BIOINFORMATICS ORIGINAL PAPER

Vol. 21 no. 3 2005, pages 338–348
doi:10.1093/bioinformatics/bti004

A more efficient search strategy for aging genes based on connectivity

Luca Ferrarini[1], Luca Bertelli[1], Jacob Feala[2], Andrew D. McCulloch[2] and Giovanni Paternostro[1,*]

[1]The Burnham Institute, 10901 North Torrey Pines Road, La Jolla , CA 92037, USA and [2]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

Received on April 1, 2004; revised on August 6, 2004; accepted on August 23, 2004
Advance Access publication September 3, 2004

ABSTRACT

Motivation: Many aging genes have been found from unbiased screens in model organisms. Genetic interventions promoting longevity are usually quantitative, while in many other biological fields (e.g. development) null mutations alone have been very informative. Therefore, in the case of aging the task is larger and the need for a more efficient genetic search strategy is especially strong.

Results: The topology of genetic and metabolic networks is organized according to a scale-free distribution, in which hubs with large numbers of links are present. We have developed a computational model of aging genes as the hubs of biological networks. The computational model shows that, after generalized damage, the function of a network with scale-free topology can be significantly restored by a limited intervention on the hubs. Analyses of data on aging genes and biological networks support the applicability of the model to biological aging. The model also might explain several of the properties of aging genes, including the high degree of conservation across different species. The model suggests that aging genes tend to have a higher number of connections and therefore supports a strategy, based on connectivity, for prioritizing what might otherwise be a random search for aging genes.

Contact: giovanni@burnham.org

INTRODUCTION

According to currently accepted evolutionary explanations of aging (Partridge and Gems, 2002; Rose, 1991), the age-related deterioration of functions is due to a generalized lack of adaptation in biological systems rather than to a specific genetic program. This is a consequence of the decreasing power of natural selection with age (Charlesworth, 2000). Other evolutionary mechanisms, for example intergenerational transfers (and related kin selection) (Lee, 2003; Maynard Smith, 1975) or sexual selection (Promislow, 2003), might play a role, but are not generally considered of major importance (Charlesworth, 2000; Partridge and Gems, 2002; Rose, 1991). This non-adaptive nature makes the study of the genetics of aging more challenging, but it might also allow us to formulate a model with general applicability because we are less limited by the historical components of biological adaptation. In other words, genetic network architecture influences any biological process, but usually it also matters if the genes under examination are involved in a specific function (e.g. cell growth if we are studying cancer). In the case of aging we might be able to infer the involvement of a gene purely from topological considerations, because natural selection for a specific function plays a lesser role.

The decline in physiological function observed during aging differs from that associated with disease: Taffet (2002) lists 147 major physiological parameters that decline with age in 22 body systems. Furthermore, the decline is progressive and gradual, initially affecting only physiological reserves (Taffet, 2002). There is no disease that has such a widespread effect on the biological function of an organism (Braunwald *et al*., 2001). In diseases, some organs and functions are usually affected to a major extent and others only secondarily and in a minor way (Braunwald *et al*., 2001). Aging-related dysfunction has therefore special properties: it is global (because of the large number of physiological functions declining), generalized (no specific function predominates) and gradual (distributed over a considerable portion of life span). No disease possesses these properties to the same extent. We suggest that aging is the biological dysfunction where network level properties of the genes have the greatest importance. In contrast, disease-related dysfunctions are likely to preferentially involve specific portions of a biological network.

Recently, many aging (or longevity) genes have been identified (Guarente and Kenyon, 2000; Hekimi and Guarente, 2003; Lin *et al*., 1998) by showing that they can retard aging when mutated or affected by interventions. Our aim has been to develop a model to better understand how these interventions or mutations can partially restore the functional decay associated with biological aging. The model does not focus on the causes of aging but on the mechanisms of the observed beneficial interventions. The necessity of using a network

*To whom correspondence should be addressed.

approach to the study of aging has been pointed out before (Promislow and Pletcher, 2002).

It is now increasingly accepted that the topology of genetic and metabolic networks is organized according to common properties. The most extensively studied is the degree distribution, which is often close to scale-free (Albert and Barabasi, 2002; Bergmann *et al*., 2003; Bray, 2003; Stuart *et al*., 2003), or in any case 'fat-tailed' (Dorogovtsev and Mendes, 2002). In these networks some nodes are hubs, with a great number of links, while the majority have few connections. The nodes in these models represent genes or metabolites and the links represent the relations among them. Much less developed are network models of biological function that also take into account the recent advances in the understanding of network architecture. Although for some networks, such as the Internet or the electric power grid, links are used for communication or transport, and can substitute for each other, genetic networks perform specific functions as a whole and the role of individual links and nodes is less easily predictable. We have therefore developed a functional network model, similar to artificial neural networks (Bhalla, 2003), that incorporates what we believe are the fundamental functional characteristics of biological networks, in order to help us understand the aging of biological systems and the effect of interventions on this process. We have used an approach similar to that used in digital evolution experiments (Lenski *et al*., 2003; Maynard Smith, 1992; O'Neill, 2003) to adapt our networks to perform a function. Digital evolution has been successfully used, for example, to investigate the evolutionary origin of complex features (Lenski *et al*., 2003) and the effect of mutations on the evolution of digital organisms of different complexity (Lenski *et al*., 1999). Our computational model differs from previous ones in many respects, for example in offering the possibility of studying artificial neural networks of different topology (scale-free versus random) as digital individuals.

We have tried to address the following questions: Should a view of aging as a generalized process of degradation lead us to be skeptical (Olshansky *et al*., 2002) of recent reports claiming substantial benefits after interventions aimed at single genes? Conversely, do these reports imply that aging is caused by specific genetic programs? How can we reconcile the observed conservation of aging genes among evolutionary very distant species with the prevailing non-adaptive view of aging? Can we propose a model that explains these facts and also makes testable predictions? Can we find initial evidence supporting these predictions? And, finally, can we use this model to devise a more efficient strategy for the search of aging genes?

SYSTEMS AND METHODS

Hardware

The software was run on three computer clusters: Falcon and Jet, located at the Burnham Institute and Blue Horizon, located at the San Diego Supercomputer Center. Jet is a 16 nodes (2 CPUs per node) cluster and Falcon a 41 nodes (2 CPUs per node) cluster. Both use parallel virtual machine (PVM) as a message-passing library. Blue Horizon is a 114 nodes (8 CPUs per node) cluster using message-passing interface (MPI) as a message-passing library.

Software

Parallel implementation The software uses parallel libraries in order to split the simulation over the nodes of the clusters. The parallelization is made by starting a number of processes equal to the number of individual networks plus one. We obtain one slave process for each individual and a master process, which coordinates the whole execution. The population is kept by the master process, which also randomly creates new sequences of symbols, sends symbols and individuals to each slave (one individual per slave) and gets results back. Each slave submits symbols to the individual network performing the classification job, and provides results back to the master.

Network implementation A network is a set of nodes (from now on the word node refers to the nodes of the network, not to the physical nodes of the cluster as above) connected by directional links that represent the flow of information. Our functional networks are composed of input nodes, output nodes, and internal nodes, each working on its input with a sigmoid function.

The value of a node, at a particular step of our evolution t, is determined as follows: given a node i, we define N as the set of nodes whose edges point to i, and W as the set of weights associated with each edge; thus, the node $N[k]$ is connected with the node i through a link whose weight is $W[k]$. Finally, let $a[t]$ be the coefficient for the sigmoid. The output for the node i is given by:

$$\text{Output\_}i(x,t) = 1/[1 + \mathrm{e}^{(-a[t]*x)}], \qquad (1)$$

where x is

$$x = \Sigma N[k]_{t-1} W[k]_{t-1},$$

where $N[k]_{t-1}$ indicates the output of the node $N[k]$ at the time $t - 1$.

Thus, each node is characterized by two values that change with time: its status (output) and its exponent modifier (a). When the population is created, the exponent modifiers are set to 1. The exponent modifier changes with time, in order to preserve a memory of previous states, by averaging with the status at each step. Input and output nodes are chosen randomly, amount to 2% of the total number of nodes and cannot coincide. The input nodes receive the sequence of numbers to be classified and the output nodes provide the classification.

When a new symbol is given to a network (to its input nodes), each input node evaluates its activation value [Equation (1)], by using the symbol as x. The activation values

are then propagated through weighted connections and at the end the activation values of all the output nodes are averaged.

A positive result indicates the input symbol is, for example, $+0.2$, a negative one leads to the opposite conclusion, -0.2. The network can also discriminate between positive numbers of different magnitude and between ranges of number.

Each individual in a population is characterized by the same architecture, which is stored in a global binary matrix. The implementation of an individual is made using a matrix of real values that represent weights associated with links. These matrices have as many rows (and columns) as the number of nodes N. In addition, an array of N elements stores the current status and exponent modifiers of each node.

Network topology The main topology we used is the scale-free model: a structure in which many nodes have a low degree of outgoing/incoming links, while a few nodes have a high degree (hubs); the link distribution follows a power law curve. The number of outgoing and incoming links for a node is decided using a power law probability distribution (Albert and Barabasi, 2002). We obtain a better power law distribution with lower values of average degree (average number of links per node). In any case, we always obtain a fat-tailed distribution with many highly connected nodes (Dorogovtsev and Mendes, 2002). Another topology we used is the random network, built by randomly selecting couples of nodes to be connected until we have the desired number of links (Erdos and Renyi, 1959). The link distribution of a random network decreases rapidly and follows a Poisson distribution with peak corresponding to the average degree of the network.

Evolution During evolution, a population of networks is selected using genetic algorithms (Holland, 1992) to perform a task: the correct classification of a sequence of different numbers in two groups (e.g. two groups of numbers of different absolute value or sign). The evolution process ends when the average success rate of the population is higher than a pre-fixed threshold (90%); if this condition is not reached, the evolution process ends after a pre-fixed number of generation steps (20), and a new population is created. During the evolution process, a genetic algorithm is used to simulate the evolution of the population. At the beginning, all of the individuals have random real values for their links (between -1 and 1), while their exponent modifiers are set to 1, and their status are set to 0. The evolution phase is divided into the following steps:

(1) A randomly created sequence of K symbols is submitted to each network of a population of 50 individual networks and the success rate is evaluated.

$$S = (s_1, \ldots, s_K), \qquad s_i = \pm 0.2.$$

(2) Given the sequence of symbols S, and defining succ_rate_j the success rate of the j-th network, we select the best individual in the population so that:

$$\text{Best\_network} = \text{argmax}_{(j=1\ldots N)}\ \text{succ\_rate\_}j;$$

a second individual is also selected, by using a tournament selection algorithm, which first creates a subgroup between the individuals (three-fourth of the entire population) and then selects the best individual within the group. This is done to avoid local maxima.

(3) These two individuals are used as parents for the offspring: we create as many offspring as the current number of individuals in the population (50) by involving two processes: crossover and mutation.

(4) In the crossover phase we set link weights, status and exponent modifiers values in each offspring. We use information coded into the parents, using the following rule: with a probability of 50%, the new value is a linear combination of the two parent values, while for the other 50%, the new value is set to the value of one of the parents (the chosen parent is decided with a 50–50% probability).

When the linear combination is chosen, we first select a random number $r \in [0..1]$ and then evaluate the new value for the offspring as:

$$\text{new\_value} = r * \text{parent1\_value} + (1 - r) * \text{parent2\_value}.$$

Thus, the linear combination is an average only when $r = 0.5$.

(5) The mutation process affects link weights, with the following rule: each individual has a probability of 0.3 to be mutated with a low mutation rate and a probability of 0.1 to be mutated with a high mutation rate. For each individual selected for mutation, the algorithm goes through the individual's links and randomly changes each of them according to the selected mutation rate: 0.1% for the high rate mutation and a 0.01% of probability for the low rate mutation. Two different mutation rates were used to generate individuals with a wider range of variation, and therefore facilitating evolution.

(6) We evaluate the performances of the new offspring by making them classify the same sequence of symbols as the original population has worked on. The final step is to create a new population by choosing the best individuals between the parent population and the offspring population. The population size does not change during the evolution phase.

Biological information is encoded in DNA and originates from natural selection (Adami, 1998; Maynard Smith, 1999). Our networks accumulate information by the analogous process of digital evolution, and this information is used to classify the inputs. Our aim is to model the degradation and restoration

of this information within individuals, to simulate aging and interventions on aging.

Node degradation After the evolution phase, node degradation is performed by changing all of the exponent modifiers that characterize the nodes: before a new sequence of symbols is submitted, each exponent modifier is set to 98% of its current value. This phase stops when only 60% of numbers or less are classified correctly by the networks.

Restoration During this phase, we reset node values (status and exponent modifiers) to the values they had before the degradation process started. In separate experiments, we restore the values of the 10% of nodes with more outgoing links (the hubs) and of the 10% with fewer outgoing links. The functional recovery in the classification ability of the networks is then measured.

Analysis of biological networks

For our systematic analysis (Section 4 of biological results), we downloaded datasets of interactions from the General Repository for Interaction Datasets (Breitkreutz *et al*., 2003). This online database contains interaction sets for *Saccharomyces cerevisae*, *Caenorhabditis elegans*, and *Drosophila melanogaster* (YeastGRID, WormGRID and FlyGRID, respectively). These datasets are dominated by physical interactions obtained by high-throughput techniques (e.g. yeast two-hybrid, affinity precipitation, etc.), but also include genetic interactions taken from the literature. Python software was written to import the lists of interactions, map genes to gene numbers and format the data into a network of nodes and links. The software has the capability to convert directed graphs into undirected graphs, eliminate duplicate links and calculate node connectivity. The Python development environment also provides interactive access to network data. The network connectivity of a sample of genes was inspected manually to verify the computer analysis.

Aging genes that had been shown to either extend or shorten life span were obtained from the SAGE database (Strauss and LaMarco, 2002). The number of links k_{aging} was found for each aging gene in the network, and the average local connectivity of aging genes $\langle k_{\text{aging}} \rangle$ was computed for each species. Since the distribution of links per node in most biological networks follow a power law rather than a normal distribution, standard statistical methods for comparing $\langle k_{\text{aging}} \rangle$ with the global mean $\langle k \rangle$ could not be used. Similar to Promislow (2004), we used a randomization procedure for evaluating the significance of $\langle k_{\text{aging}} \rangle$ for a set of genes of size m. Random samples of size m were extracted from the network and the mean connectivity $\langle k_{\text{rand}} \rangle$ of each sample was calculated and compared with $\langle k_{\text{aging}} \rangle$. A Python script was written to perform this test a large number of times ($N = 100\,000$; see below) and obtain the percentage of trials in which $\langle k_{\text{rand}} \rangle$ exceeded $\langle k_{\text{aging}} \rangle$. We then reported a P-value as a measure of statistical significance.

A simple analysis was performed to support our choice of N, the number of random samples. The mean value of $\langle k_{\text{rand}} \rangle$ over N random samples was computed for a range of N. This average was seen to quickly converge to the global network average for values of N much smaller (only a few hundred iterations) than our choice of 100 000. The P-values obtained after 10 000 and 100 000 iterations were also virtually identical.

Statistical analysis

All results are expressed as mean ± standard error of the mean. For comparisons of two groups we used paired or unpaired t-tests. The data presented in Figure 2 were analyzed with two-way analysis of variance. For analyses of some biological data we used χ^2 tests. The statistical software used was Prism (GraphPad).

RESULTS

The computational model

The main features of our computational model are as follows:

(1) The networks perform a function, and we reasoned that the simplest function shared by biological networks is classification, given that they usually have to produce the correct output in response to different signals coming from their internal or external environments. Among the fundamental properties of living systems are homeostasis (which implies a response to changes in the internal environment) and a capacity for response to stimuli (Mayr, 1997). They are similar to artificial neural networks, which, however, normally have a regular topology (Haykin, 1999).

(2) The topological structure of the networks can take different forms, including a scale-free degree distribution.

(3) The links and nodes are weighted and the links are directional.

(4) The function is not designed by us but evolves using the principles of natural selection [as in genetic algorithms (Holland, 1992)], that is, we have populations of networks undergoing mutations, crossing-over and selection.

(5) The degradation of function, simulating aging, is generalized and quantitative and does not necessitate the removal of nodes.

(6) Interventions on specific nodes can partially restore function, similarly to interventions on aging or longevity genes.

The model is computationally very intensive and was implemented using parallel programming on a cluster computer.

We have been able to show a clear difference between strategies aimed at restoring the function of a network acting

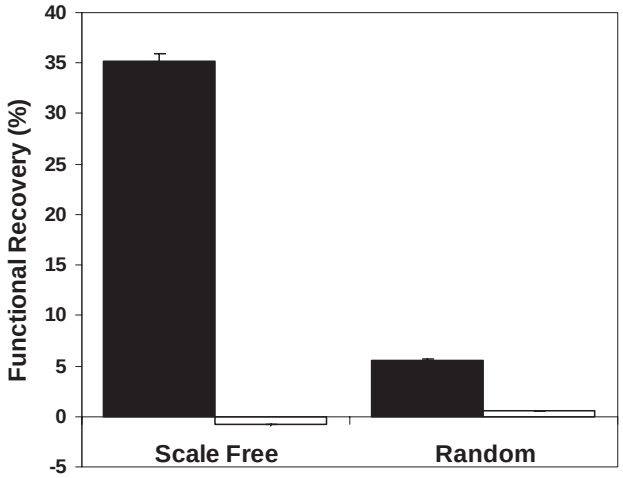

Fig. 1. The effect of restoring the values of the hubs after degradation. The hubs (defined as the top 10% of the link distribution) (filled bars), and the nodes with fewer outgoing links (defined as the bottom 10% of link distribution) (open bars) were restored to their values before degradation. The scale-free and random networks shown have 150 nodes, with an average 9 links per node. In scale-free networks, we demonstrate a large functional effect of interventions optimizing the hubs. For each experiment, $n = 200$.

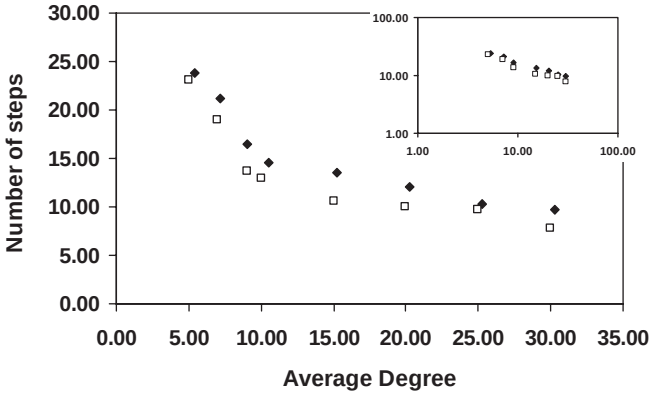

Fig. 2. The number of steps required for degradation of function (to an accuracy $\leq$60%), after damage equally affecting every node. Experiments are shown for networks with a scale-free topology (filled rhombuses) and with a random topology (open squares). Each data point is the average of 200 independent experiments. The average degree of the network is the average number of links per node. Using two-way analysis of variance, scale-free networks are shown to be functionally more resistant to generalized degradation, compared to random networks. The effect is significant ($P < 0.0001$). Additionally, networks with more links per node degrade more rapidly ($P < 0.0001$). Interestingly, this curve seems also to approximate a power law (inset) although we only studied it over a very limited range.

on highly connected nodes (hubs) versus less connected ones. This difference is much more pronounced in networks with a scale-free rather than a random distribution (Fig. 1). For example, if we restore the hubs (defined as the top 10% of the link distribution) to the value they had before degradation, in a scale-free network of 150 nodes (Fig. 1), we can recover $35 \pm 3.3\%$ of lost function. If, however, we restore the values of the least connected nodes (defined as the bottom 10% of link distribution) we do not improve the function of the network ($-0.86 \pm 1.3\%$ recovery) ($P < 0.0001$ versus most connected nodes, $n = 200$). We also performed the same experiment in a random network of comparable size and number of links (Fig. 1). In this case, restoring the most connected nodes only improves function by $5.6 \pm 1.9\%$ while restoring the least connected nodes does not improve function ($0.45 \pm 1.6\%$ recovery) ($P = 0.049$ versus most connected nodes, $n = 200$). The difference between scale-free and random network in functional recovery, after restoration of the hubs, was significant ($P < 0.0001$).

We explored the consequences of changing the number of nodes and the number of links per node in the networks. In a large number of experiments on scale-free networks, in which we studied networks ranging in size from 50 to 200 nodes and with average degree ranging from 5 to 30, the average functional recovery after restoring the hubs was 30.2% and after restoring the least connected nodes was -0.5%($n = 1800$). In every experiment the difference between these two interventions was significant ($P < 0.0001$). In a comparable large set of experiments on random networks, the functional improvements were on average 7.1 and 2.7%, respectively for interventions on most and least connected nodes ($n = 1800$). In the majority of experiments, the difference between these interventions was small and did not reach statistical significance.

We then measured the effect of generalized damage (node degradation affecting every node) on scale-free and random networks (Fig. 2) and found that scale-free networks were functionally more resistant. This is consistent with the results obtained by other authors (Albert and Barabasi, 2002; Albert *et al*., 2000; Jeong *et al*., 2001; Maslov and Sneppen, 2002), who, however, studied connectivity rather than function. In our model we do not remove nodes or links and the structure is left intact. The structure of a network has a powerful effect on its function but does not explain it completely. Several authors (Newman, 2003; Strogatz, 2001) have emphasized the importance of progressing from structural to functional studies.

Another factor affecting the functional degradation of the networks in our model was the average number of links per node (Fig. 2). This figure suggests that our computational model might be used to investigate the relation between network complexity and functional degradation.

The computational model shows that limited interventions on the hubs can cause substantial restoration of function in generally degraded networks that have the following

properties: a scale-free structure and evolution of function by a process similar to natural selection.

Testing the predictions of the model by analyses of aging genes and biological networks

An increasing number of studies have analyzed the architecture of biological networks (Albert and Barabasi, 2002; Bray, 2003). Comparative assessments have concluded that available data are still incomplete and distorted by high rates of false positives and false negatives (Alm and Arkin, 2003; von Mering *et al*., 2002). A database of aging genes, the Sage database (Strauss and LaMarco, 2002), also exists but it is likely to be incomplete, given that the rate of identification of aging genes does not seem to be slowing down. Even with these limitations, initial evidence for our model can be found using four different types of analysis:

(1) Evidence supporting our results is found in the recent network analysis of expression profiles of Bergmann *et al*. (2003). We have also confirmed this finding analyzing the original data from Arbeitman *et al*. (2002). Bergmann *et al*. report that the most connected gene in *Drosophila* (barkai-serv.weizmann.ac.il/ComparativeAnalysis/) is RPD3, a histone deacetylase regulating transcription. RPD3 is also one of only 20 genes that can retard aging in flies (Sage database) (Strauss and LaMarco, 2002). There are around 14 000 fly genes and the probability that this could happen by chance is only 1 in 700 ($P < 0.01$). Furthermore, RPD3 is one of only three fly genes for which evidence of involvement in aging is especially strong, because data exist in more than one species (Kim *et al*., 1999; Rogina *et al*., 2002).

(2) Further evidence supporting our results derives from the increased abundance of metabolic energy genes in both aging genes and hubs lists. Among the genetic interventions listed as promoting longevity in the Sage database (Strauss and LaMarco, 2002), 108 are genes of known function. Among these, the two largest groups by far are mitochondrial genes (31 were listed) and genes acting on insulin/carbohydrate metabolism (30 genes were listed). A statistical analysis can be performed by looking at individual species. Both in *Drosophila* and in *C.elegans* there is a significant increase in the number of genes involved in energy metabolism in the Sage database, compared to the entire genome data from Flybase (www.flybase.org) and Wormbase (www.wormbase.org). We used (both here and in all subsequent analyses) the Gene Ontology (Harris *et al*., 2004) classifications for genes involved in carbohydrate metabolism, insulin signaling and mitochondrial respiration. In *Drosophila* there were 4 energy metabolism genes out of 20 aging genes (20%), while in the whole genome data there were 125 energy metabolism genes out of 7938 annotated genes (1.6%) ($P < 0.0001$). In *C.elegans* there were 18 energy metabolism genes out of 50 aging genes (36%) while in the whole genome there were 114 energy metabolism genes out of 7022 annotated genes (1.6%) ($P < 0.0001$). Similar results are reported within a large-scale longevity RNAi screen in *C.elegans* where Lee *et al*. (2003) found a much larger proportion of mitochondrial genes compared to genomic data. Furthermore, the most widely tested intervention that can prolong life span is caloric restriction, obviously affecting energy metabolism (Masoro and Austad, 2001). Using DNA microarray data from humans, flies, worms and yeast, Stuart *et al*. (2003) have found 22 163 coexpression relationships that are conserved across evolution. The number of links among these conserved genes (which the author called metagenes) was distributed according to a power law (Stuart *et al*., 2003). We examined all of the genes that had more than 20 links (the largest hubs) from this dataset. There were 263 such genes. In this list there were 18 energy metabolism genes (7%), significantly more than in both the *Drosophila* and the *C.elegans* genome (1.6%) ($P < 0.0001$). Also in the *Drosophila* hub list (expression data) of Bergmann *et al*. (2003) there were 5 energy metabolism genes out of 43 (11.6%), significantly more than in the fly genome (1.6%) ($P < 0.0001$).

(3) Another well-characterized biological network is the metabolic network (Fell and Wagner, 2000; Jeong *et al*., 2000; Wagner and Fell, 2001; Wuchty and Stadler, 2003). All these studies [one comparing 43 organisms from all three domains of life (Jeong *et al*., 2000)] have found that the largest hubs in this network are ATP, ADP, inorganic phosphate and NAD. These are, of course, the central molecules involved in metabolic energy transfer, and are clearly affected by the products of aging genes involved in energy metabolism.

(4) A more systematic analysis of the connectivity of aging genes was also performed for the three main model organisms used in aging research yeast (*S.cerevisiae*), *C.elegans* and *D.melanogaster*. All the genes that were both in the SAGE database of aging genes (Strauss and LaMarco, 2002) and in the GRID database (Breitkreutz *et al*., 2003) (the general repository for interactions datasets) were examined. GRID contains physical and genetic interactions, but not expression network data.

Yeast. For yeast we substantially confirmed the results reported in a recent article by Promislow (2004), which appeared while our paper was under revision. We used similar methods to analyze a larger set of aging genes as well as a larger interaction dataset. As shown in Table 1, aging genes

Table 1. Connectivity of *S.cerevisae* aging genes

Gene	Gene no.	GRID links
S.cerevisiae		
ADA1	YPL254W	9
ATP2	YJR121W	3
BCY1	YIL033C	20
CCR4	YAL021C	28
CDC25	YLR310C	8
CDC6	YJL194W	7
CDC7	YDL017W	77
CTF4	YPR135W	116
CTT1	YGR088W	1
CYR1	YJL005W	13
DNA2	YHR164C	3
FOB1	YDR110W	13
GPA2	YER020W	8
GPD1	YDL022W	1
GPR1	YDL035C	1
HAP4	YKL109W	5
HAP5	YOR358W	12
HDF1	YMR284W	10
HDF2	YMR106C	65
HOG1	YLR113W	33
HXK2	YGL253W	6
LAG1	YHL003C	3
MPT5	YGL178W	12
MPT5	YGL178W	12
NMT1	YLR195C	1
NNT1	YLR285W	3
NPT1	YOR209C	2
PDE2	YOR360C	5
PHB1	YGR132C	1
PNC1	YGL037C	6
POL1	YNL102W	12
RAD27	YKL113C	59
RAD51	YER095W	39
RAD52	YML032C	40
RAD9	YDR217C	25
RAS1	YOR101W	6
RAS2	YNL098C	15
RIF1	YBR275C	2
RPD3	YNL330C	26
RSR1	YGR152C	9
RTG3	YBL103C	4
SGS1	YMR190C	43
SIM1	YIL123W	1
SIP2	YGL208W	11
SIR2	YDL042C	11
SIR3	YLR442C	16
SIR4	YDR227W	15
SNF1	YDR477W	26
SNF4	YGL115W	36
SOD1	YJR104C	18
SOD2	YHR008C	5
SSD1	YDR293C	8
SUN4	YNL066W	3
SWI4	YER111C	23
TPK2	YPL203W	8
UTH1	YKR042W	2
ZDS1	YMR273C	10
ZDS2	YML109W	24

Table 1. *Continued...*

Gene	GRID links
Aging links	981
Aging genes	58
Aging mean	16.9
P-value ($N = 100\,000$)	0.00035
Network links	38078
Network genes	4909
Network mean	7.8

This table lists the set of yeast aging genes from SAGE which were also in GRID (the General Repository of Interaction Datasets). The average number of links of aging genes (Aging mean) was larger than the average number of links for the entire network (Network mean). Aging links and Network links are the total number of links in each dataset (number of interactions per node multiplied by the number of nodes). The P-value was obtained by comparing the aging genes mean with that of 100 000 random samples of the same size.

have a higher average number of interactions compared to the entire network ($P = 0.00035$). We found interaction data for 58 out of 67 aging genes from SAGE (87%). The interaction dataset covers 4909 genes out of the 5773 genes in the refined yeast genome (Cliften *et al*., 2003) (85%).

C.elegans. For *C.elegans* the interaction dataset is very limited, and it covers only 2519 out of 19 542 genes in the genome (13%). Additionally, the average number of interactions in this dataset is only 3, compared to 7.3 for *Drosophila* and 7.8 for yeast. Therefore, we had interaction data for only 9 out of the 97 genes (9%) in the SAGE database and we could not detect any difference in connectivity compared to the entire network.

Drosophila. For *Drosophila*, the interaction dataset includes 7230 out of 14 015 genes (52%) and we could find interaction data for 18 out of 29 genes (62%) from SAGE. The average connectivity of aging genes was increased, with $P = 0.07$. These data are shown in Table 2.

DISCUSSION

It is useful to remember that the phenotypes caused by single gene mutants (and chemical treatments) have been shown to result from changes in the expression of many genes, up to several hundreds (Featherstone and Broadie, 2002; Hughes *et al*., 2000). This is why we designed our functional network model to include the optimization of multiple hubs. These co-regulated genes often are functionally related. In fact, one of the main co-regulated clusters in the study by Hughes *et al*. (2000) was composed of genes involved in mitochondrial respiration. Therefore, when we analyze the enrichment of energy metabolism genes in aging and hub lists we are not simply describing a property of a class, but it is likely that at least some of the same genes are affected by anti-aging interventions and by coexpression links.

Table 2. Connectivity of *D.melanogaster* aging genes

Gene	Gene no.	GRID links
D.melanogaster		
EcR	FBGN0000546	1
hsp68	FBGN0001230	8
chico	FBGN0024248	7
Sug	FBGN0036191	1
RPD3	FBGN0015805	12
cct1	FBGN0035230	4
indy	FBGN0036816	4
DPOSH	FBGN0040294	5
SOD2	FBGN0010213	1
cher	FBGN0014141	44
EF-1	FBGN0000556	43
hep	FBGN0010303	40
MSRA	FBGN0000565	3
ovo	FBGN0003028	21
VhaSFD	FBGN0027779	5
SOD1	FBGN0003462	2
cat	FBGN0000261	3
Dmp53	FBGN0039044	11
Aging links		215
Aging genes		18
Aging mean		11.9
P-value ($N = 100\,000$)		0.07
Network links		52826
Network genes		7230
Network mean		7.3

This table lists the set of *Drosophila* aging genes from SAGE which were also in GRID (the General Repository of Interaction Datasets). The average number of links of aging genes (Aging mean) was larger than the average number of links for the entire network (Network mean). Aging links and Network links are the total number of links in each dataset (number of interactions per node multiplied by the number of nodes). The P-value was obtained by comparing the aging genes mean with that of 100 000 random samples of the same size.

The two most common theories (Charlesworth, 2000; Partridge and Gems, 2002; Rose, 1991), not mutually exclusive, for the evolution of aging are as follows:

(1) Mutation accumulation (accumulation of mutations with late onset because of the decreasing power of natural selection at advanced ages).

(2) Antagonistic pleiotropy or trade-off (mutations that are beneficial early in life but damaging later are selected for).

Our model is compatible with these theories because they both predict that the genetic network is not adapted to the extra- and intracellular environment of the aged organism, as a consequence of the decreasing power of natural selection with age. These evolutionary mechanisms do not predict the existence of a specific genetic pathway solely responsible for aging. In any case generalized damage, affecting every physiological system, is what we observe in aged organism (Hazzard *et al*., 1999), and when damage is widespread even gene products not directly affected will not be optimized for the changed cellular environment. We are not modeling the causes of aging but the mechanism of its partial restoration, which many authors have observed after interventions on genes or on their products. Theses two aspects (cause and remedy) might not necessarily coincide.

Within our model, aging or longevity genes are the genes with more connections (the hubs), which contribute more to organ or body function restoration when their expression or activity is optimized. This offers a general explanation for the extensive degree of conservation of aging genes among very distant species (Guarente and Kenyon, 2000; Hekimi and Guarente, 2003), which might be puzzling in the absence of selective pressure. Other proposed explanations have limitations. The mutation accumulation theory does not seem to explain conservation (Partridge and Gems, 2002). The antagonistic pleiotropy (trade-off) theory would explain the conservation of aging genes if the link between early benefit and late damage is based on some conserved physiological mechanism (Partridge and Gems, 2002). The problem of conservation of aging genes is transformed into the problem of conservation of this link. Also, this theory requires that interventions improving aging are accompanied by a cost at an earlier age. Another proposal is that a conserved survival mechanism, to be activated in times of scarcity, might explain the conservation of the involvement of energy metabolism genes in aging (Hekimi and Guarente, 2003). This is relevant for only some aging genes and again it implies a cost, to explain why the survival mechanism is not active all the time. There are several apparent examples of cost-free longevity that would not be consistent with these models (Arantes-Oliveira *et al*., 2003; Dillin *et al*., 2002; Holzenberger *et al*., 2003; Lithgow and Gill, 2003). According to our model what is actively conserved is not the effect of a gene on aging but the central role of hub genes in biological systems. Highly connected genes are often evolutionary conserved (Bergmann *et al*., 2003).

Optimal interventions on aging based on our model would be quantitative and would act on multiple hubs. Almost all (19 out of 20) *Drosophila* aging genes from the Sage database (Strauss and LaMarco, 2002) have been found after experiments involving overexpression, heterozygous mutants or hypomorphic mutants. All these are quantitative interventions. Only one null mutant is listed, *chico*, a substrate for the insulin receptor, and the authors (Clancy *et al*., 2001) who described it argue that the mutants have only a mild reduction in the insulin receptor pathway. This seems likely, since severe mutations of the insulin receptor are lethal (Clancy *et al*., 2001). Therefore the effect of this genetic intervention is also quantitative. If a pathway that had aging as the only function existed, complete genetic knockouts would also be beneficial. The efficacy of all these quantitative interventions support a network model of aging, in which genes that do not have an optimal level of activity do not achieve optimal function.

If we view aging as a generalized accumulation of random damage (Olshansky *et al*., 2002), we might conclude that the lack of a specific genetic program for aging might imply that treatments are necessarily very problematic or impossible. Our computational model shows that, even in the worst case scenario, where aging results from damage to (or lack of optimization of) every gene, a substantial benefit can be obtained from a targeted intervention aimed at optimizing hub function. Any action of natural selection on aging (Lee *et al*., 2003; Maynard Smith, 1975; Promislow, 2003) would act on a pre-existing [since we can also find it in prokaryotes (Jeong *et al*., 2000)] scale-free network structure, which would represent a constraint on any evolutionary change. Strong departures from our non-adaptive model would represent a trace of the action of natural selection and might be used to estimate its effects on aging.

Limitations of the model

Are all aging genes always hubs? The biological data we analyzed do not support an affirmative answer. There are three reasons that might explain this:

(1) Limitations of existing biological datasets. Comparisons of different protein–protein interaction datasets in yeast show very little overlap, suggesting that the coverage might still be incomplete (von Mering *et al*., 2002). The authors of the recent paper describing the interaction map in *Drosophila* stated clearly that they estimate only 40% of the interactions in their high confidence subset are likely to be biologically relevant (Giot *et al*., 2003). There are also different classes of molecular networks [e.g. protein–protein physical interactions and expression networks, see Xia *et al*. (2004) for a more comprehensive list] and it is not clear which are more relevant to our model. The fact that some of the biological evidence for our model is stronger in yeast, where the existing interaction data are much more extensive (Bork *et al*., 2004), seems to support the relevance of this consideration.

(2) Other network properties beside local connectivity might be important. By local connectivity we mean the number of links of a node, for example the number of interactions of a protein. It is becoming clear that there might be more than one type of hub (Han *et al*., 2004). Measures of global connectivity can also be obtained (Chin and Samanta, 2003). All the large biological datasets only list the presence or absence of an interaction but it is clear that we would also need to know by how much one protein affects another to make completely accurate functional predictions.

(3) Natural selection might affect aging in a way that distorts network properties. Many believe, however, that non-adaptive forces might predominate in shaping biological aging (Partridge and Gems, 2002; Rose, 1991). In any case, since natural selection acts on existing structures (Futuyma, 1998), it might just reinforce the actions of genes that affect aging because of their topological properties, for example by reinforcing the role of the hubs.

In biological aging, gene products (proteins and their substrates) change with age and beyond a certain level the change affects function significantly. Some genetic changes or interventions can modify the functional decline (at least in model organisms). For interventions acting when aging is already under way our model applies directly, while for genetic mutations, and for interventions starting from earlier life stages, we can interpret any effect as due to an increased resistance to change of the gene products. Computationally, the scenarios of having a gene product not change or having it change and then restore it are equivalent as far as their final effect is concerned. But it is certainly true that there is much that we do not know about the aging process and the applicability of this model. Having alternative models and investigating their relevance to biological reality might help us to understand aging better.

We have studied the effect of changing some parameters of our model (e.g. the number of nodes and the number of links). The parameters used during evolution were optimized to make the evolution of functional networks possible during a realistic computational time frame (Mitchell, 1998). Our present computational model can only be used to derive a very general qualitative conclusion: restoring some nodes of a network can have a large effect on the function of the network uniquely because of their topological properties. Future studies might try to more closely approximate biological reality, for example by using real biological networks, with higher number of nodes, for functional experiments, but computational requirements might be large. Another feature that should be incorporated in future studies is the hierarchical model of biological networks (Barabasi and Oltvai, 2004).

CONCLUSIONS

Many aging genes have been found from unbiased screens in model organisms. As discussed above, genetic interventions promoting longevity are usually quantitative, and might require a large number of experiments to find the optimal level, while in many other biological fields (e.g. development) null mutations alone have been very informative. In the case of aging therefore the task is larger and the need for a more efficient genetic search strategy is especially strong. We suggest that in those species in which large collections of known mutants are available [e.g. *Drosophila*, where a genome wide collection is at a very advanced stage of development (Spradling *et al*., 1999)], it might be advantageous to first examine the genes that have the largest number of links. This might also be a useful strategy in human centenarian studies, where researchers are now searching within chromosomal

regions that are likely to contain alleles affecting human aging (Puca *et al*., 2001).

Our model suggests that aging genes tend to have a higher number of connections and therefore supports a strategy for prioritizing, with limited additional effort, what might otherwise be a random search. Limitations of existing interaction datasets and the possibility that other network properties besides local connectivity might be relevant should, however, be considered.

ACKNOWLEDGEMENTS

G.P. is grateful for the support received from the Ellison Medical Foundation and the American Federation for Aging Research.

REFERENCES

Adami,C. (1998) *Introduction to Artificial Life*. Springer-Verlag, NY.

Albert,R. and Barabasi,A.L. (2002) Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**, 47–97.

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

Alm,E. and Arkin,A.P. (2003) Biological networks. *Curr. Opin. Struct. Biol.*, **13**, 193–202.

Arantes-Oliveira,N., Berman,J.R. and Kenyon,C. (2003) Healthy animals with extreme longevity. *Science*, **302**, 611.

Arbeitman,M.N., Furlong,E.E., Imam,F., Johnson,E., Null,B.H., Baker,B.S., Krasnow,M.A., Scott,M.P., Davis,R.W. and White,K.P. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science*, **297**, 2270–2275.

Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.

Bergmann,S., Ihmels,J. and Barkai,N. (2003) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol.*, **2**, E9.

Bhalla,U.S. (2003) Understanding complex signaling networks through models and metaphors. *Prog. Biophys. Mol. Biol.*, **81**, 45–65.

Bork,P., Jensen,L.J., von Mering,C., Ramani,A.K., Lee,I. and Marcotte,E.M. (2004) Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.*, **14**, 292–299.

Braunwald,E., Fauci,A.S., Kasper,D.L., Hauser,S.L., Longo,D.L. and Jameson,J.L. (2001) *Harrison's Principles of Internal Medicine*, 15th edn. McGraw-Hill, Boston.

Bray,D. (2003) Molecular networks: the top-down view. *Science*, **301**, 1864–1865.

Breitkreutz,B.J., Stark,C. and Tyers,M. (2003) The GRID: the General Repository for Interaction Datasets. *Genome Biol.*, **4**, R23.

Charlesworth,B. (2000) Fisher, Medawar, Hamilton and the evolution of aging. *Genetics*, **156**, 927–931.

Chin,C.S. and Samanta,M.P. (2003) Global snapshot of a protein interaction network—a percolation based approach. *Bioinformatics*, **19**, 2413–2419.

Clancy,D.J., Gems,D., Harshman,L.G., Oldham,S., Stocker,H., Hafen,E., Leevers,S.J. and Partridge,L. (2001) Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science*, **292**, 104–106.

Cliften,P., Sudarsanam,P., Desikan,A., Fulton,L., Fulton,B., Majors,J., Waterston,R., Cohen,B.A. and Johnston,M. (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science*, **301**, 71–76.

Dillin,A., Crawford,D.K. and Kenyon,C. (2002) Timing requirements for insulin/IGF-1 signaling in *C.elegans*. *Science*, **298**, 830–834.

Dorogovtsev,S.N. and Mendes,J.F.F. (2002) *Evolution of Networks*. Oxford University Press, Oxford.

Erdos,P. and Renyi,A. (1959) On random graphs. *Publicationes Mathematicae*, **6**, 290–297.

Featherstone,D.E. and Broadie,K. (2002) Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *BioEssays*, **24**, 267–274.

Fell,D.A. and Wagner,A. (2000) The small world of metabolism. *Nat. Biotechnol.*, **18**, 1121–1122.

Futuyma,D.J. (1998) *Evolutionary Biology*, 3rd edn. Sinauer, Sunderland.

Giot,L., Bader,J.S., Brouwer,C., Chaudhuri,A., Kuang,B., Li,Y., Hao,Y.L., Ooi,C.E., Godwin,B., Vitols,E. *et al*. (2003) A protein interaction map of *Drosophila melanogaster*. *Science*, **302**, 1727–1736.

Guarente,L. and Kenyon,C. (2000) Genetic pathways that regulate ageing in model organisms. *Nature*, **408**, 255–262.

Han,J.D., Bertin,N., Hao,T., Goldberg,D.S., Berriz,G.F., Zhang,L.V., Dupuy,D., Walhout,A.J., Cusick,M.E., Roth,F.P. *et al*. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, **430**, 88–93.

Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al*. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32** (Database issue), D258–D261.

Haykin,S.S. (1999) *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice-Hall, Upper Saddle River, NJ.

Hazzard,W.R., Blass,J.P., Ettinger,W.H., Halter,J.B. and Ouslander,J.G. (1999) Principles of geriatric medicine and gerontology. McGraw-Hill, NY.

Hekimi,S. and Guarente,L. (2003) Genetics and the specificity of the aging process. *Science*, **299**, 1351–1354.

Holland,J.H. (1992) *Adaptation in Natural and Artificial Systems*, 2nd edn. MIT Press, Cambridge.

Holzenberger,M., Dupont,J., Ducos,B., Leneuve,P., Geloen,A., Even,P.C., Cervera,P. and Le Bouc,Y. (2003) IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*, **421**, 182–187.

Hughes,T.R., Marton,M.J., Jones,A.R., Roberts,C.J., Stoughton,R., Armour,C.D., Bennett,H.A., Coffey,E., Dai,H., He,Y.D. *et al*. (2000) Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109–126.

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

Jeong,H., Tombor,B., Albert,R., Oltvai,Z.N. and Barabasi,A.L. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

Kim,S., Benguria,A., Lai,C.Y. and Jazwinski,S.M. (1999) Modulation of life-span by histone deacetylase genes in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, **10**, 3125–3136.

Lee,R.D. (2003) Rethinking the evolutionary theory of aging: transfers, not births, shape senescence in social species. *Proc. Natl Acad. Sci. USA*, **100**, 9637–9642.

Lee,S.S., Lee,R.Y., Fraser,A.G., Kamath,R.S., Ahringer,J. and Ruvkun,G. (2003) A systematic RNAi screen identifies a critical role for mitochondria in *C.elegans* longevity. *Nat. Genet.*, **33**, 40–48.

Lenski,R.E., Ofria,C., Collier,T.C. and Adami,C. (1999) Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, **400**, 661–664.

Lenski,R.E., Ofria,C., Pennock,R.T. and Adami,C. (2003) The evolutionary origin of complex features. *Nature*, **423**, 139–144.

Lin,Y.J., Seroude,L. and Benzer,S. (1998) Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science*, **282**, 943–946.

Lithgow,G.J. and Gill,M.S. (2003) Physiology: cost-free longevity in mice? *Nature*, **421**, 125–126.

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

Masoro,E.J. and Austad,S.N. (2001) *Handbook of the Biology of Aging*. Academic Press, San Diego, CA.

Maynard Smith,J. (1975) Altruism, social behaviour, and sex. In Maynard Smith, J. (ed.), *The Theory of Evolution*, 3rd edn. Cambridge University Press, Cambridge, pp. 193–215.

Maynard Smith,J. (1992) Evolutionary biology. Byte-sized evolution. *Nature*, **355**, 772–773.

Maynard Smith,J. (1999) The 1999 Crafoord Prize Lectures. The idea of information in biology. *Quat. Rev. Biol.*, **74**, 395–400.

Mayr,E. (1997) *This is Biology*. Belknap/Harvard, Cambridge, MA.

Mitchell,M. (1998) *An Introduction to Genetic Algorithms*. MIT Press, Boston, MA.

Newman,M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

O'Neill,B. (2003) Digital evolution. *PLoS Biol.*, **1**, E18.

Olshansky,S.J., Hayflick,L. and Carnes,B.A. (2002) No truth to the fountain of youth. *Sci. Am.*, **286**, 92–95.

Partridge,L. and Gems,D. (2002) Mechanisms of ageing: public or private? *Nat. Rev. Genet.*, **3**, 165–175.

Promislow,D. (2003) Mate choice, sexual conflict, and evolution of senescence. *Behav. Genet.*, **33**, 191–201.

Promislow,D.E. and Pletcher,S.D. (2002) Advice to an aging scientist. *Mech. Ageing Dev.*, **123**, 841–850.

Promislow,D.E. (2004) Protein networks, pleiotropy and the evolution of senescence. *Proc. R. Soc. Lond. B Biol. Sci.*, **271**, 1225—1234.

Puca,A.A., Daly,M.J., Brewster,S.J., Matise,T.C., Barrett,J., Shea-Drinkwater,M., Kang,S., Joyce,E., Nicoli,J., Benson,E. *et al*. (2001) A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc. Natl Acad. Sci. USA*, **98**, 10505–10508.

Rogina,B., Helfand,S.L. and Frankel,S. (2002) Longevity regulation by *Drosophila* Rpd3 deacetylase and caloric restriction. *Science*, **298**, 1745.

Rose,M.R. (1991) *Evolutionary Biology of Aging*. Oxford University Press, NY.

Spradling,A.C., Stern,D., Beaton,A., Rhem,E.J., Laverty,T., Mozden,N., Misra,S. and Rubin,G.M. (1999) The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics*, **153**, 135–177.

Strauss,E. and LaMarco,K. (2002) Science of aging knowledge environment: one-stop shopping for researchers in the field of aging. *Exp. Gerontol.*, **37**, 1297–1301.

Strogatz,S.H. (2001) Exploring complex networks. *Nature*, **410**, 268–276.

Stuart,J.M., Segal,E., Koller,D. and Kim,S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.

Taffet,G.E. (2002) Physiology of aging. In Cassel,C.K. (ed.), *Geriatric Medicine*. Springer, NY, pp. 27–35.

Taylor,C. and Jefferson,D. (1995) Artificial life as a tool for biological inquiry. In Langton,C.G. (ed.), *Artificial Life*. MIT Press.

von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale datasets of protein–protein interactions. *Nature*, **417**, 399–403.

Wagner,A. and Fell,D.A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. B Biol. Sci.*, **268**, 1803–1810.

Wuchty,S. and Stadler,P.F. (2003) Centers of complex networks. *J. Theor. Biol.*, **223**, 45–53.

Xia,Y., Yu,H., Jansen,R., Seringhaus,M., Baxter,S., Greenbaum,D., Zhao,H. and Gerstein,M. (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.*, **73**, 1051–1087.