_s = \hat{p}_u = \frac{\alpha}{(1 + 2\alpha)} \quad (1a)$$

$$\hat{p}_r = \frac{1}{(1 + 2\alpha)} \quad (1b)$$

where the subscripts refer to self-reliant (s), upstream-dependent (u) and redundantly regulated (r) alleles (FIG. 2). These results show that, as the relative rate of gain of regulatory sites becomes high ($\alpha << 1$, as for multicellular species), the population will generally be fixed for the redundantly regulated allele. If $\alpha = 1$, all three alternative allelic states are equally likely, with transitions between adjacent states occurring on average every $1/u_g$ generations. However, as regulatory-site loss rates begin to predominate ($\alpha >> 1$, as for microbial species), a situation is approached in which the population spends equal amounts of time fixed for either the self-reliant or the upstream-dependent allele, with the intermediate redundant state becoming diminishingly rare. The mean transition time between these two extreme states is:

$$\bar{t} = \left(\frac{1}{u_g} + \frac{1}{u_l}\right)\sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^i \cdot i = 2\left(\frac{1}{u_g} + \frac{1}{u_l}\right) \quad (2)$$

which gives ~ $2/u_g$ generations when $\alpha >> 1$. To an order of magnitude, $u \approx 10^{-9}$ for unicellular species, 10^{-8} for invertebrates and 10^{-7} for vertebrates[13]; so, with $n \approx 8$, these results imply that neutral transitions between alternative regulatory mechanisms can occur on timescales of a few million generations or less.

We next consider the situation in which Nu_g and Nu_l are >>1. As noted above, such conditions are unlikely to be met in many species, but the results are nevertheless informative. In this case, the population is expected to reach an equilibrium level of polymorphism, as defined by the rates of mutational conversion between alternative allelic states and the weak mutational advantage (u_l) of the redundantly regulated allele. Assuming that there is no recombination between the elements involved in self-reliance and upstream regulation:

$$\hat{p}_r = \frac{\sqrt{(1+\alpha)^2 + 4\alpha} - (1+\alpha)}{2\alpha} \quad (3a)$$

$$\hat{p}_s = \hat{p}_u = \frac{(1-\hat{p}_r)}{2} \quad (3b)$$

Other than a slight elevation in the equilibrium frequency of the redundantly regulated allele at intermediate α, the expectation for a large N value is not greatly different from that for a small N value (FIG. 2). As in small populations, if the relative rate of gain of regulatory sites is high, the population is expected to be nearly monomorphic for the redundantly regulated allele but, if the relative gain rate is low, a polymorphism will exist, with the self-reliant and upstream-regulated alleles each having frequencies near 0.5.

This near population-size independence of the expected frequencies of alternative allelic states raises significant questions about the commonly held belief that redundant gene regulation evolves as a buffering mechanism against mutational inactivation. Because such an advantage is on the order of the mutation rate, it can be efficiently promoted only in large populations, but even then the functional immunity of redundantly regulated alleles to single degenerative mutations is essentially balanced by their rate of transformation to simpler allelic architectures.

The effects of recombination. In the previous analyses, the capacities for self-regulation and for control by an upstream factor were assumed to be completely linked, as can be the case in an asexual population or if the key mechanisms for regulation involve the presence or absence of closely adjacent upstream binding sites for transcription factors. However, if the recruitment of an upstream transcription factor B results not from the acquisition of novel regulatory sites by A, but from the acquisition by B of an ability to interact with pre-existing (latent) regulatory elements of A then, in a sexually reproducing species, the two modes of regulation of A need not be co-inherited. Recombination has the potential to substantially alter the evolution of the mode of gene regulation, because half of the recombinants between a gamete in which A is strictly self-reliant and another in which A is strictly upstream dependent are nulls (with no mechanism for gene activation) (FIG. 3). By contrast, all recombinants involving at least one redundantly regulated parental allele are fully functional. For this indirect recombinational advantage of a redundantly regulated allele to be influential, the population must be sufficiently large to harbour multiple allelic types. So, for genomes experiencing recombination, we can anticipate an evolutionary tendency towards different modes of gene regulation, depending on the population size and the relative rates of transcription-factor binding site gain and loss.

For the most extreme situation, that is, an effectively infinite population, the equilibrium frequencies of the three viable gametic types can be obtained analytically. Again assuming that both types of gain-of-function mutations occur at rate u_g and both types of loss-of-function mutations occur at rate u_l, the equilibrium polymorphism under free recombination is:

$$\hat{p}_s = \hat{p}_u = -(u_g + 3u_l) + \sqrt{(u_g + 3u_l)^2 + 2u_l} \simeq \sqrt{2u_l} \quad (4a)$$

$$\hat{p}_r \simeq 1 - 2\hat{p}_s \quad (4b)$$

Under this setting, in dramatic contrast to the case of no recombination, nearly all individuals are expected to carry redundantly regulated alleles. Recalling that $n \approx 8$ and $u_l = nu$, the total expected frequency of non-redundant alleles is $\approx 8\sqrt{u}$, which will generally be <0.001. Although equations 4a,b give the upper bound

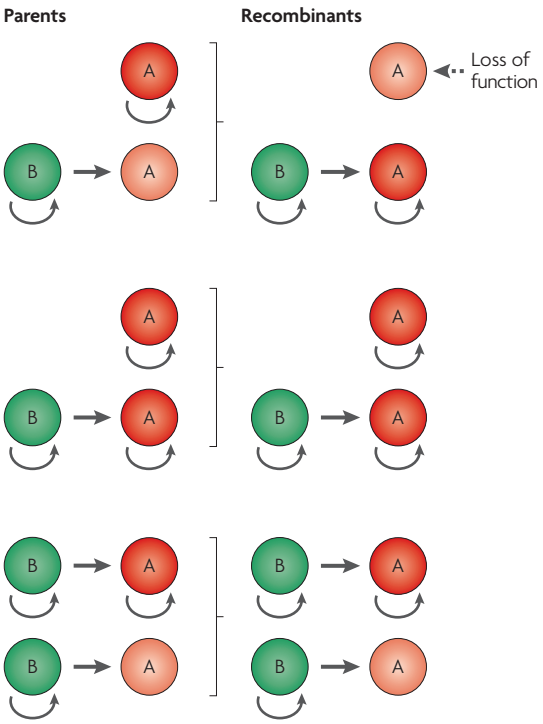

Figure 3 | **A consideration of the pairs of recombinant progeny for all pairs of alternative parental regulatory states reveals the recombinational advantage of redundantly regulated alleles.** A non-functional allele is produced only when recombination occurs between alternative states involving single regulatory modes, although this case also produces a redundantly regulated recombinant.

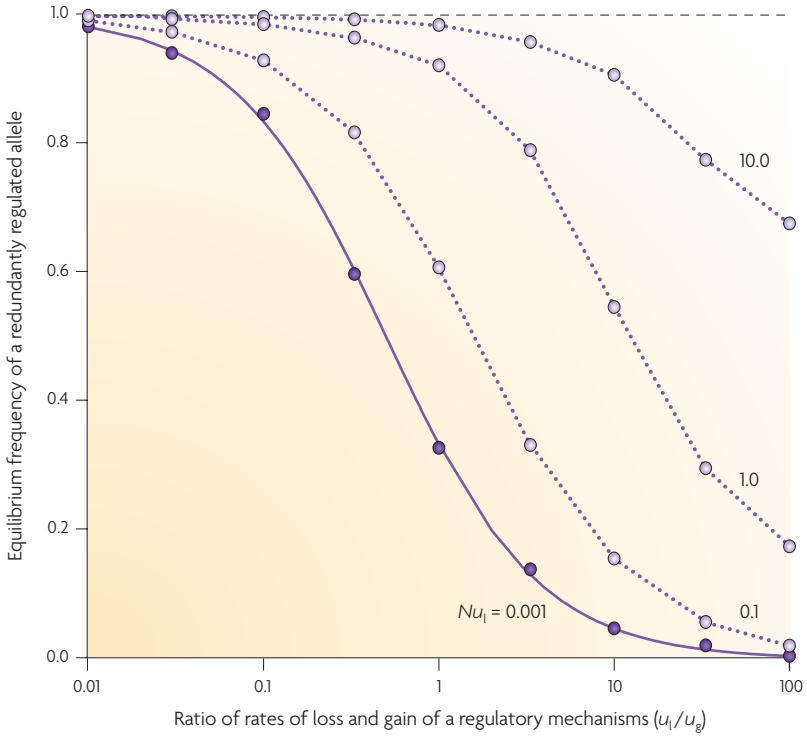

Figure 4 | **Recombination encourages the evolution of redundantly regulated alleles.** The expected frequencies of redundantly regulated alleles are shown, under the assumption of free recombination between the DNA-level elements that are responsible for self-regulation and those responsible for upstream regulation. The solid line denotes the theoretical expectation for small populations (equation 1b), whereas the dashed line denotes the expectation for effectively infinite populations (equation 5b). Data points are derived from stochastic simulations with $u_l = 10^{-6}$ and $N = 10^3$ (solid circles), and $N = 10^5$, $N = 10^6$ and $N = 10^7$ (open circles). The inset numbers denote Nu_l.

to the effect of recombinational inactivation, a significant influence on the abundance of alternative pathway configurations is still expected if $Nu_l = Nnu > 0.01$ (FIG. 4). Recalling the estimates of Nu given above, this condition is fulfilled in most species.

For intermediate rates of recombination ($0 < r < 0.5$, which here for simplicity are assumed to be homogeneous over sites), the equilibrium regulatory-state frequencies for effectively infinite populations are closely approximated by the solution of:

$$\hat{p}_s^3 r + \hat{p}_s^2 (2u_l + ru_g - r) - \hat{p}_s(3u_l + u_g) + u_l = 0 \qquad (5)$$

The form of this equation shows that the influence of recombination largely depends on whether r is small or large relative to u_l. If $r/u_l < 0.1$, the equilibrium frequencies at high N are close to those expected under complete linkage, as in Equations 3a,b, whereas, if $r/u_l > 10$, then:

$$\hat{p}_s = \hat{p}_u \simeq \sqrt{\frac{u_l}{r}} \qquad (6a)$$

$$\hat{p}_r \simeq 1 - 2\hat{p}_s \qquad (6b)$$

Some insight into the power of recombination with respect to factors that are linked on the same chromosome can be acquired by noting that $r \approx 0.5(1-e^{2dc})$, where d is the number of nucleotides separating two locations and c is the recombination frequency per base pair. For small distances, $r \approx dc$, and the ratio r/u_l is equivalent to $(c/u)(d/n)$. Letting $n \approx 10$, this implies that recombinational inactivation can be effective if $c/u > 100/d$, but it will be inconsequential if $c/u < 1/d$. These expressions are useful because estimates of c/u, the ratio of the rates of recombination and mutation per base pair, are available for numerous species. In eukaryotes, c/u scales inversely with genome size, being on the order ≈2500, ≈20 and ≈1 for genome sizes of 10 Mb (a typical unicellular eukaryote), 100 Mb (an invertebrate), and 1,000 Mb (a vertebrate or land plant), respectively[13]. Although incapable of meiotic recombination, prokaryotes can exchange DNA in other ways, and they have an average c/u of ~3 (REF. 13). Thus, even in the most non-recombinogenic backgrounds known ($c/u \approx 1$), recombinational inactivation will promote redundantly regulated alleles, provided that the regulatory sites are at least 100 bp apart and the effective population size is sufficiently large.

In summary, the preceding results suggest that there are at least two dramatically different pathways by which regulatory redundancy can be promoted by non-adaptive processes. Owing to the inefficiency of selection at eradicating insertions[13], species with small effective population sizes accumulate substantial amounts of excess DNA, providing adjacent genes with a natural susceptibility to the mutational origin of redundant regulatory structures. Such species are expected to evolve high levels of regulatory redundancy for purely physical reasons, regardless of any inherent selective advantage of redundant transcription-factor binding sites. By contrast, although the genomes of most microbial species harbour little non-coding DNA, and therefore provide few targets for the mutational origin of new regulatory sites, the effective population sizes of such species are often large enough that the physical force of recombination can promote the abundance of redundantly regulated alleles; again, this is not because of any inherent advantage of such alleles from a metabolic or developmental standpoint, but because of the increased susceptibility of their simpler counterparts to recombinational inactivation.

Higher-order systems

The preceding examples will be oversimplifications for cellular contexts in which the recruitment of multiple alternative transcription factors is possible. However, even the addition of one more potential regulatory factor can greatly complicate the array of possible outcomes, as the two transcription factors might or might not regulate each other's expression, and one factor might lose self-regulatory capacity. In addition, the downstream target gene A might be redundantly regulated by one or both transcription factors, be dependent on one of them, or be entirely self-reliant. Even if the ordering of the two upstream transcription factors is ignored, 31 alternative network configurations ensure the activation of all three genes (assumed to be essential for viability). Systems

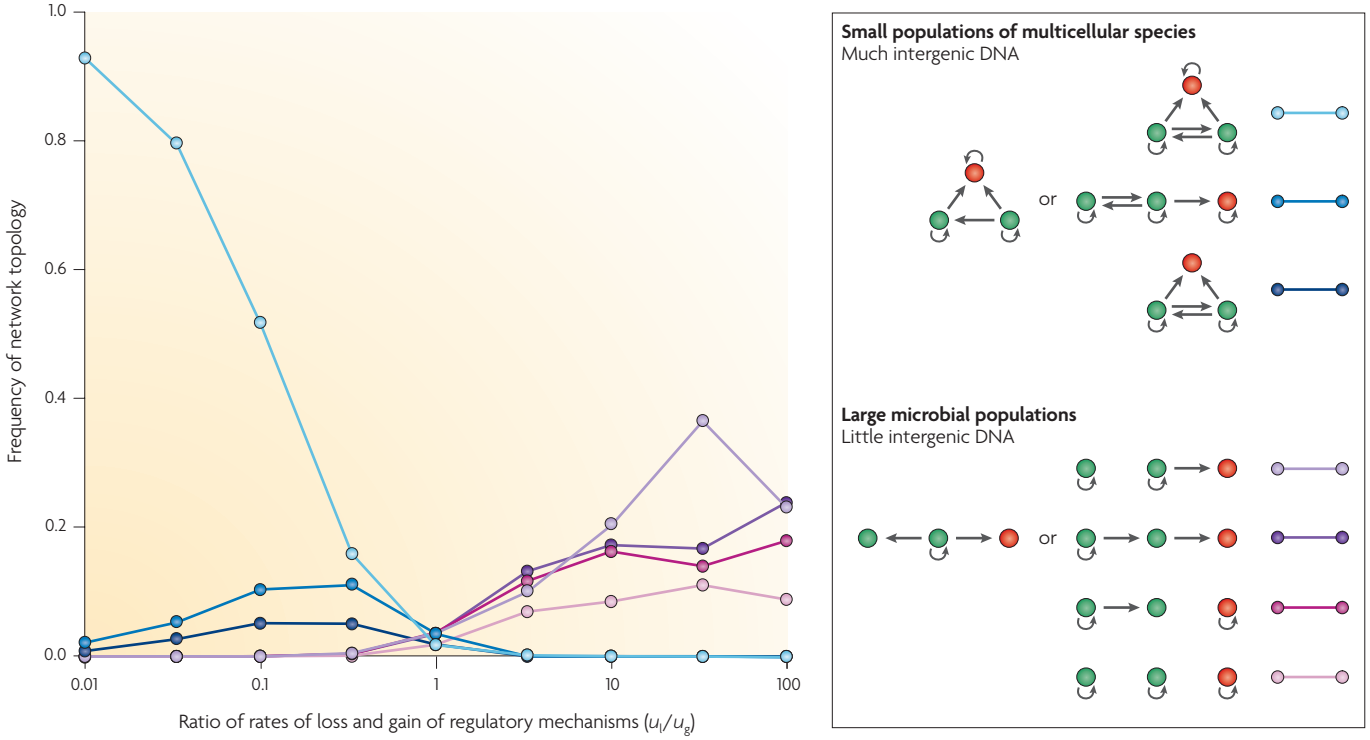

Figure 5 | **Average frequencies of several alternative three-gene network configurations under the assumptions of complete linkage of the DNA-level elements involved in the activation of gene A.** Results were obtained by stochastic simulations of 10^9 generations, with $N = 100$ and u_l and u_g in the range of 10^{-7} to 10^{-5}. On the right hand side of the figure, transcription factors are denoted by green circles, and the downstream gene by a red circle.

with this level of complexity do not yield simple analytical solutions, although the expected frequencies of alternative configurations are readily obtained by computer simulations of the joint processes of mutation and random genetic drift. Such analyses reveal behaviour with similarities to that noted for two-factor systems.

Considering first the situation in which all three genes are assumed to be completely linked: regardless of the population size, when the rate of gain of regulatory elements exceeds that of loss, the system evolves towards a high level of interconnectivity and redundancy (FIG. 5); when $u_l/u_g \approx 1$, all possible configurations have nearly equal probabilities; and when $u_l/u_g > 1$, the system becomes dominated by linear pathways and/or self-reliant components. Here it should be noted that particular linear pathways are equally likely to have either of the two potential transcription factors in the central position. For example, pathways with a B→C linkage are readily replaced by C→B, after passing through the intermediate state B↔C. Alternatively, as with two-factor systems, when population sizes are large, recombination alters the equilibrium network frequencies through the production of recombinant progeny that lack function in one or more genes. This leads to populations that are increasingly dominated by genotypes with high levels of redundant regulation (conferring immunity to recombinational breakdown) when r is high (FIG. 6). Interconnected triads like those noted in FIG. 6, known as feed-forward loops, are common in *Escherichia coli*[46].

Although systems containing still larger numbers of potentially interacting genes are not uncommon, owing to the rapid increase in the number of potential pathways with the dimensionality of the system, such extensions are computationally challenging, particularly when recombination is allowed for. Nonetheless, for systems within the domain of low u_l/u_g, for which mutation pressure is the primary governing force, as seems to be the case for multicellular eukaryotes, the complete probability distribution of all alternative pathways can be approximated by using a matrix description for pathway architecture and stochastically introducing regulatory-element gains and losses (Supplementary information S1 (box)). Such extensions to larger numbers (k) of potential transcription factors confirm all of the previously discussed results, while also revealing patterns that are relevant to the overall levels of node connectivity (FIG. 7).

First, the variance in levels of connectivity exhibited by alternative alleles can be quite substantial, especially when $u_l/u_g \approx 1$, enough so that in many cases the most common regulatory type is represented less than 50% of the time. Such high levels of regulatory variation, which are consistent with the observations cited above, raise obvious questions regarding the adaptive basis of pathway architectures that are associated with any particular species. Second, distributions of connectivity levels can take on dramatically different shapes depending on the relative levels of loss and gain of transcription-factor

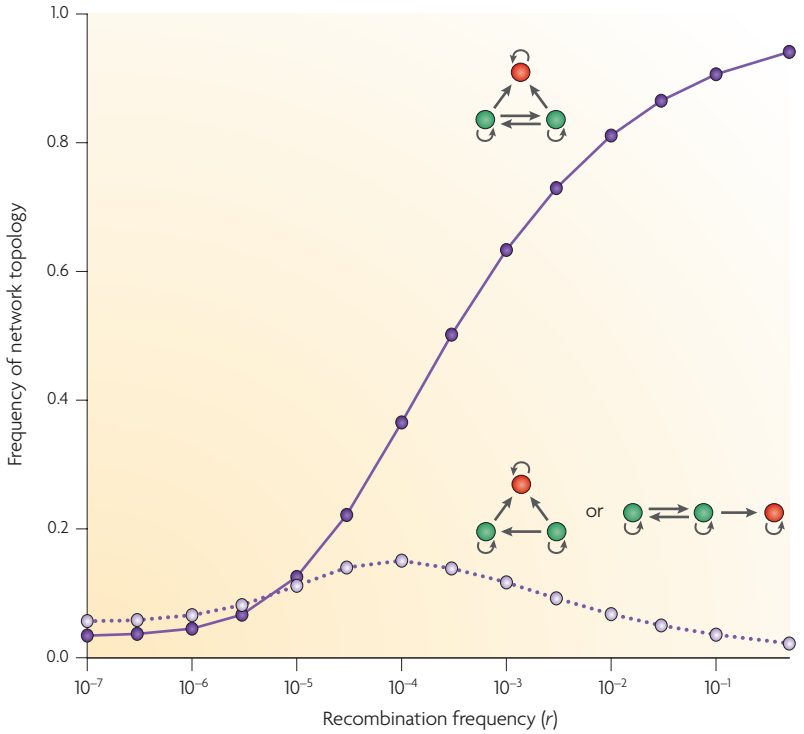

Figure 6 | **Average frequencies of the three most common three-gene network configurations in large, recombining populations.** Results were obtained by stochastic simulations of >10^9 generations, with $N = 10^9$ and $u_l = u_g = 10^{-6}$. For these mutational properties, as $r \to 0$, all possible network configurations converge on nearly identical frequencies. Transcription factors are denoted by green circles, and the downstream gene by red circles.

binding sites: such distributions are nearly symmetrical when $u_l = u_g$, decline with increasing connectivity when $u_l > u_g$ and increase with increasing connectivity when $u_l < u_g$.

Because most microbial species seem to be in the domain in which $u_l > u_g$, it might be tempting to draw a parallel between the distributions in the lower part of FIG. 7 with the power-law distributions that have been inferred for levels of connectivity in such species (for protein-interaction networks[17,47] and transcriptional regulatory networks[16,48]). However, although these theoretical results do show that commonly observed emergent features of regulatory networks can be generated entirely by non-adaptive evolutionary forces, a conclusion that has previously been reached in numerous independent studies (for example, REFS 13,24,49), patterns of within-gene temporal variation in regulatory connections (the subject of this study) need not always reflect the standing patterns for entire collections of genes at single points in time (the subject of all previous network surveys), which will also depend on among-gene variation in k and α. Unless something is known of the last-mentioned features, the attachment of evolutionary interpretations to observed patterns of node connectivity seems to be problematical.

Taken together, the preceding results again predict not only substantial dependence of average pathway architectures on the population-genetic environment, but also considerable variation around the expected patterns within any particular context. In addition, it can be predicted that, even with constant pathway architectures, genes will frequently experience temporal turnover in the identities of the component members of their regulatory repertoires. These expectations, which are entirely consistent with the growing list of examples of 'developmental systems drift' noted above, can be further evaluated with observations on allelic variation in regulatory mechanisms in natural populations. Interestingly, the first analysis of this type, involving a highly complicated regulatory network that had previously been presented as a fixed pattern in an echinoderm[50], revealed substantial variation for key regulatory sites in natural populations[51]. A related study revealed substantial interspecific divergence in both *cis*-regulatory sites and *trans*-acting transcription factors[52]. Such observations raise questions about the utility and feasibility of precise typological descriptions of specific regulatory pathways, motivating concern that "…developmental evolution is a series of case studies with few overarching laws."[53]

One issue that has not been formally resolved in previous analyses concerns the extent to which an upper limit to pathway complexity might eventually result from either the magnified mutational vulnerability of highly interdependent systems or the simple attainment of a birth–death equilibrium for regulatory sites. Some evidence that this might be the case can be seen in FIG. 7, in which it is apparent that the movement of the upper ends of the distribution of connectivity tends to diminish with increasing numbers of potential regulatory factors (k). In an analysis that has some resemblance to that used here, but that imposes a cost to component addition (a weak form of selection), Soyer and Bonhoeffer[54] found that, without any direct selection for pathway architecture, pathways eventually reach an equilibrium average size, although they also tend to grow to sizes beyond the minimal requirement (as found here).

Finally, it has been observed that numbers of transcription-factor genes scale disproportionately with genome size in both prokaryotes and eukaryotes (in comparison with other gene classes)[55]. Although an increase in the size of a species' transcription-factor repertoire might reflect the action of selection to promote increased numbers of alternative regulatory states of cells, it is unclear why such adaptive promotion should be increasingly effective with increasing genome size. One possibility that merits further exploration is that the types of processes outlined above have a snowball effect, as individual transcription factors are passively recruited into networks and at least transiently preserved by the embedding process, particularly in genomes that have elevated numbers of targets for the origin of novel regulatory sites.

Concluding remarks

Contrary to widespread belief, there is no compelling empirical or theoretical evidence that complexity, modularity, redundancy or other features of genetic pathways are promoted by natural selection. For example, despite

Power-law distribution
A distribution for a variable x that follows the form ax^b.

the continued fascination with the putative adaptive basis of redundancy, this is not the first time that the difficulties of maintaining redundancy by natural selection have been pointed out. It is well known that the advantage of masking mutations with a backup duplicate gene is approximately equal to the per-locus mutation rate, which is effectively offset by the rate at which the duplicate itself is inactivated[56–58]. Certainly, there is substantial evidence that duplicate genes can mask the effects of mutations at paralogous loci[13,21]; however, contrary to what is expected for genes that are maintained by selection, paralogous genes in all species undergo considerable turnover, and their presence is more readily explained by a steady-state birth–death process[59,60], analogous to that presented above for regulatory elements. Such observations raise significant questions about whether distributed robustness (the tendency for single-gene knockouts to have less-than-lethal effects as a consequence of excess network connections) evolves by natural selection[21,61], as opposed to being an unavoidable by-product of the purely physical processes of mutation and duplication.

Although the models presented here are simplistic in many ways, ignoring, for example, the details of molecular kinetics and the likelihood that small networks will often be imbedded in a larger set of interactions, they are far more biologically realistic than previous attempts to model network evolution as a series of saltational (instantaneous and stepwise) events (for example, REFS 2,3,62–66). Invariably, such studies simulate the construction of network topologies by pre-specified rules for stochastic node attachment, gene duplication and/or transcription-factor binding site origin, with no attention being given to the intermediate states that necessarily exist in the context of actual populations. Relying on this approach, deviations of observed patterns of network connectivity from those that are generated with randomly assembled networks have often been invoked as evidence for the guiding hand of natural selection in the evolution of network architecture[2,3,67,68]. However, the preceding results clearly demonstrate that models that are devoid of population-genetic considerations can be misleading with respect to network evolution. Fixations of new mutations are not instantaneous, and null models that are generated by randomization procedures or simulations of saltational change are by no means equivalent to neutral models of evolution. As amply described above, the plausible distributions of outcomes under neutral evolution are defined by the prevailing population-genetic background, a case in point being the driving force of recombinational inactivation, which can be revealed only in a framework that allows for population-level polymorphism.

Although the preceding arguments demonstrate numerous mechanisms for the passive emergence of complex gene-level transactions, this does not necessarily rule out the promotion of specific gene-interaction configurations in very large populations by natural selection. However, it should be kept in mind that the conditions for the adaptive promotion of changes in

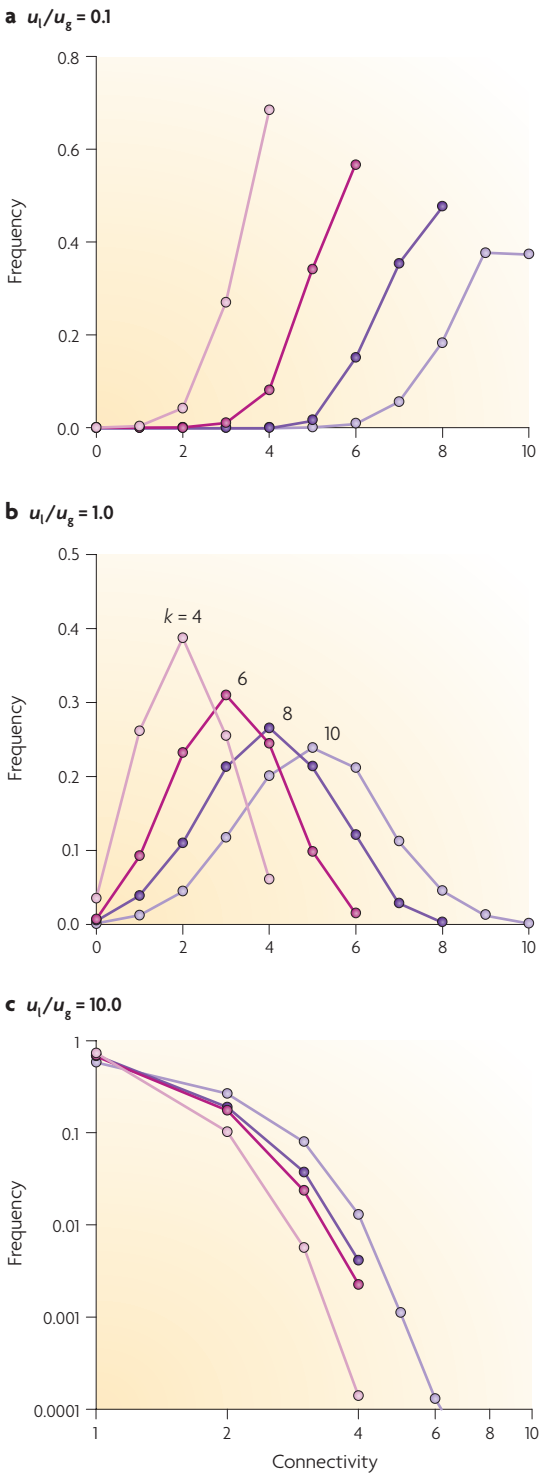

Figure 7 | **Expected null distributions of node density (number of transcription factors that actually interact with downstream gene A).** Results are given for situations in which there are k = 4, 6, 8 or 10 potential transcription factors, with no direct selection for the form of the network pathway except for the requirement that all genes remain active. Patterns are shown for three ratios of rates of regulatory-site loss and gain, u_l/u_g. Note that for u_l/u_g = 10, the results are plotted logarithmically to demonstrate the approximation of the distribution to a power-law expectation.

gene-regulatory structure are more stringent than for coding-region mutations, a point that is yet to enter the current debate about whether adaptive evolution is more likely to involve changes in gene structure or gene regulation[69,70]. Whereas a simple amino-acid substitution does not alter the target size of an allele for degenerative mutations, the addition of a regulatory embellishment will generally impose an increased mutational burden on a gene. Consider a modification of a gene's regulatory structure that confers a selective advantage equal to s. The net long-term advantage is then $s–u_d$, where u_d is the mutational burden of the modified allele (the excess mutation rate to defective alleles owing to the larger mutational target; equal to zero in the case of a nucleotide substitution). If the mutational disadvantage is sufficiently large that $(s–u_d) < 0$, fixation will be opposed regardless of s. If $(s–u_d)$ is positive, because $1/N$ defines the magnitude of stochastic fluctuations in allele frequencies resulting from drift, a reasonable probability of fixation by natural selection requires that the net selective advantage exceeds the power of random genetic drift[71], that is, $(s–u_d) > 1/N$. Thus, if a regulatory site based on a 10-bp motif is to be maintained by selection, the advantage must exceed $10u+(1/N)$.

In summary, a mechanistic understanding of network evolution is unlikely to be achieved without specific reference to the ubiquitous non-adaptive forces of evolution that operate at the levels of DNA (mutation and recombination) and populations (random genetic drift). A growing body of evidence supports the idea that many aspects of complexity at the genomic, molecular and cellular levels in multicellular species are likely to owe their origins to these non-adaptive forces, representing little more than passive outcomes of the unique population-genetic environments that are presented by such lineages[13]. Genomic features that are consistent with such evolution include the preservation of duplicate genes, the evolution of modular gene architecture, the establishment of introns and untranslated regions and, now, the evolution of network architecture and redundant regulation. Although demonstrating the feasibility of such neutral evolution is not equivalent to verifying it as the only determinant of genomic architecture, the common view that natural selection alone is responsible for the evolution of organismal complexity seems to be not only unnecessary, but misleading. The failure to recognize this fact breeds a false sense of security in our understanding of evolutionary processes.

1. Gerhart, J. & Kirschner, M. *Cells, Embryos and Evolution* (Blackwell Science, Malden, 1997).
2. Barab'asi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
3. Babu, M. M., Teichmann, S. A. & Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *Mol. Biol.* **358**, 614–633 (2006).
4. Balaji, S., Iyer, L. M., Aravind, L. & Babu, M. M. Uncovering a hidden distributed architecture behind scale-free transcriptional regulatory networks. *J. Mol. Biol.* **360**, 204–212 (2006).
5. Davidson, E. H. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, New York, 2006).
6. Alon, U. Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867 (2003).
7. Adami, C. Digital genetics: unravelling the genetic basis of evolution. *Nature Rev. Genet.* **7**, 109–118 (2006).
8. Wagner, A. Does selection mold molecular networks? *Science STKE* **202**, pe41 (2003).
One of the first papers to raise questions about the adaptive paradigm for the architecture of genetic networks.
9. Sole, R. V. & Valverde, S. Are network motifs the spandrels of cellular complexity? *Trends Ecol. Evol.* **21**, 419–422 (2006).
10. Keller, E. F. Revisiting 'scale-free' networks. *Bioessays* **27**, 1060–1068 (2005).
11. Lynch, M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl Acad. Sci. USA* **104**, S8597–S8604 (2007).
This paper raises questions about the rationale and objectivity of numerous arguments that nearly all aspects of molecular, cellular and developmental complexity have arisen by adaptive mechanisms.
12. Wilkins, A. S. *The Evolution of Developmental Pathways* (Sinauer, Sunderland, 2002).
An excellent overview of our knowledge (or lack thereof) of the evolutionary forces that mould developmental pathways.
13. Lynch, M. *The Origins of Genome Architecture* (Sinauer, Sunderland, 2007).
14. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).
15. Thieffry, D., Huerta, A. M., Perez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20**, 433–440 (1998).
16. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
17. Wuchty, S. & Almaas, E. Evolutionary cores of domain co-occurrence networks. *BMC Evol. Biol.* **5**, 24 (2005).
18. Johnson, N. A. & Porter, A. H. Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J. Theor. Biol.* **205**, 527–542 (2000).
19. Force, A. *et al.* The origin of subfunctions and modular gene regulation. *Genetics* **170**, 433–446 (2005).
20. Haag, E. S. & Molla, M. N. Compensatory evolution of interacting gene products through multifunctional intermediates. *Evolution* **59**, 1620–1632 (2005).
21. Wagner, A. *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, Princeton, 2005).
A broad and relatively balanced view of the evolutionary mechanisms that can lead to the robustness of living systems to external and internal perturbations.
22. Conant, G. C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genet.* **34**, 264–266 (2003).
23. Artzy-Randrup, Y., Fleishman, S. J., Ben-Tal, N. & Stone, L. Comment on 'Network motifs: simple building blocks of complex networks' and 'Superfamilies of evolved and designed networks'. *Science* **305**, 1107 (2004).
24. van Noort, V., Snel, B. & Huynen, M. A. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Rep.* **5**, 280–284 (2004).
25. de Silva, E. & Stumpf, M. P. Complex networks and simple models in biology. *J. R. Soc. Interface* **2**, 419–430 (2005).
26. Wilkins, A. S. Recasting developmental evolution in terms of genetic pathway and network evolution and the implications for comparative biology. *Brain Res. Bull.* **66**, 495–509 (2005).
27. Sommer, R. J. Evolution and development — the nematode vulva as a case study. *Bioessays* **19**, 225–231 (1997).
28. Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
29. Ruvinsky, I. & Ruvkun, G. Functional tests of enhancer conservation between distantly related species. *Development* **130**, 5133–5142 (2003).
30. Coulson, R. M. & Ouzounis, C. A. The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res.* **31**, 653–660 (2003).
31. Goltsev, Y., Hsiong, W., Lanzaro, G. & Levine, M. Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. *Dev. Biol.* **275**, 435–446 (2004).
32. Coulson, R. M., Touboul, N. & Ouzounis, C. A. Lineage-specific partitions in archaeal transcription. *Archaea* **2**, 117–125 (2006).
33. Hill, R. C. *et al.* Genetic flexibility in the convergent evolution of hermaphroditism in *Caenorhabditis* nematodes. *Dev. Cell* **10**, 531–538 (2006).
34. Mazurie, A., Bottani, S. & Vergassola, M. An evolutionary and functional assessment of regulatory network motifs. *Genome Biol.* **6**, R35 (2005).
35. Moses, A. M. *et al.* Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput.Biol.* **2**, 1219–1231 (2006).
36. Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006).
An elegant demonstration of how dramatic changes in regulatory mechanisms can be brought about by intermediate, neutral steps involving functional redundancy.
37. Marino-Ramirez, L., Jordan, I. K. & Landsman, D. Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol.* **7**, R122 (2006).
38. Tanay, A., Regev, A. & Shamir, R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl Acad. Sci. USA* **102**, 7203–7208 (2005).
39. Lozada-Chavez, I., Janga, S. C. & Collado-Vides, J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.* **34**, 3434–3445 (2006).
40. Perez-Rueda, E., Collado-Vides, J. & Segovia, L. Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. *Comput. Biol. Chem.* **28**, 341–350 (2004).
41. Riechmann, J. L. *et al. Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
42. Messina, D. N., Glasscock, J., Gish, W. & Lovett, M. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res.* **14**, 2041–2047 (2004).
43. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
44. Stone, J. R. & Wray, G. A. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol. Biol. Evol.* **18**, 1764–1770 (2001).

45. Hahn, M. W., Stajich, J. E. & Wray, G. A. The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol.* **20**, 901–906 (2003).
This work shows that even non-functional DNA can be under selection for the avoidance of spurious regulatory sites
46. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
47. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
48. Rodriguez-Caso, C., Medina, M. A. & Sole, R. V. Topology, tinkering and evolution of the human transcription factor network. *FEBS J.* **272**, 6423–6434 (2005).
49. Chung, F., Lu, L., Dewey, T. G. & Galas, D. J. Duplication models for biological networks. *J. Comput. Biol.* **10**, 677–687 (2003).
50. Yuh, C. H., Bolouri, H. & Davidson, E. H. Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896–1902 (1998).
51. Balhoff, J. P. & Wray, G. A. Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proc. Natl Acad. Sci. USA* **102**, 8591–8596 (2005).
52. Romano, L. A. & Wray, G. A. Conservation of *Endo16* expression in sea urchins despite evolutionary divergence in both *cis* and *trans*-acting components of transcriptional regulation. *Development* **130**, 4187–4199.
References 51 and 52 provide dramatic evidence that typological descriptions of the regulatory structure of genes on the basis of narrow model systems ignore important aspects of variation that are found within and among natural populations.
53. True, J. R. & Haag, E. S. Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* **3**, 109–119 (2001).
A thoughtful account of how major modifications of developmental processes can come about by neutral processes.
54. Soyer, O. S. & Bonhoeffer, S. Evolution of complexity in signaling pathways. *Proc. Natl Acad. Sci. USA* **103**, 16337–16342 (2006).
55. van Nimwegen, E. Scaling laws in the functional content of genomes. *Trends Genet.* **19**, 479–484 (2003).
56. Clark, A. G. Invasion and maintenance of a gene duplication. *Proc. Natl Acad. Sci. USA* **91**, 2950–2954 (1994).
57. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of fixation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
58. Proulx, S. R. & Phillips, P. C. The opportunity for canalization and the evolution of genetic networks. *Am. Nat.* **165**, 147–162 (2005).
References 56–58 formally demonstrate the difficulties of evolving genetic redundancy by natural selection.
59. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1154 (2000).
60. Wagner, A. The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* **154**, 1389–1401 (2000).
61. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361 (2000).
62. Bhan, A., Galas, D. J. & Dewey, T. G. A duplication growth model of gene expression networks. *Bioinformatics* **18**, 1486–1493 (2002).
63. Cordero, O. X. & Hogeweg, P. Feed-forward loop circuits as a side effect of genome evolution. *Mol. Biol. Evol.* **23**, 1931–1936 (2006).
64. Foster, D. V., Kauffman, S. A., & Socolar, J. E. Network growth models and genetic regulatory networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73**, 031912 (2006).
65. Ciliberti, S., Martin, O. C. & Wagner, A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput. Biol.* **3**, e15 (2007).
66. Pagel, M., Meade, A. & Scott, D. Assembly rules for protein networks derived from phylogenetic-statistical analysis of whole genomes. *BMC Evol. Biol.* **7**, S16 (2007).
67. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
68. Yeger-Lotem, E. *et al.* Network motifs in integrated cellular networks of transcription-regulation and protein–protein interaction. *Proc. Natl Acad. Sci. USA* **101**, 5934–5939 (2004).
69. Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3**, e245 (2005).
The author argues that most interesting aspects of evolution are associated with changes at the level of gene regulation, rather than with changes in coding DNA.
70. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo–devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
This paper presents a counterview to reference 69.
71. Crow, J. F. & Kimura, M. *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).

Acknowledgements
I am very grateful to E. Haag, M. Hahn and three anonymous reviewers for helpful comments. This work has been supported by US National Science Foundation and US National Institutes of Healthe grants to the author.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Michael Lynch's homepage: http://www.bio.indiana.edu/facultyresearch/faculty/Lynch.html

SUPPLEMENTARY INFORMATION
See online article: S1 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF.