Supplementary Methods

Sequence Analysis

S. cerevisiae and S. paradoxus alignments were obtained from the *Saccharomyces* Genome Database (www.yeastgenome.org) and annotated with respect to a recent comparative study (Kellis et al. 2003). All alignments were then rerun using the software DIALIGN 2.2 (Morgenstern 1999), and a subset was visually inspected for accuracy. *C. elegans-C. briggsae* alignments were kindly provided by J. Stajich and have been reported in a recent publication (Stein et al. 2003). Estimates of amino acid divergence between *D. melanogaster* and *D. pseudoobscura* genes were provided by K. Thornton and appear in Richards et al. (2005). Orthologs for each comparison were generally found by reciprocal best-Blast hit or by synteny (Kellis et al. 2003; Stein et al. 2003; Richards et al. 2005). We calculated divergence estimates for all three datasets using the maximum likelihood method of Goldman and Yang (1994) as implemented in the PAML software package (Yang 1997).

Statistical Analysis

All statistical analyses were done using JMP version 5.0 (SAS Institute Inc.). Betweenness was log-transformed and connectivity was Box-Cox transformed before regression analyses in order to comply with normality assumptions; non-parametric statistical methods were used for all other analyses.

Network Analysis

Protein-Protein interaction networks for fly, yeast, and worm were obtained from the GRID database (Breitkreutz et al. 2003; http://biodata.mshri.on.ca/grid). This protein-protein interaction data comes from a variety of methods that detect interactions among proteins including: affinity precipitation, affinity chromatography, yeast two-hybrid, purified complex, reconstituted complex, biochemical assay, synthetic lethality, synthetic rescue, dosage lethality, dosage suppression, chemical lethality, and chemical rescue (Breitkreutz et al. 2003). GRID database data were converted to undirected adjacency matrices for simplicity. All network statistics were calculated using the Pajek software package (Batagelj and Mrvar 1998). Three measures of centrality were examined to provide slightly different perspectives on the position of a protein within the larger protein-protein interaction network: connectivity (also called "degree"; Freeman 1979) measures the number of interactors a given protein has, and thus corresponds to the row sums from the adjacency matrix. Specifically if we consider the indicator variable x_{ij}, *where* x_{ij}= 1 if vertex *i* connects to vertex *j* and 0 otherwise, then the connectivity of node n_i is defined as

$$C_D(n_i) = \sum_j x_{ij} .$$

If the size (i.e. the number of vertices) of a graph is *g*, than the maximum value for the connectivity of any given vertex is $g - 1$.

Closeness is a measure of distance between a vertex an all other vertices in a graph. The notion is that a vertex is central if it can quickly interact with all other vertices in a network. If we let $d(n_i,n_j)$ be the number of steps in the shortest path linking vertices *i* and *j*, then the total distance of *i* from all other vertices is $\sum_{j=1}^{g} d(n_i,n_j)$, for $j \neq i$. This leads to Sabidussi's (1966) definition of relative closeness:

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i,n_j)\right]^{-1} \cdot (g-1).$$

The maximum value for Sabidussi's closeness is thus $(g\text{-}1)^{-1}$, where a vertex is adjacent to all other vertices; relative closeness thus goes between 0 and 1.

A third measure of network centrality is betweenness. Betweenness measures the amount of information that passes through a node when interactions occur among vertices in a graph. By way of example, suppose that for node *j* and node *k* to interact in a given network, node *i* must be used as an intermediate. In such a network node *i* conveys a certain amount of "information" between nodes *j* and *k*. If we were to count all of the minimum paths that pass through node *i*, then we would have a measure of the amount of information (sometimes called 'stress') which node *i* carries throughout the network. Freeman (1979) defines betweenness in the following way: let g_{jk} be the number of geodesics (paths) linking nodes *j* and *k*. If all these geodesics are used with equal frequency, the probability of using any one geodesic is clearly $1/ g_{jk}$. We then are interested in the probability that interaction between nodes *j* and *k* occurs in such a way that the geodesic chosen goes through node *i*. If we let $g_{jk}(n_i)$ be the number of geodesics linking nodes *j* and *k* that include node *i*, than we can estimate the above probability as $g_{jk}(n_i)/ g_{jk}$. Note that this definition rests on the assumption that all geodesics are equally likely to be chosen. The normalized betweenness centrality for node n_i is therefore the sum of this probability over all pairs of nodes with the exception of the *i*th:

$$C_B(n_i) = \frac{\left(\sum_{j<k} g_{jk}(n_i)/g_{jk}\right)}{\left(\frac{(g-1)(g-2)}{2}\right)}$$

for i≠j≠k. As betweenness is just a sum of probabilities, the minimum value is zero, and the maximum value is $(g$-$1)(g$-$2)/2$, which is the number of pairs of vertices not including n_i. This maximum is attained when the *i*th node falls on all geodesics through the network; for relative betweenness, values fall between 0 and 1.

References for Supplementary Methods

Batagelj, V., and A. Mrvar. 2003. Pajek - Analysis and visualization of large networks. Pp. 77-103 *in* M. Junger, and P. Mutzel, eds. Graph Drawing Software. Springer, Berlin.

Breitkreutz, B. J., C. Stark, and M. Tyers. 2003. The GRID: The general repository for interaction datasets. Genome Biology **4**.

Freeman, L. C. 1979. Centrality in social networks. I. Conceptual clarification. Social Networks **1**:215-239.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution **11**:725-736.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423**:241-254.

Morgenstern, B., K. Frech, A. Dress, and T. Werner. 1998. DIALIGN: Finding local similarities by multiple sequence alignment. Bioinformatics **14**:290-294.

Richards, S., Y. Liu, B. R. Bettencourt et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene and cis-element evolution. Genome Research. **In press.**

Sabidussi, G. 1966. The centrality index of a graph. Psychometrika **31**:581-603.

Stein, L. D., Z. R. Bao, D. Blasiar et al. 2003. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLoS Biology **1**:166-192.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555-556.