

# An Exponential Core in the Heart of the Yeast Protein Interaction Network

José B. Pereira-Leal,<sup>1</sup> Benjamin Audit,<sup>2</sup> José M. Peregrin-Alvarez,<sup>3</sup> and Christos A. Ouzounis

Computational Genomics Group, The European Bioinformatics Institute, EMBL Cambridge Outstation, Wellcome Trust Genome Campus, Cambridge, UK

Protein interactions in the budding yeast have been shown to form a scale-free network, a feature of other organized networks such as bacterial and archaeal metabolism and the World Wide Web. Here, we study the connections established by yeast proteins and discover a preferential attachment between essential proteins. The essential-essential connections are long ranged and form a subnetwork where the giant component includes 97% of these proteins. Unexpectedly, this subnetwork displays an exponential connectivity distribution, in sharp contrast to the scale-free topology of the complete network. Furthermore, the wide phylogenetic extent of these core proteins and interactions provides evidence that they represent the ancestral state of the yeast protein interaction network. Finally, we propose that this core exponential network may represent a generic scaffold around which organism-specific and taxon-specific proteins and interactions coalesce.

## Introduction

The advent of the genomics era is changing our focus from describing the molecular components of life in terms of their individual function to a more systems-level approach, where the focus is on the interactions between these components. The study of the complex networks they define is critical for the understanding of cellular-level and organism-level processes. Protein-protein interactions underlie the majority of cellular mechanisms. They form complex networks that have been shown to have a scale-free topology (Jeong et al. 2001; Wagner 2001), a property shared with other organized networks. This topological property is characterized by a power-law distribution of the number of connections established by each node. This attribute is particularly important in biological systems, as it conveys robustness (i.e., tolerance to errors) (Albert, Jeong, and Barabasi 2000). In scale-free networks, most nodes can be removed with little or no effect to the network; the network remains fully connected. Only the targeted removal of the most central (connected) nodes causes the collapse of the system (Albert, Jeong, and Barabasi 2000). In fact, it has been shown that connectivity ( $k$ ) and essentiality ( $e$ ) in the yeast protein interaction network are positively correlated (Albert, Jeong, and Barabasi 2000). Thus, the most highly connected proteins are more likely to be essential (i.e., to have a lethal phenotype).

Independently, it was observed that the yeast interaction network displays an anticorrelation between the connectivity of a protein ( $k_0$ ) and that of its binding partners ( $k_1$ ) (Maslov and Sneppen 2002). Hence, highly connected proteins are connected to proteins of low

connectivity, and vice-versa. Considering these two results, it follows that essential proteins are more highly connected, and, therefore, bind to proteins of lower connectivity that are more likely to be nonessential ( $n$ ). Thus, essential proteins should rarely bind other essential proteins. This prediction is, however, at odds with many cases of essential proteins that interact with other essential proteins (e.g., in the context of multiprotein complexes). Examples are RNA polymerase subunits (Shilatifard, Conaway, and Conaway 2003), ribosomal proteins (Garrett 1999), and the components of the origin recognition complex (ORC) (Bielinsky and Gerbi 2001), to name just a few. Thus, an apparent discrepancy emerges between topological studies of proteomics data, which suggest that essential proteins do not frequently interact with each other, and experimental studies of protein complexes, which indicate the opposite. This problem arises from the integration of distinct functional genomics data, which is critical to our understanding of the evolution of biological networks (Eisenberg and Levanon 2003; Qin et al. 2003; Kunin, Pereira-Leal, and Ouzounis 2004; Yook, Oltvai, and Barabasi 2004).

## Methods

### Interaction and Phenotypic Data

We obtained protein interaction information for *Saccharomyces cerevisiae* from the Database of Interacting Proteins (DIP) (Xenarios et al. 2002). Proteins were classified as essential ( $e$ ) or nonessential ( $n$ ) in glucose-rich medium, as determined in a genome-wide gene-deletion study (Winzeler et al. 1999). These form a network of 15,114 interactions (2,543  $e:e$ , 5,580  $e:n$  or  $n:e$ , and 6,991  $n:n$ ) involving 4,716 proteins, of which 826 are essential.

### Assessing Correlations

When assessing the existence of correlations between two variables, we compute Pearson's correlation coefficient  $r$ . We further report the slope of  $\alpha$  linear fit  $a$  and the ranges in which these quantities are computed.

### Fitting Connectivity Distributions

We fit the connectivity distribution to two concurrent models: power-law ( $P(k) \propto k^{-\gamma}$ ) and exponential ( $P(k) \propto$

<sup>1</sup> Present address: MRC-Laboratory of Molecular Biology, Hills Road, Cambridge, UK.

<sup>2</sup> Present address: Laboratoire de Physique, Ecole Normale Supérieure, Lyon, France.

<sup>3</sup> Present address: Genetic and Genomics Biology, Hospital for Sick Children, Toronto, Canada.

José B. Pereira-Leal and Benjamin Audit contributed equally to this work.

Key words: protein interactions, essential proteins, network evolution, *Saccharomyces cerevisiae*.

E-mail: jleal@mrc-lmb.cam.ac.uk.

*Mol. Biol. Evol.* 22(3):421–425. 2005

doi:10.1093/molbev/msi024

Advance Access publication October 20, 2004

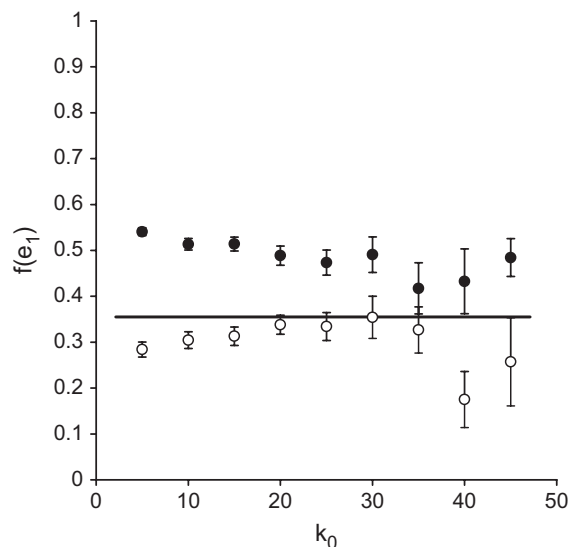


FIG. 1.—High frequency of essential:essential interactions. Frequency  $f(e_l)$  of essential neighbors as a function of the connectivity  $k_0$  of the reference central node. These data represent the relative frequency of essential neighbors, averaged over all central nodes with the same connectivity level and phenotype: filled circles correspond to essential central nodes ( $f(e_l | e_0)$  versus  $k_0$ ); open circles correspond to nonessential central nodes ( $f(e_l | n_0)$  versus  $k_0$ ). Error bars are the standard deviation from the mean. The solid line represents the expected frequency  $f_{\text{exp}}$  of essential neighbors, based on the observed connectivities of essential ( $k_e$ ) and nonessential ( $k_n$ ) proteins:  $f_{\text{exp}} = \frac{\sum k_e}{\sum k_e + \sum k_n}$ . This model assumes that there is no correlation between the phenotype of neighbors, but takes into account the higher connectivity of essential proteins and, hence, availability to establish connections.

$\exp(-\beta k)$ , using standard regression methods. To assess the goodness-of-fit for each model, we calculate the  $\chi^2_{\text{obs}}$  between the data and the model as well as its associated probability  $P(\chi^2 > \chi^2_{\text{obs}})$ —the smaller the observed  $\chi^2_{\text{obs}}$ , the larger the  $P$  value; that is, the better the analytical formulations explain the observed distribution.

### Phylogenetic Extent

The phylogenetic extent of each protein was determined as described previously (Peregrin-Alvarez, Tsoka, and Ouzounis 2003), using as reference the nonredundant protein sequence database SwAll (SwissProt and TrEMBL) (Bairoch and Apweiler 2000), which includes 837,986 protein sequences (Bacteria, 336,041; Archaea, 40,517; Eukaryota, 461,428; Fungi, 27,331; Metazoa, 28,2613; Protista, 30,467; and Viridiplantae, 121,017).  $P(h)$  is defined as the relative frequency of essential or nonessential proteins that have a homolog in the considered taxonomical groups.

An interaction is estimated to be conserved if both interacting proteins have homologs in a given model organism with complete genome sequence available, using Blast (Altschul et al. 1990) with default parameters and at a threshold value of  $E < 10^{-5}$ . Similar results are obtained when more stringent threshold values are used (i.e.,  $E < 10^{-10}$  and  $E < 10^{-20}$  [data not shown]).

## Results and Discussion

To address the apparent contradiction between the prediction that essential proteins do not frequently interact based on topological analysis of proteomics data and the experimental studies of protein complexes that indicate otherwise, we employ a parallel statistical analysis of the distribution of connectivity and essentiality in the protein interaction graph of *Saccharomyces cerevisiae*. Because the amount of data is steadily growing, in a first step, we confirm that the two topological features of the yeast interaction network described above apply to the current data set (see *Methods*). Indeed, we recover the previously observed positive correlation between the probability of essentiality and connectivity (Jeong et al. 2001) (data not shown) and the anticorrelation between the connectivities of neighboring nodes (Maslov and Sneppen 2002) (see below). We study the relative frequency of neighbors with an essential deletion phenotype ( $f(e_l)$ ) as a function of the phenotype ( $e_0$  or  $n_0$ ) and connectivity ( $k_0$ ) of the central node. We observe that essential proteins are, on average, more frequently connected to other essential proteins than are the nonessential proteins:  $f(e_l | e_0) = 0.48 \pm 0.10$  and  $f(e_l | n_0) = 0.27 \pm 0.10$ , respectively ( $P < 10^{-69}$ , two-way Student's  $t$ -test). This observation ( $f(e_l | e_0) > f(e_l | n_0)$ ) is valid over the full range of connectivity levels considered (fig. 1). These results indicate a preferential binding between essential proteins, as represented by the high frequency at which essential proteins bind each other.

From the model discussed above, we expected to observe an anticorrelation between the relative frequency of neighbors with an essential deletion phenotype ( $f(e_l)$ ) and the connectivity ( $k_0$ ) of the central node. We fail to detect such anticorrelation (data not shown):  $r = +0.24$  and  $\alpha = +0.001 \pm 0.001$  for  $k_0 \leq 60$  (see *Methods*), indicating that globally the connectivity of a node is not correlated to the phenotype of its binding partners. When we decompose the data by the phenotype of the central node ( $e_0$  or  $n_0$ ), we observe that at lower connectivity values ( $k_0 < 30$ ), a weak trend can be detected for essential ( $e_0$ ) central nodes, and an opposite trend can be detected for the neighbors of nonessential ( $n_0$ ) nodes (fig. 1).

The preferential attachment of essential proteins and the absence of anticorrelations between neighbor phenotypes and central-node connectivities raises the question of whether the global anticorrelation between  $k_1$  and  $k_0$  is independent of the node phenotype; that is, if it is applicable to all possible types of connections in the network ( $e_0:e_1$ ,  $e_0:n_1$ ,  $n_0:e_1$ , and  $n_0:n_1$ ). We address this issue by analyzing the mean connectivity of neighboring nodes ( $\langle k_1 \rangle$ ) as a function of the connectivity of the central, reference node ( $k_0$ ), for all types of pairs in the network. The complete network displays, as previously reported (Maslov and Sneppen 2002), an anticorrelation between  $k_0$  and  $\langle k_1 \rangle$  (fig. 2, open circles). This is characterized by a correlation coefficient  $r_{k_0:k_1} = -0.79$  and a slope  $\alpha_{k_0:k_1} = -0.22 \pm 0.02$  between  $\log(\langle k_1 \rangle)$  and  $\log(k_0)$  for  $k_0 \leq 60$ . Accordingly, the  $\log(\langle k_1 \rangle):\log(k_0)$  correlations for the pairs  $n_0:n_1$  are characterized by  $r_{k_{n0}:k_{n1}} = -0.86$  and  $\alpha_{k_{n0}:k_{n1}} = -0.31 \pm 0.03$ , for  $k_0 \leq 60$  (data not shown). In contrast,  $e_0:e_1$  connections do not follow this trend, displaying markedly weaker

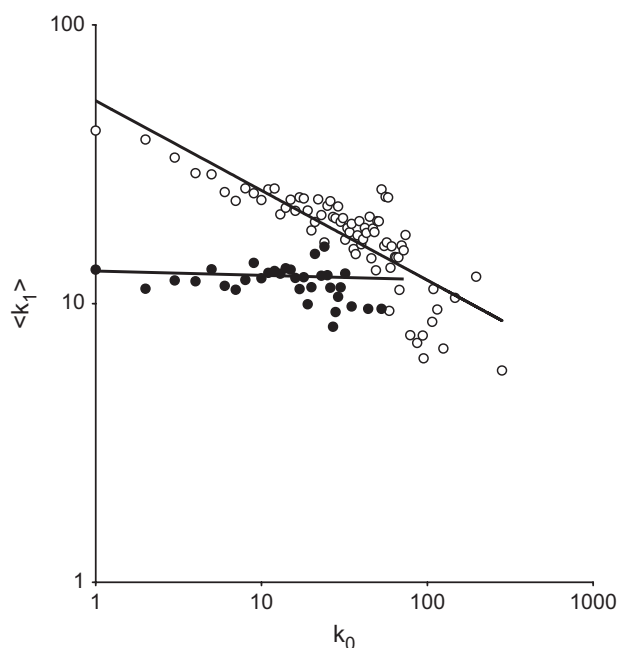


FIG. 2.—Essential proteins do not obey the global  $k_1:k_0$  anticorrelation rule. Average value of  $k_1$  as a function of  $k_0$ , displayed in a log-log scale for the global network (open circles) and the subnetwork formed only between essential proteins (filled circles); that is, all the nonessential proteins from the graph are removed to account for interactions present in this subnetwork.

correlations:  $r_{ke0:ke1} = -0.42$  and  $\alpha_{ke0:ke1} = -0.09 \pm 0.03$  for  $k_0 \leq 60$  (fig. 2, filled circles). For the pairs  $n_0:e_1$  and  $e_0:n_1$ , these correlations display intermediate strength:  $r_{kn0:ke1} = -0.73$  and  $\alpha_{kn0:ke1} = -0.28 \pm 0.04$  and  $r_{ke0:kn1} = -0.59$  and  $\alpha_{ke0:kn1} = -0.18 \pm 0.04$  for  $k_0 \leq 60$ , respectively (data not shown).

This finding shows that  $e_0:e_1$  connections between essential proteins in the network display specific topological properties that, because of their small relative weight (approximately 17% of the interactions), are hidden in the global trend. Moreover, this result provides an explanation to the preferential attachment of essential proteins. These proteins, even though they are, on average, more highly connected than nonessential proteins, are not as constrained by the anticorrelation of connectivity between neighboring nodes as other nodes in the network. Thus, there is no restriction on their frequent binding to other essential, highly connected proteins.

The distinct topological properties of  $e_0:e_1$  binding within the global network raises the possibility that they form a subnetwork with distinct topological properties. In fact, when we remove all the nonessential proteins from the network, we observe that approximately 97% of the essential proteins (801/826) are still connected to each other, forming a fully connected component that is significantly bigger than any fully connected component created by 1,000 random samplings of the same number of proteins from the global network (mean size of the largest connected component is  $437 \pm 43$ ,  $P < 0.001$ ). Furthermore, in this subnetwork, the correlations between  $\log(\langle k_1 \rangle)$  and  $\log(k_0)$  are characterized by  $r'_{ke0:ke1} = -0.19$  and  $\alpha'_{ke0:ke1} =$

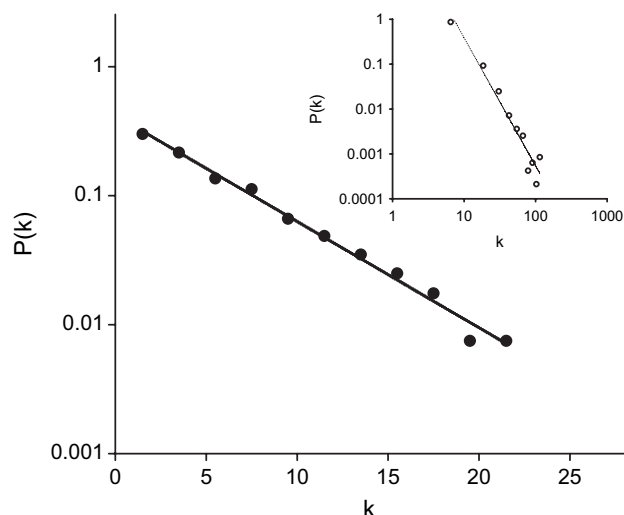


FIG. 3.—The essential subnetwork is an exponential network. Connectivity distribution  $P(k)$  in the essential subnetwork (filled circles) plotted in a log-linear scale, showing a clear exponential behavior. Fitting with an exponential model  $P(k) \propto \exp(-\beta k)$ , we obtain  $\beta = 0.18 \pm 0.01$  and  $\chi^2_{\text{obs}} = 3.0$  ( $P(\chi^2 \geq 3.0) = 0.96$ ). Note that when fitting with a power law model  $P(k) \propto k^{-\gamma}$ , we get a significantly larger  $\chi^2 = 77$  ( $P(\chi^2 \geq 77) < 10^{-12}$ ); that is, the exponential model explains much better the data. The inset panel shows in a log-log plot the relative connectivity distribution in the global network (open circles) displaying a power-law decay, indicating that the global network is a scale-free network. Note that the  $\chi^2$  value obtained when fitting with the power-law model ( $\chi^2 = 43$ ) is significantly lower than that obtained with the exponential model ( $\chi^2 = 511$ ). In other words, the power-law model is better suited to explain connectivity distribution of the global network.

$-0.03 \pm 0.03$  for  $k \leq 30$ , clearly indicating that these two variables are not (or only very weakly) correlated (fig. 2). This argues for the weak anticorrelation observed for  $e_0:e_1$  pairs in the global network being a consequence of this intrinsic property of the essential protein subnetwork. Surprisingly, the connectivity distribution in the essential subnetwork is well approximated by an exponential curve (fig. 3), indicating that this subnetwork does not display the scale-free topology observed in the global network (Jeong et al. 2001; Wagner 2001). It is interesting to note that the absence of the scale-free topology appears to be associated with the absence of anticorrelation between  $k_0$  and  $k_1$ . This finding raises the possibility that these two properties share a common biological origin.

By analyzing the topological properties of the interactions between essential proteins, we detected a preferential attachment between these proteins, resulting in an almost fully connected exponential network. This network includes those proteins involved in processes that have been proposed to have appeared early in evolution (e.g., transcription, translation, and replication) (Kyrpides, Overbeek, and Ouzounis 1999; Makarova et al. 1999; Harris et al. 2003). Moreover, essential proteins have been proposed to be under purifying selection; that is, their amino acid sequence evolves at a slower rate than other proteins (Hirsh and Fraser 2001), and they may be of wider phylogenetic extent (Kobayashi et al. 2003), suggesting that they may be of earlier evolutionary origin than

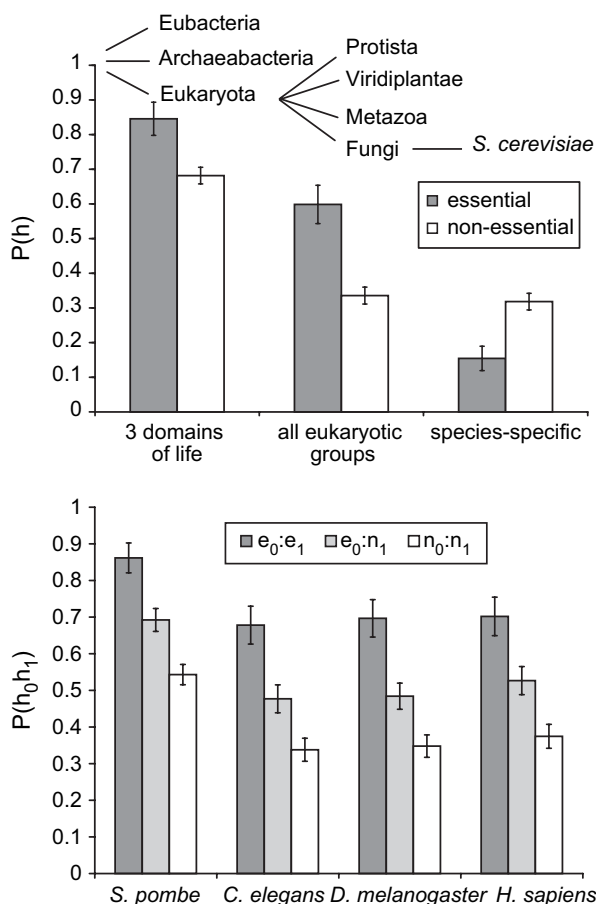


FIG. 4.—Essential proteins and their interactions are of earlier origin. (a) Probability  $P(h)$  of essential and nonessential yeast proteins being conserved in the considered taxonomic groups, based on their individual phylogenetic extent (see *Methods*). This result indicates that essential proteins are more phylogenetically extended and more likely to be of earlier origin than nonessential proteins. (b) Probability  $P(h_0, h_1)$  of both yeast proteins in a pair of interacting proteins to have homologs in model eukaryotic species with an available complete genome sequence. Essential:essential pairs are more likely to have homologs, suggesting that their interactions are more likely to be conserved. This is not a simple consequence of higher conservation of essential proteins, as the observed probabilities of pairwise conservation of essential interacting proteins are 20% to 50% higher than expected for all the species here considered (not shown). Note that directionality of interactions is not considered; that is, pairs  $n_0:e_1$  and  $e_0:n_1$ , are equivalent (represented as  $e_0:n_1$ ).

nonessential proteins. This condition would indicate that the network of essential proteins may represent the ancestral “core” of protein interactions in eukaryotes. Several lines of evidence support this hypothesis. First, if the essential proteins correspond to the components of an ancestral network, they must be of earlier evolutionary origin than other proteins. We observe that essential proteins have a wider phylogenetic extent than nonessential proteins (fig. 4a), indicating that they are indeed of earlier origin. They are more likely to be present in the three domains of life, including the four eukaryotic taxa considered than are nonessential proteins (Peregrin-Alvarez, Tsoka, and Ouzounis 2003). Conversely, nonessential proteins are more likely to be species specific than are essential proteins, and, hence, of more recent origin (fig. 4a). Second, if the essential subnetwork indeed

represents an earlier stage of the extant network, then the interactions between essential proteins must be of earlier origin. In fact, we observe that the pairwise conservation of  $e:e$  pairs in the genomes of model eukaryotic organisms and human is significantly higher than for other types of pairs (fig. 4b). Similarity of phylogenetic profiles is a way of predicting association between proteins (Pellegrini et al. 1999) and, thus, can be used to infer that the connections between essential proteins are specifically conserved and of earlier origin than those between nonessential proteins or those between essential and nonessential proteins. Interestingly, we also observe that  $e:n$  pairs appear to be more phylogenetically extended than  $n:n$  pairs (fig. 4b).

These arguments strongly infer that the network of interactions formed by essential proteins represent the ancestral form of the extant yeast protein interaction network. We further speculate, based on the wide phylogenetic extent of both essential proteins and their interactions, that the core of the yeast protein interaction network is shared by other organisms. It may represent a scaffold that is common to interaction networks in other species, around which interactions that are organism specific or cell-type specific coalesce. Indeed, several examples suggest that speciation has been accompanied by the addition of species-specific or taxon-specific components to existing core components of various processes, such as in the case of the transcriptional apparatus (Coulson, Enright, and Ouzounis 2001; Coulson and Ouzounis 2003), the translation apparatus (Lecompte et al. 2002), or metabolism (Peregrin-Alvarez, Tsoka, and Ouzounis 2003). Furthermore, the notion of a conserved genetic core composed of phylogenetically extended gene families is commonly accepted (Makarova et al. 1999; Harris et al. 2003; Peregrin-Alvarez, Tsoka, and Ouzounis 2003). Our results suggest that a core set of genes in ancestral genomes coded, as might be expected, for a fully connected protein interaction network, whose interactions have been conserved.

The scale-free topology has been suggested to have emerged early in evolution (Wagner 2003). Our results, however, do not support this idea—the subnetwork representing an early stage of the extant network displays an exponential connectivity distribution (fig. 3). One possibility is that this development is just a consequence of “historical noise” and that the primitive network was indeed a scale-free network. However, our interpretation is supported by previous work, suggesting that this exponential character may represent the state of an ancestral network. In mathematical models of network growth, it has been shown that if the preferential attachment of new nodes to existing nodes of high connectivity is relaxed, exponential topologies emerge (Krapivsky, Redner, and Leyvraz 2000). Furthermore, other modeling results suggest that an exponential network can also emerge if rewiring of existing nodes is also considered (Albert and Barabasi 2000). Note that in this model, rewiring is not random—it is still constrained by preferential attachment. In biological terms, the implication is that the gain and loss of interactions between existing components (rewiring) within the early network may have played a predominant role, compared with the addition of new components. This scenario is plausible, for example, assuming that replication

is not faithful and that new generations have slightly changed protein repertoires—some of the components will lose connections but gain new connections to other nodes.

By combining phenotypic and interaction data, we have been able to propose novel descriptive attributes of the yeast protein interaction network. More quantitative studies, along the lines of the work presented here, will be needed to determine to what extent the peculiar topological properties of the essential subnetwork follow from different underlying mechanisms. Furthermore, it remains to be determined whether in nonbiological systems, such as the World Wide Web or social networks, there is also a specific conservation of a core of components and interactions and whether these also display distinct topological properties. The analysis presented here may have consequences for our understanding of the overall structure and evolution of biological and other organized networks.

### Acknowledgments

We thank Eduardo Rocha and Cedric Vaillant for valuable discussions. J.B.P.-L. acknowledges support from the Foundation of Science and Technology—Portugal. B.A. was supported by a Marie Curie Fellowship of the European Community program “Improving Human Research Potential and the Socioeconomics Knowledge Base” under contract number HPMF-CT-2001-01321. J.M.P.A. was supported by the EUWOL network. C.A.O. thanks the UK Medical Research Council and IBM Research for additional support.

### Literature Cited

Albert, R., and A. L. Barabasi. 2000. Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* **85**:5234–5237.

Albert, R., H. Jeong, and A. L. Barabasi. 2000. Error and attack tolerance of complex networks. *Nature* **406**:378–382.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.

Bairoch, A., and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**:45–48.

Bielinsky, A. K., and S. A. Gerbi. 2001. Where it all starts: eukaryotic origins of DNA replication. *J. Cell Sci.* **114**:643–651.

Coulson, R. M., A. J. Enright, and C. A. Ouzounis. 2001. Transcription-associated protein families are primarily taxon-specific. *Bioinformatics* **17**:95–97.

Coulson, R. M., and C. A. Ouzounis. 2003. The phylogenetic diversity of eukaryotic transcription. *Nucleic Acids Res.* **31**:653–660.

Eisenberg, E., and E. Y. Levanon. 2003. Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **91**:138701.

Garrett, R. 1999. Mechanics of the ribosome. *Nature* **400**:811–812.

Harris, J. K., S. T. Kelley, G. B. Spiegelman, and N. R. Pace. 2003. The genetic core of the universal ancestor. *Genome Res.* **13**:407–412.

Hirsh, A. E., and H. B. Fraser. 2001. Protein dispensability and rate of evolution. *Nature* **411**:1046–1049.

Jeong, H., S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. 2001. Lethality and centrality in protein networks. *Nature* **411**:41–42.

Kobayashi, K., S. D. Ehrlich, A. Albertini et al. (96 co-authors) 2003. Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* **100**:4678–4683.

Krapivsky, P. L., S. Redner, and F. Leyvraz. 2000. Connectivity of growing random networks. *Phys. Rev. Lett.* **85**:4629–4632.

Kunin, V., J. B. Pereira-Leal, and C. A. Ouzounis. 2004. Functional evolution of the yeast protein interaction network. *Mol. Biol. Evol.* **21**:1171–1176.

Kyrpides, N., R. Overbeek, and C. Ouzounis. 1999. Universal protein families and the functional content of the last universal common ancestor. *J. Mol. Evol.* **49**:413–423.

Lecompte, O., R. Ripp, J. C. Thierry, D. Moras, and O. Poch. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* **30**:5382–5390.

Makarova, K. S., L. Aravind, M. Y. Galperin, N. V. Grishin, R. L. Tatusov, Y. I. Wolf, and E. V. Koonin. 1999. Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* **9**:608–628.

Maslov, S., and K. Sneppen. 2002. Specificity and stability in topology of protein networks. *Science* **296**:910–913.

Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* **96**:4285–4288.

Peregrin-Alvarez, J. M., S. Tsoka, and C. A. Ouzounis. 2003. The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* **13**:422–427.

Qin, H., H. H. Lu, W. B. Wu, and W. H. Li. 2003. Evolution of the yeast protein interaction network. *Proc. Natl. Acad. Sci. USA* **100**:12820–12824.

Shilatifard, A., R. C. Conaway, and J. W. Conaway. 2003. The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* **72**:693–715.

Wagner, A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* **18**:1283–1292.

———. 2003. How the global structure of protein interaction networks evolves. *Proc. R. Soc. Lond. B Biol. Sci.* **270**:457–466.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff et al. (52 co-authors). 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**:901–906.

Xenarios, I., L. Salwinski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg. 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**:303–305.

Yook, S. H., Z. N. Oltvai, and A. L. Barabasi. 2004. Functional and topological characterization of protein interaction networks. *Proteomics* **4**:928–942.

Brian Golding, Associate Editor

Accepted October 13, 2004