

Uncovering a Hidden Distributed Architecture Behind Scale-free Transcriptional Regulatory Networks

S. Balaji, Lakshminarayan M. Iyer, L. Aravind* and M. Madan Babu*

National Center for
Biotechnology Information
National Library of Medicine
National Institutes of Health
Bethesda, MD 20894, USA

Numerous studies in both prokaryotes and eukaryotes have shown that, under standard growth conditions, less than 20% of the protein-coding genes are essential for survival. This suggests that biological systems have evolved to have a high degree of robustness to mutational disruptions that can affect the majority of their genes. This mutational robustness could arise either due to redundancy, i.e. direct backup, or due to distributed architecture, i.e. indirect backup where multiple genes contribute to the functioning of a process in the system. Despite clear evidence for direct backup, the prevalence of indirect backup is poorly understood. In this study, we reveal the existence of a hidden distributed architecture behind the scale-free transcriptional regulatory network of yeast by applying a unique network transformation procedure and show that the network is tolerant even to mutations that disrupt regulatory hubs. Contrary to what is generally accepted, our observation that hubs can be lost or replaced in evolution suggests that this hidden distributed architecture behind scale-free networks protects the overall transcriptional program of the organism from mutations affecting major regulatory hubs. We show that the distributed architecture has been provided by an unexpectedly large number of coordinating partners for any regulatory protein. On the basis of these findings, we propose that the existence of such architecture can allow organisms to explore the adaptive landscape in changing environments by providing the plasticity required to reprogram levels of expression of specific genes that may enhance survival. Thus, an “over-engineered” backup system in the form of distributed architecture is likely to be a major determinant of the “evolvability” of the gene expression in organisms faced with environmental diversity.

Published by Elsevier Ltd.

*Corresponding authors

Keywords: transcriptional network; co-regulation network; distributed robustness; scale-free structure; evolution

Introduction

Several studies in both eukaryotes^{1–5} and prokaryotes^{6–8} suggest that, under standard growth conditions, more than 80% of the protein-coding genes are not essential for their survival. While a subset of the non-essential genes may be required for specialized conditions, various studies have demonstrated that a major fraction of the non-essential genes may be inherently backed up in cellular systems.⁹ Moreover, in both prokaryotes and eukaryotes, which are at a similar level of organizational complexity, there can be large differences in the

number of genes encoded,^{10,11} implying that biological systems have a relatively small core of absolutely essential genes, and a much larger population of genes that might be individually dispensable. This suggests that biological systems might have evolved to have a high degree of robustness to mutations that disrupt gene function.

Earlier studies on mutational robustness have argued that this phenomenon arises either due to redundancy or distributed robustness.⁹ In the former case, a system may be unaffected upon removal of a gene as a duplicate or a functionally equivalent gene can compensate directly for the affected gene. In the latter case, different genes with distinct roles work together to form a functional system; hence, disruption of a gene in such a system is compensated indirectly by unrelated genes

E-mail addresses of the corresponding authors:
madanm@ncbi.nlm.nih.gov; aravind@ncbi.nlm.nih.gov

participating in the functioning of the system. The latter case differs from the former due to the indirect nature of the backup arising as a result of cooperative interactions among the genes in the system. The first phenomenon, redundancy, is very tangible and has been measured in terms of genes with overlapping functions, i.e. gene duplicates (paralogs) and analogous proteins.^{12–17} However, the role of distributed robustness remains poorly understood.⁹ Earlier studies on mutational robustness have concentrated primarily on individual genes, or on a set of genes that act in the same functional pathway.^{9,18–21} Even though further studies have addressed the role of redundancy^{9,12,22,23} in various molecular interaction networks, including protein regulatory networks, very few studies have investigated the collective properties of real biological networks, which are required to evaluate the significance of concepts like distributed robustness. In this context, the technological advances in high-throughput genomics and proteomics place us in a unique position to reconstruct the interaction network of most genes, and to study them as an integrated system rather than in isolation or as a small group of interacting proteins.^{24–27}

Production of proteins in a cell is regulated primarily at the level of transcription, where a subset of regulatory proteins, the DNA-binding transcription factors, bind to specific sequences on the DNA associated with a gene and either activate or repress its expression. In this work, we compiled the most comprehensive regulatory network of transcription factors and their target genes in the model yeast *Saccharomyces cerevisiae* from various sources. In this transcriptional network, also re-

ferred to as the protein regulatory network, nodes denote either transcription factors or target genes and the links represent direct transcriptional regulatory interactions.²⁸ The network, which represents the blueprint for the control of gene expression in yeast, consisted of 12,873 regulatory interactions between 157 DNA-binding transcription factors and 4410 target genes, and is one of the largest assembled regulatory networks for any organism.

We investigated the inherent structural properties of this transcriptional network to uncover any possible role for distributed robustness arising from cooperative functional interactions between transcription factors. We show that a specific network transformation procedure provides convincing evidence for the presence of an inherent distributed robustness in the transcriptional regulatory network hidden behind its scale-free structure.

Results and Discussion

Paradoxical observations on the yeast transcriptional regulatory network

Very few regulatory hubs are essential for viability

It is now known that networks with different topology have very different properties, and their structure provides insights into their robustness, i.e. their ability to function in the face of disruption of their nodes.²⁹ Previous theoretical studies have shown that networks with a power-law-like degree distribution (Figure 1(a)) are tolerant to random removal of nodes, but are particularly vulnerable to targeted removal of hubs.²⁹ On the contrary,

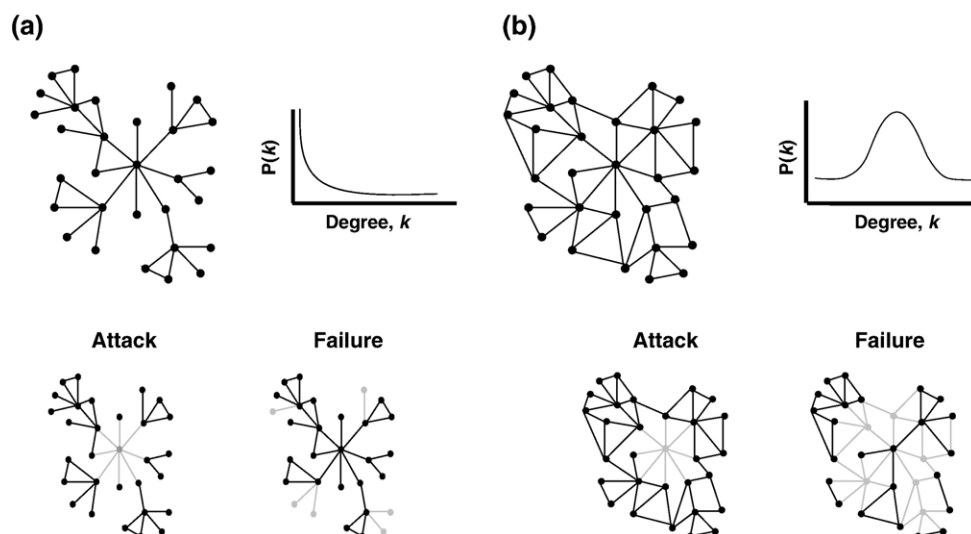


Figure 1. A diagram of networks with different topologies. (a) A network with a decentralized architecture shows a power-law behavior (inhomogeneous distribution) in the degree distribution. Networks with such behavior are robust to random removal of nodes, but are vulnerable to targeted attacks of hubs. (b) A network with a distributed architecture shows a peaked behavior (homogeneous distribution) in the degree distribution. Such networks are tolerant to removal of highly connected proteins, but are sensitive to random removal of a large fraction of nodes.

networks that are homogeneously wired (Figure 1 (b)) are sensitive to random removal of a large fraction of nodes, but are robust to targeted attacks of the most highly connected nodes. This feature is attributed to their distributed architecture.²⁹

The connectivity distribution of the compiled transcriptional regulatory network was best approximated by a power-law equation, suggesting that it resembles a scale-free network (Figure 3(a)). This suggested that they might withstand disruption of transcription factors that regulate a few genes, but could be vulnerable to disruption of regulatory hubs. It should be noted, however, that not all networks with power-law degree distribution are vulnerable to removal of hubs, as was shown recently by Doyle *et al.*³⁰ and Li *et al.*^{31†}. Hence, we assessed this possibility through direct computational simulations (see below). We defined the hubs in this network as the transcription factors that had 150 or more target genes, as this constituted the top 20% of the transcription factors with a high number of interactions in the network. It is interesting to note that these hubs account for more than half the regulatory interactions in the network. To test if the organism is really vulnerable to disruptions of regulatory hubs, we checked if transcription factors that are regulatory hubs are essential for survival. By integrating gene essentiality data^{3,5} with the transcriptional network, we observed that only nine of the 157 DNA-binding transcription factors are essential for survival. Surprisingly, we found that only five of the 33 regulatory hubs are essential. These include the heat-shock response transcription factor Hsf1p, and the autonomous replication site-binding factor, Abf1p. In fact, 28 of the 33 global regulators, such as Cin5p and Tos8p that regulate 251 and 248 target genes, are not essential for the survival of the organism. Our results are consistent with studies conducted recently on much smaller networks of yeast.^{31,32} This observation implied that organisms can tolerate mutations that disrupt many of their regulatory hubs in the transcriptional network under standard laboratory growth conditions. To assess the essentiality of hubs in different conditions, we used the dataset made available by Dunn *et al.* in the yeast genome consortium website‡. In their study, the authors monitored the viability of deletion mutants under five different conditions in this study. This showed that almost no regulatory hub, except Aft2p (a regulator of genes involved in iron homeostasis), was essential under all conditions, suggesting that even the essential hubs are largely condition-specific.³³ Importantly, we found that there is no enrichment for hubs to be more essential than non-hubs, even under specific conditions. For instance 12 of the 33 regulatory hubs (~1/3) were essential in at least one of the five

conditions; this fraction is comparable to our observation that 37 of the 124 non-hub transcription factors (~1/3 of all non-hub transcription factors) are essential under at least one of the five conditions.

To understand what really happens when hubs are disrupted, we analyzed the expression dataset reported by Hughes *et al.*³⁴ In their study, the authors had deleted ~290 genes independently and monitored gene expression changes for the entire yeast genome for each of the mutant strains. A total of 21 transcription factors of the 290 genes were also represented in their dataset. For each of the 21 mutant strains, we evaluated the correlation between the number of genes they regulate in the network and the number of genes whose expression was affected (twofold change in expression) and found a lack of correlation. This suggested to us that transcriptional regulatory interactions might be backed up by regulations involving other transcriptional regulators, which might not necessarily be hubs.

A small fraction of essential genes are regulated by hubs alone

For essential genes that were not transcription factors, we investigated whether they tended to be uniquely regulated by hubs in order to assess if disruption of hubs can affect expression of essential genes in an organism. Interestingly we find that a majority of essential genes (60%, 437 genes, *p*-value to attain this value or lower purely by chance is 0.005) are regulated by more than one transcription factor, which includes a non-hub regulator. Only 26% (190 genes) of the genes are regulated by hubs alone. As the current network is not complete, the availability of additional data is likely to add additional non-hub regulators for essential genes. Hence, the estimate of 26% of essential genes regulated by hubs alone should be treated as an upper limit, and is likely to decrease with further data. These observations could imply the following: hubs rarely regulate essential gene by themselves, thus disruption of a hub can possibly affect gene expression changes of many genes but not necessarily the viability of the organism. These results then suggest that an organism can indeed evolve new regulatory states by withstanding mutations that affect regulatory hubs, since such mutation of hubs generally does not affect viability, even if the hub is disrupted. This would mean that transcriptional regulatory hubs need not be conserved across organisms, and might be lost or displaced selectively.

Regulatory hubs can be lost in evolution

We systematically analyzed regulatory hubs to test the above hypothesis regarding their dispensability in the course of evolution. For this purpose, we identified the ortholog of each of the 157 regulators in 15 fungal genomes and looked for a correlation between the number of genes a

† <http://arxiv.org/abs/cond-mat/0501169>

‡ http://sgd-lite.princeton.edu/download/yeast_datasets/phenotypes/footprinting/growthScores.tab

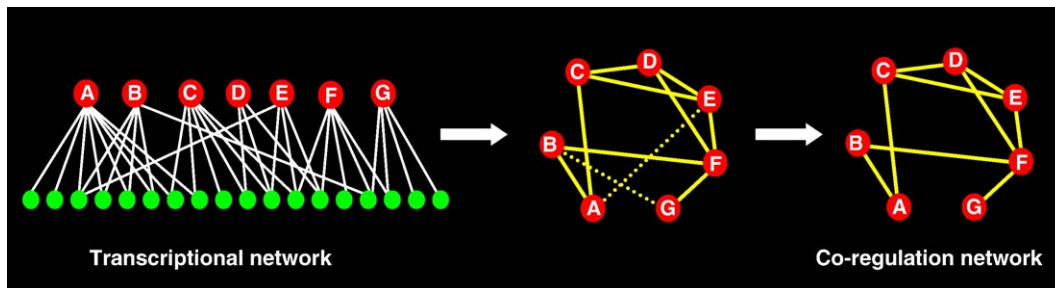


Figure 2. Schematic of the network transformation procedure. Starting from a transcriptional regulatory network, where red circles denote transcription factors, and green circles denote target genes, we construct a co-regulation network where nodes are only transcription factors. In the co-regulation network, we link two transcription factors if they share a common target gene. We remove links between a pair of transcription factors in the co-regulation network if they share the same number of target genes in 10,000 random trials, to ensure that only greater than non-random pairs of combinations are maintained.

transcription factor regulates and the number of organisms in which a regulator was conserved in evolution. If a correlation existed, it would mean that hubs are probably important for an organism and hence are more conserved than the regulators that regulate only a small subset of genes. However, the plot of conservation and number of genes regulated by a transcription factor (Figure 4) revealed a lack of correlation. This means that regulatory hubs are not essential in an organism, and can be subjected to loss or replacement by new contenders, thus providing the possibility for an organism to adapt and survive in changing conditions. This interpretation is supported independently by the earlier observations that each of the major lineages in eukaryotes has evolved its own set of transcriptional regulators (and hubs) by lineage-specific expansions.^{28,35,36} For instance, the C6 fungal-type zinc finger is abundant in fungi, but is entirely absent from other eukaryotes. Many of the key transcriptional controls performed by the C6 fungal-type zinc finger transcription factors are performed by unrelated proteins in other eukaryotes, suggesting that their rise to a position of a hub occurred relatively late in the evolution of eukaryotes. Taken together, these observations suggest that the apparent functional robustness of the transcriptional network must arise from causes other than what is apparently provided by the scale-free structure of the network. Given that this does not emerge from the obvious redundancy of hubs, it might emerge from a non-trivial feature hidden with the architecture of the transcriptional regulatory network.

Network transformation procedure reveals hidden properties

To address the above paradoxical observations, we explored the yeast transcriptional regulatory network for the existence of any hidden distributed architectural feature. Towards this goal, we developed a network transformation procedure that allowed us to arrive at a co-regulation network

starting from a transcriptional regulatory network. In simple terms, this transformation produces an undirected network, referred as the co-regulation network, where the nodes are only transcription factors, and two transcription factors are linked if they show a specific tendency to share a common set of target genes (Figure 2). Thus, starting with a transcriptional network composed of 157 transcription factors, 4410 target genes, involving 12,873 interactions, we arrived at the co-regulation network involving 157 transcription factors and 3459 links between them (Figure 3(a) and (b)). We compared the topological features of the transcriptional network and the co-regulation network to obtain insights into their properties, like robustness. This comparison yielded interesting and unexpected results with potentially important implications for functional robustness that we discuss in the following sections.

Uncovering a distributed architecture within a scale-free network

Interestingly, the degree of distribution of the co-regulation network suggested that the co-regulation network has a completely different architecture from the transcriptional regulatory network (Figure 3(a) and (b)). The observed distribution showed a clear central tendency similar to those observed for idealized distributed networks (Figure 1(b)), suggesting that the co-regulation network has a distributed architecture. We also calculated the average clustering co-efficient for the two networks and found that the value for the co-regulation network (average clustering coefficient, $CC=0.48$, $p=0.004$) was much higher than that for the transcriptional network (average $CC=0.12$, $p < 10^{-3}$). This supported the idea that a distributed architecture of co-regulation is hidden within a decentralized topology of gene regulation in yeast (Figure 3).

We suspected that this distributed architecture of the co-regulation network, which lies hidden in the decentralized architecture of the transcriptional regulatory network, might provide additional

robustness to the transcriptional regulatory network. In the transcriptional regulatory network, we find that the average number of target genes regulated by a transcription factor is 82, and in the co-regulation network, the number of other transcription factors with which a given transcription factor shares a target gene is 44. To evaluate the significance of these observations, we carried out simulation experiments in which we generated 10,000 random scale-free networks with a degree of distribution similar to that in the real network and applied the above transformation procedure to derive the co-regulation network. The simulations showed that the degree of distribution of co-regulation networks derived from random scale-free networks had very different shapes, suggesting that the observed distributed architecture for the real network has not arisen by chance. The distribution of the average connectivity in the co-regulation networks derived from random scale-free networks, i.e. the average number of transcription factors with which a given transcription factor share a target gene, is around 38. These simulations showed also that the higher value for target gene sharing (44 partners) encountered in the real yeast co-regulation network occurs with

$p < 0.009$, suggesting that the real transcriptional network has evolved a stronger tendency for co-regulation than expected by chance alone. This implies that there is a strong tendency for backup in the system due to the high level of co-regulation observed.

A co-regulation network is more robust to attack of hubs

Robustness of a network can be tested by removing nodes in one of two ways: (i) random removal of nodes, termed failure; and (ii) removal of highly connected nodes, called attack. We carried out simulations where both the transcriptional regulatory network and the transformed co-regulation network were subjected to attack and failure, and the fraction of interactions lost were monitored. As shown in Figure 5(a), when subjected to failure, the scale-free structure of the transcriptional network provides the observed robustness. However, when subjected to attack, which disrupts the hubs, the co-regulation network holds out, despite the breakdown of the scale-free structure of the basic transcriptional network (Figure 5(b)). The results remain unchanged even

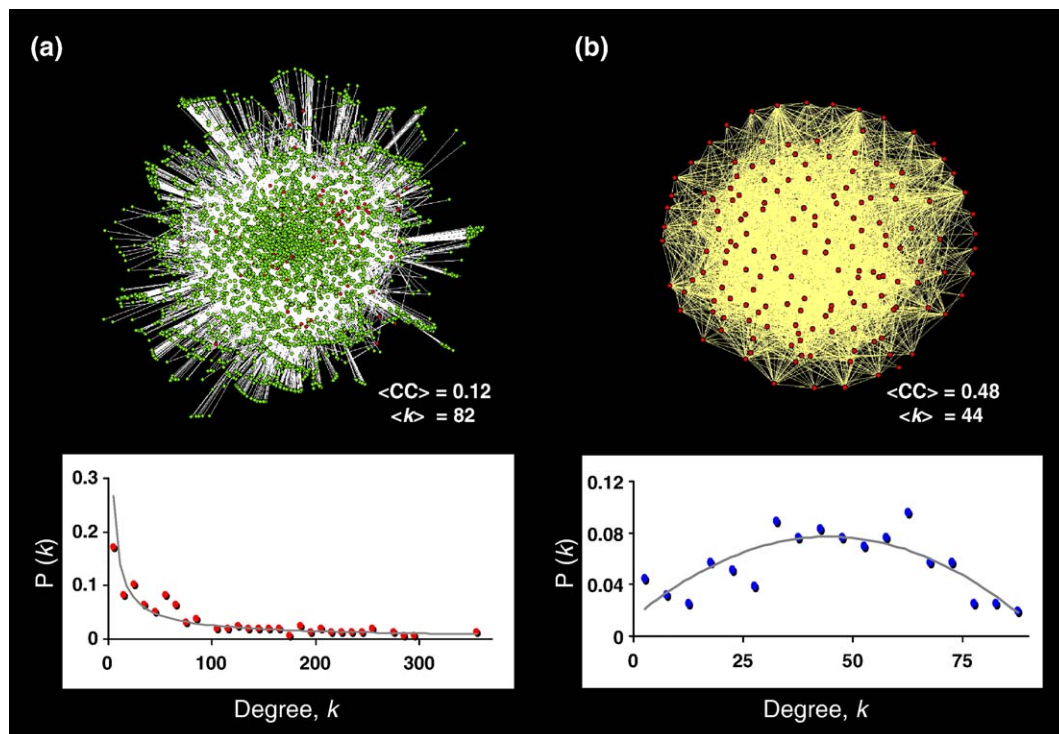


Figure 3. Network representation of (a) a transcriptional regulatory network in yeast with 12,873 regulatory interactions. This network is similar to the decentralized network in Figure 1, which is supported by an inhomogeneous degree distribution shown below. (b) The co-regulation network obtained after the transformation procedure described in Figure 2 is similar to the network with a distributed architecture shown in Figure 1(b). The degree distribution is homogeneous in the co-regulation network and shows a peaked behavior, as is expected for a network with a distributed architecture. Thus, the transformation procedure reveals the hidden distributed behavior among the transcription factors, which is not apparent in the original representation of the transcriptional network.

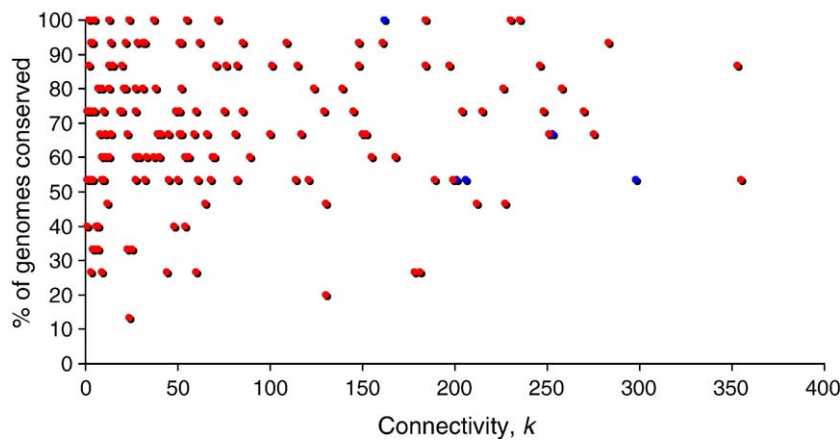


Figure 4. Plot showing the lack of correlation between connectivity and evolutionary conservation of regulatory proteins ($r^2=0.0079$). The absence of correlation suggests that the scale-free network can sustain mutations that disrupt regulatory hubs during evolution. The colored data points represent essential genes. Note that many regulatory hubs are not essential genes, indicating clearly that hubs in these scale-free networks can be lost or replaced in an organism. These observations suggest that such scale-free networks should have other features that should be

able to provide the additional robustness to withstand these mutations. This is uncovered by the network transformation procedure, which reveals the distributed architecture, where removal of highly connected proteins can still be tolerated and can be backed up by other transcription factors.

if we monitor the size of the largest component after attack or failure. Thus, the hidden co-regulation network appears to provide the additional robustness, which is not apparent in the transcriptional network, and might explain the apparently paradoxical observations mentioned above. Similar calculations done on 1000 random scale-free networks with an identical degree of distribution as the real transcriptional network reveal that this is not a generic property of all scale-free networks. In fact, in none of the 1000 random trials did we find this extent of robustness of the co-regulation network. This would mean that evolutionary forces have probably shaped individual interactions such that the co-regulation network can provide the additional robustness

against attack disrupting highly connected nodes. The higher values for the average connectivity of a transcription factor in the co-regulatory network suggest that this process has most probably taken place due to evolutionary pressure leading to selection for higher co-regulatory interactions between individual transcription factors.

Adaptive evolution by affecting regulatory hubs

A recent study by Fong *et al.*,³⁷ showed the first fixed mutations occurring during adaptation to environmental changes, typically affect global regulatory proteins. Such mutations, possibly affecting global regulators, alter the expression of many genes simultaneously, thereby increasing the

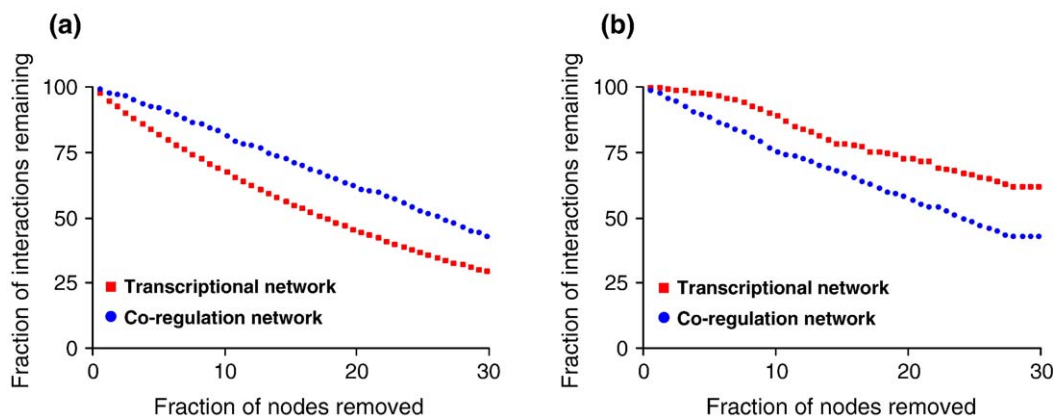


Figure 5. Robustness as measured by removal of nodes in the transcriptional network and the co-regulation network. Trends for the transcriptional network and for the co-regulation networks are shown in red and blue. (a) The trends observed when transcription factors are removed randomly (failure) reveals that robustness is provided by the transcriptional network. (b) The trends observed when regulatory hubs are removed first (attack) would show that the transcriptional network is vulnerable and disintegrates; however, the trend for the co-regulation network shows they can withstand removal of hubs and hence provide additional robustness, which can explain why living systems are tolerant to removal or replacement of regulatory hubs in their transcriptional regulatory network.

chance of arriving rapidly at the altered expression of a particular protein, which might confer the necessary adaptation.^{37–39} Subsequent mutations that become fixed are those that reset the expression levels of the irrelevant genes to baseline levels, but retain the differential expression of the adaptively relevant genes. Our result concerning the buffering provided by the distributed architecture of the co-regulation network against attack of the hubs in a transcriptional network is consistent with this study. The innate backup coming from the co-regulation network is likely to protect the overall transcriptional program of the organism from alterations affecting major regulators. At the same time, it provides alternative inputs that can restore expression of particular genes while leaving others unaffected, even if both groups of targets were under a common major regulator. Thus, “over-engineered” backup systems are likely to be a major determinant of the “evolvability” of gene expression in organisms faced with environmental diversity.

Conclusions

A simple transformation procedure uncovers the presence of a distributed architecture in the co-regulation network hidden behind the observed scale-free structure of the transcriptional network. This distributed architecture of the underlying co-regulation network differs from random scale-free networks with similar degrees of distribution as the real transcription network has a significantly higher tendency for co-regulation. This feature might account for the observation that regulatory hubs are not always essential and can be replaced or lost during evolution. This architecture provides a possible explanation for how different families of transcription factors have emerged in a lineage-specific manner across eukaryotes. It might account for the ability of organisms facing new environmental conditions to attain adaptive states of expression of particular genes rapidly *via* alterations of global transcriptional regulators, followed by a slower mutational resetting of expression of other genes. We believe that these principles inferred from the yeast network might be prevalent in other eukaryotic systems, providing a general model to understand the tremendous diversity of transcriptional interactions observed during the evolution of eukaryotes.

Materials and Methods

Data sets

The transcriptional regulatory network was assembled from the results of genetic, biochemical and ChIP-chip experiments.^{24–27,33,40} Only the interactions with $p < 0.001$ and $S.D. > 4$ from these high-throughput experiments

were considered. We were able to assemble a network of 4441 genes, which include 157 DNA-binding transcription factors, 4410 target genes and 12,873 regulatory interactions. Regulatory hubs were identified as proteins that regulate more than 150 target genes (the top 20% of the transcription factors with high out-going connectivity that cumulatively account for about more than half the number of regulatory interactions in the network). Of the 157 DNA-binding transcription factors, 33 qualified as hubs in our network. The complete genome sequence for the yeast species and other fungi compared in this study that include *Saccharomyces bayanus*, *Saccharomyces mikatae*, *Saccharomyces kluyveri*, *Saccharomyces castellii*, *Saccharomyces kudriavzevii*, *Kluyveromyces waltii*, *Candida albicans*, *Ashbya gossypii*, *Saccharomyces paradoxus*, *Schizosaccharomyces pombe*, *Aspergillus nidulans*, *Neurospora crassa*, *Magnaporthe grisea*, *Ustilago maydis* and *Cryptococcus neoformans* were obtained from the NCBI genome website. Information about essential genes was obtained from the Internet§. Gene expression data for the 21 transcription factor disruption mutant strains were obtained from the supplementary material accompanying Hughes *et al.*³⁴

Network transformation and other algorithms

In the transformation procedure, we link two transcription factors in the co-regulation network if the number of genes shared by the pair of transcription factors is greater than the average number of genes shared by the same pair in 10,000 random networks with a similar degree of distribution. This procedure is similar to that used by Porter *et al.* in their recent network analysis of the committees in the U.S. House of Representatives.⁴¹ It is of importance that the correct randomization procedure is employed, as an unrealistic randomization procedure can introduce bias into the estimation of statistical significance.⁴² To this end, all scale-free networks for the statistical tests used in the analysis have been generated from the original transcription network by randomly rewiring the network edges between transcription factors while maintaining the out-going and in-coming degree distributions of all the transcription factors and target genes. Such a procedure ensures that the in-degree and out-degree distributions of the random networks show identical pattern of behavior as observed in the original transcriptional network.^{43,44}

Ortholog detection procedure

Orthologs were identified using the bi-directional best hit approach. Whenever ortholog detection was not resolvable by this method, phylogenetic trees based on the relationships between the sequences of proteins that were picked up in the blast runs were built. Orthologs were then defined as the group of proteins that clustered together in the tree.

Statistical significance of our observations

To ensure that the observed phenomenon is not a property of the network structure, we carried out all calculations reported here by generating random scale-free networks with a degree of distribution similar to

that seen in the real transcriptional network of yeast. The results showed that the transformation procedure on the random networks does not yield the observed trends seen in the original dataset. *p* values were calculated as the fraction (over 10,000 trials) of the number of times a value observed in the random network was equal to or higher than that observed in the real network.

Acknowledgements

The authors gratefully acknowledge the Intramural research program of National Institutes of Health, USA for funding their research. We thank S. Geetha for carefully reading the proofs.

References

- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P. *et al.* (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.
- Boutros, M., Kiger, A. A., Armknecht, S., Kerr, K., Hild, M., Koch, B. *et al.* (2004). Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science*, **303**, 832–835.
- Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S. *et al.* (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M. *et al.* (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **421**, 231–237.
- Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B. *et al.* (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
- Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D. *et al.* (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684.
- Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M. *et al.* (2003). Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA*, **100**, 4678–4683.
- Song, J. H., Ko, K. S., Lee, J. Y., Baek, J. Y., Oh, W. S., Yoon, H. S. *et al.* (2005). Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol. Cells* **19**, 365–374.
- Wagner, A. (2005). *Robustness and Evolvability in Living Systems*, Princeton University Press, Princeton, NJ.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, **423**, 241–254.
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J. *et al.* (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature*, **406**, 959–964.
- Kafri, R., Bar-Even, A. & Pilpel, Y. (2005). Transcription control reprogramming in genetic backup circuits. *Nature Genet.* **37**, 295–299.
- Conant, G. C. & Wagner, A. (2004). Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. Biol. Sci.* **271**, 89–96.
- Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W. & Li, W. H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature*, **421**, 63–66.
- Bergman, A. & Siegal, M. L. (2003). Evolutionary capacitance as a general feature of complex gene networks. *Nature*, **424**, 549–552.
- Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. (1997). Evolution of genetic redundancy. *Nature*, **388**, 167–171.
- Wagner, A. (1996). Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol. Cybern.* **74**, 557–567.
- Papp, B., Pal, C. & Hurst, L. D. (2004). Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, **429**, 661–664.
- Papin, J. A., Price, N. D., Edwards, J. S. & Palsson, B. B. (2002). The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.* **215**, 67–82.
- Kitami, T. & Nadeau, J. H. (2002). Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genet.* **32**, 191–194.
- Edwards, J. S. & Palsson, B. O. (2000). Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* **16**, 927–939.
- Prachumwat, A. & Li, W. H. (2006). Protein function, connectivity, and duplicability in yeast. *Mol. Biol. Evol.* **1**, 30–39.
- Wagner, A. (2000). Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361.
- Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W. *et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K. *et al.* (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
- Horak, C. E., Luscombe, N. M., Qian, J., Bertone, P., Piccirillo, S., Gerstein, M. & Snyder, M. (2002). Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16**, 3017–3033.
- Svetlov, V. V. & Cooper, T. G. (1995). Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast*, **11**, 1439–1484.
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14**, 283–291.
- Albert, R., Jeong, H. & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.
- Doyle, J. C., Alderson, D. L., Li, L., Low, S., Roughan, M., Shalunov, S. *et al.* (2005). The “robust yet fragile” nature of the Internet. *Proc. Natl Acad. Sci. USA*, **102**, 14497–14502.
- Evangelisti, A. M. & Wagner, A. (2004). Molecular evolution in the yeast transcriptional regulation network. *J. Expt. Zool. B* **302**, 392–411.
- Maslov, S. & Sneppen, K. (2005). Computational architecture of the yeast regulatory network. *Phys. Biol.* **2**, S94–S100.

33. Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. & Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, **431**, 308–312.
34. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D. *et al.* (2000). Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109–126.
35. Coulson, R. M., Enright, A. J. & Ouzounis, C. A. (2001). Transcription-associated protein families are primarily taxon-specific. *Bioinformatics*, **17**, 95–97.
36. Lespinet, O., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* **12**, 1048–1059.
37. Fong, S. S., Joyce, A. R. & Palsson, B. O. (2005). Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **15**, 1365–1372.
38. Babu, M. M. & Aravind, L. (2006). Adaptive evolution by optimizing expression levels in different environments. *Trends Microbiol.* **14**, 11–14.
39. Dekel, E. & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**, 588–592.
40. Teichmann, S. A. & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genet.* **36**, 492–496.
41. Porter, M. A., Mucha, P. J., Newman, M. E. & Warmbrand, C. M. (2005). A network analysis of committees in the U.S. House of Representatives. *Proc. Natl Acad. Sci. USA*, **102**, 7057–7062.
42. Artzy-Randrup, Y., Fleishman, S. J., Ben-Tal, N. & Stone, L. (2004). Comment on “Network motifs: simple building blocks of complex networks” and “Superfamilies of evolved and designed networks”. *Science*, **305**, 1107; author reply 1107.
43. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.
44. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet.* **31**, 64–68.

Edited by J. Karn

(Received 9 February 2006; received in revised form 1 April 2006; accepted 14 April 2006)
Available online 27 April 2006