Protein networks, pleiotropy and the evolution of senescence

Daniel E. L. Promislow

Department of Genetics, University of Georgia, Athens, GA 30602-7223, USA (promislow@uga.edu)

The number of interactions, or connectivity, among proteins in the yeast protein interaction network follows a power law. I compare patterns of connectivity for subsets of yeast proteins associated with senescence and with five other traits. I find that proteins associated with ageing have significantly higher connectivity than expected by chance, a pattern not seen for most other datasets. The pattern holds even when controlling for other factors also associated with connectivity, such as localization of protein expression within the cell. I suggest that these observations are consistent with the antagonistic pleiotropy theory for the evolution of senescence. In further support of this argument, I find that a protein's connectivity is positively correlated with the number of traits it influences or its degree of pleiotropy, and further show that the average degree of pleiotropy is greatest for proteins associated with senescence. I explain these results with a simple mathematical model combining assumptions of the antagonistic pleiotropy theory for the evolution of senescence with data on network topology. These findings integrate molecular and evolutionary models of senescence, and should aid in the search for new ageing genes.

Keywords: senescence; networks; antagonistic pleiotropy; yeast; *Saccharomyces cerevisiae*

1. INTRODUCTION

Over the past half-century, evolutionary biologists have developed powerful explanatory models for why we age (Medawar 1952; Williams 1957) and have been relatively successful in predicting age-specific patterns of genetic variation for fitness traits (Rose & Charlesworth 1980; Hughes 1995; Charlesworth 2001). However, evolutionary theory has been less successful in predicting the specific classes of genes that should be associated with the ageing process. Theory suggests that senescence should be affected by very many genes with relatively small effects (but see Kowald & Kirkwood 1996), and that genes involved in senescence should not necessarily be similar across taxa (Rose 1991). In contrast with these evolutionary arguments, molecular geneticists have successfully identified a few genes and gene pathways that, when mutated, lead to dramatic increases in longevity (Lithgow *et al.* 1995; Hekimi & Guarente 2003; Tatar *et al.* 2003) and have shown that some of these pathways influence longevity across a taxonomically diverse range of model organisms, including yeast, worms, flies and mice (Bartke *et al.* 2003; Hekimi & Guarente 2003).

Until now, genetic studies of senescence have focused primarily on the search for single genes or relatively simple gene pathways. However, recent work is beginning to uncover gene networks associated with senescence (Murphy *et al.* 2003), and some studies have shown that gene and protein interaction networks can be used to explain detailed patterns of complex traits. For example, networks have been used to predict the developmental patterns of segment polarity in *Drosophila* (von Dassow *et al.* 2000; Albert & Othmer 2003) and tooth morphology in mammals (Salazar-Ciudad & Jernvall 2002).

In the past few years, researchers have put significant effort into defining the large-scale structure of these biological networks, especially in the yeast *Saccharomyces cerevisiae* (Jeong *et al.* 2001). The structure of the yeast

protein–protein interaction network can be described in part by its ‘degree distribution’, $p(k)$, where k , the protein’s ‘connectivity’, is the number of proteins with which each protein interacts. Connectivity for yeast protein networks and many other networks (Albert & Barabási 2002), is well described by a power law with an exponential term truncating the right-hand side of the distribution,

$$p(k) = Ck^{-\gamma}e^{-k/k_c}, \quad (1.1)$$

where γ is the characteristic exponent, k_c is the exponential cut-off parameter and C is a normalizing constant. In these networks, also known as ‘scale-free’ networks (Albert & Barabási 2002), most nodes are weakly connected, interacting with only one or two other nodes, whereas a few nodes are highly connected, with tens or even hundreds of interactions (Jeong *et al.* 2001; Albert & Barabási 2002). Studies have shown that the overall structure of scale-free networks makes them resistant to random damage (Albert *et al.* 2000; Wagner 2000) and that network structure can be used to predict biological function (von Dassow *et al.* 2000; Stelling *et al.* 2002). If we think about senescence as the gradual failure not simply of individual biological components, but rather of complex networks, then studies on the structure and function of networks may provide a useful framework to understand the genetic basis of senescence (Kowald & Kirkwood 1996).

Here, I analyse the yeast protein–protein interaction network to determine if genes associated with senescence are a non-random sample of all genes in terms of connectivity. I then ask whether the observed patterns of molecular interactions are consistent with Williams’s classic antagonistic pleiotropy theory for the evolution of senescence (Williams 1957). The antagonistic pleiotropy theory posits that senescence evolves because at least some early-acting beneficial alleles favoured by selection will have deleterious pleiotropic effects late in life, leading to an inevitable decline in late-age fitness. With this model

in mind, I hypothesize that if more highly connected proteins are more pleiotropic, these proteins will be most likely to evolve an association with senescence. I argue that an analysis of protein networks may enable us to better integrate evolutionary theories of senescence with molecular data, eventually leading to a predictive model for ageing genes.

2. MATERIAL AND METHODS

(a) Data

I obtained data on the number of protein interactions (k) for each of 3575 proteins within the yeast protein–protein interaction network from Fraser *et al.* (2002). This dataset, which represents over 60% of the yeast genome (Cliften *et al.* 2003), included a total of 13 231 pairwise interactions, excluding 538 self-protein interactions. I used these data to determine average connectivity for proteins associated with six quantitative traits in *S. cerevisiae* for which I was able to obtain lists of known mutants. These traits included replicative senescence (Kaeberlein *et al.* 2002), the cell cycle (Stevenson *et al.* 2001), reduced cell size (*whi* mutants) (Jorgensen *et al.* 2002), ultra-violet (UV) radiation sensitivity (Birrell *et al.* 2001), salt tolerance (de Jesus Ferreira *et al.* 2001) and DNA silencing (Smith *et al.* 1999). The last four traits are independent of senescence, although cell cycle is functionally related to senescence (Gershon & Gershon 2000), and so may show similar patterns (see Appendix A for a full list of genes). Where a protein in a trait dataset was not found in the full dataset of 3575 proteins, it was excluded from all further analysis.

My first aim was to determine if the distribution of connectivity for any of the six traits examined was significantly different from that of the full dataset. A simple model (see § 2b) predicts that proteins associated with senescence should have a higher mean and a lower variance for ln-transformed values of k . I first calculated the mean and variance for ln-transformed values of k ($\bar{k}$ and σ_k^2 , respectively) for the known yeast protein–protein interaction network (Fraser *et al.* 2002), and then used the same k -values to determine the mean and variance of ln(k) for each of the six traits. Values of k were log-transformed to better approximate normality.

In addition to values of k , I also determined the degree of pleiotropy for each protein using the online database at the Munich Information Center for Protein Sequences (MIPS) (see <http://mips.gsf.de>). I defined the degree of pleiotropy for each of the 3575 proteins in the network database as the number of functional classifications (e.g. nucleotide metabolism, signal transduction, stress response, etc.) listed in the ‘FunCat’ category of the MIPS database. In all, 2611 proteins had both k -values and a functional classification number equal to or greater than 1. To test for a significant relationship between these two parameters, I used linear regression on log-transformed values of k and degree of pleiotropy in JMP v. 5.0 (SAS Institute 1995).

Proteins involved in nuclear processes tend to have higher connectivity than those which are active outside of the nucleus (supplementary material in Yu *et al.* (2004)). Thus, proteins associated with senescence may have relatively high connectivity if they occur preferentially inside the nucleus. To test this possibility, I analysed 2774 proteins for which I had information on protein localization (Huh *et al.* 2003) and connectivity. I controlled for the effects of localization by comparing k in ageing-related proteins and all other proteins found within the nucleus, and for proteins not found in the nucleus.

(b) Model and statistical inference

I developed a model to analyse these data in the framework of the antagonistic pleiotropy theory for the evolution of senescence (Williams 1957). Williams’s model assumes that genes exist with early-acting beneficial effects and late-acting deleterious effects. Thus, we expect that the greater the degree of pleiotropy a gene has, the more likely that it evolves to influence senescence. A more pleiotropic gene is simply more likely to meet Williams’s criterion, exhibiting trade-offs between early-life and late-life characteristics.

The model is based on two assumptions. First, I assume that in any mutagenesis search for genes related to senescence, genes are sampled at random in terms of the connectedness (k) of their protein product, distributed as $p(k)$ (equation (1.1)). Second, I assume that proteins that interact with many other proteins at a molecular level (i.e. high k -values) are more likely to exhibit pleiotropic effects (Price & Schluter 1991). A test of this second assumption is provided below.

To test the hypothesis that the probability that a protein affects a particular trait increases with the protein’s connectivity, I first assume that the relative probability that a protein influences ageing increases with connectivity according to a sigmoidal function,

$$\theta(k) = \frac{e^{4sk}}{e^{4sk} + e^{4si}} \quad (2.1)$$

In equation (2.1), i is the inflection point of the curve and s is the slope of $\theta(k=i)$. The sigmoidal function has the desirable characteristics that it is monotonic and bounded by 0 and 1.

The expected frequency distribution of connectivity, $f(k)$, among proteins that influence the trait of interest is thus given by the product of equations (1.1) and (2.1), $f(k) = p(k) \times \theta(k)$ or

$$f(k) = \frac{C_2 e^{4sk} k^{-\gamma} e^{-k/k_c}}{(e^{4sk} + e^{4si})}, \quad (2.2)$$

where C_2 is a normalization constant.

If more highly connected proteins are more likely to be associated with senescence, then when we fit the distribution of connectivity for proteins associated with senescence to the model in equation (2.2), we should observe that the slope, $s > 0$, and the inflection point, $i \geq 1$.

For scale-free networks, the relationship between the mean and variance of ln(k) is positive and linear (data not shown), and the distribution of k is highly right-skewed. Numerical analysis of equation (2.2) reveals that for most of the parameter space in this model, if connectivity increases the likelihood that a gene is associated with a given trait, then the mean ln(k) for that trait will be greater and the variance lower compared with the full dataset (figure 1). This provides an explicit testable hypothesis for these datasets.

To determine if mean and variance for log-transformed values of k for a particular trait dataset of N proteins were significantly different from those expected by chance, I used a randomization procedure (Manly 1997). I sampled with replacement N proteins from the full dataset of 3575 proteins and calculated the mean and variance for ln(k) among the resampled dataset. I repeated this sampling procedure 15 000 times for each trait. I then determined whether the observed trait mean and variance for ln(k) was significantly different from the random samples. To correct for the positive mean-variance correlation between $\bar{k}$ and σ_k^2 , I determined the probability by calculating the $\beta\%$ confidence ellipse for the randomized data at which the observed data point intersected, with $P = 1 - \beta/100$.

Figure 1. Over most of the parameter space, a positive slope for $\theta(k)$ (i.e. $s > 0$) leads to an increase in mean (a) and a decrease in variance (b) of $\ln(k)$ relative to the complete dataset. The figures compare mean and variance for equation (2.2) (where $s > 0$ and $i = 0, 1$ or 2 : dotted, dashed and solid lines, respectively) with equation (1.1) ($s = 0$). The horizontal line indicates equal mean or variance for both equations. Results are generally consistent for varying values of i . If connectivity increases the likelihood that a protein is associated with a particular trait, the model predicts that the trait should show an increase in mean and decrease in variance for $\ln(k)$.

At the same time, we can determine directly if $s > 0$ by fitting equation (2.2) to each dataset. This fitting procedure consists of a two-part process. First, I determine the values of k_c and γ by fitting equation (1.1) to the entire dataset of 3575 proteins using maximum-likelihood estimation (MLE). I assume that the subsets of data have the same underlying values of k_c and γ as the full dataset. In the second step, I apply these predetermined values of k_c and γ to equation (2.2), and use MLE to determine the values of the slope and inflection point, s and i , respectively.

If connectivity influences the probability that a protein influences a particular trait, then we should find that incorporating $\theta(k)$ into equation (2.2) provides a significantly improved fit to the distribution of k in the trait of interest compared with equation (1.1) alone. To determine if s and i improved the fit significantly, I compared the two models in equations (1.1) and (2.2) using a likelihood ratio test. Under the likelihood ratio test, $\Delta LL = 2\{(-LL_{\text{equation (2.2)}}) - (-LL_{\text{equation (1.1)}})\}$, where LL is the log likelihood for the function, and ΔLL is distributed as χ^2 with d degrees of freedom, where d is the number of parameters by which the two nested functions differ.

Figure 2. Variance (σ_k^2) and mean ($\bar{k}$) for log-transformed values of protein connectedness. The small dots are estimates of σ_k^2 and $\bar{k}$ for 15 000 random samples of proteins, with sample size determined for the replicative senescence dataset (see text). The contour denotes the 99.999 97% confidence ellipse. The open circles denote the observed values of σ_k^2 and $\bar{k}$ for the traits indicated, with parameter values from table 1 (a, cell cycle; b, ageing; c, UV sensitivity; d, salt tolerance; e, DNA silencing; f, cell size). To increase clarity, the randomized estimates are shown only for replicative senescence. Confidence ellipses and p -values in table 1 for all other traits were estimated from separate randomizations with appropriate sample size.

To illustrate the patterns graphically, I determined the joint 50%, 75% and 95% confidence intervals (CI) for s and i numerically for each trait. This was done by calculating the values of s and i for which ΔLL was equal to $\chi^2_{1-\alpha}$ (where $\alpha = 0.5, 0.25$ or 0.05 for 50%, 75% and 95% CI, respectively).

3. RESULTS

In comparison with random samples of proteins, those proteins associated with senescence had a higher $\bar{k}$ and lower σ_k^2 ($n = 38$, $p = 0.000\ 03$; figure 2; table 1). Among the other five traits, the four traits that are independent of senescence did not show this pattern. Proteins associated with the cell cycle, which is very closely associated biologically with replicative senescence (Gershon & Gershon 2000), showed a relatively high connectivity ($p < 0.000\ 01$; table 1). All other traits had values of $\bar{k}$ and σ_k^2 that were not significantly different from random expectation ($p > 0.1$ in all cases; table 1).

The significantly higher mean and lower variance for $\ln(k)$ observed in proteins associated with senescence is unlikely to be a result of one or two outliers with very high k . We see a notable deficit of proteins with $k = 1$ or $k = 2$ in the senescence dataset (figure 3). In the full yeast dataset for 3575 proteins, 1557 (43.6%) have one or two interactions. In the dataset for genes associated with senescence, only five out of 38 (13.2%) have one or two interactions ($\chi^2_1 = 10.24$, $p = 0.0014$). A similar pattern is seen for cell cycle proteins ($k < 3$ for 15 out of 66

Table 1. Parameter estimates for protein datasets.

(The first p -value ‘(random)’ is the probability of observing a mean and variance of $\ln(k)$ ($\bar{k}'$ and $\sigma_{k'}^2$, respectively) for a given trait, based on a randomization test. The second p -value ‘(LL)’ is based on a likelihood ratio test to determine whether the probability that a randomly chosen protein is associated with a trait of interest increases with its connectivity (i.e. if the slope, s , of the function $\theta(k)$ is significantly greater than zero). The dataset for proteins associated with ageing was divided into mutations that have been shown to extend longevity (‘slow ageing’) and those that have decreased longevity (‘fast ageing’).)

trait	N	$\bar{k}'$	$\sigma_{k'}^2$	p -value (random)	i	s	χ^2 d.f.	p -value (LL)
ageing	38	1.970	1.069	0.00005	2.074	0.527	18.66	0.00008
slow ageing	16	1.926	0.768	0.004	2.051	5.199	14.29	0.000787
fast ageing	22	2.002	1.332	0.0025	3.789	0.123	8.16	0.0169
cell cycle	66	2.032	1.850	< 0.00001	3.760	0.082	17.1	0.00019
salt tolerance	26	1.336	1.193	0.69	0.948	3.627	0.74	0.691
silencing	10	1.767	1.172	0.29	1.728	0.445	2.83	0.24
UV sensitivity	30	1.704	1.463	0.11	1.428	0.250	3.89	0.143
cell size	16	1.411	1.009	0.55	1.051	3.647	2.92	0.282
all data	3575	1.289	1.348	n.a.	2.7×10^{-7}	4.6×10^{-4}	0.002	0.999

Figure 3. Frequency distribution of protein connectivity for 3575 yeast proteins (filled circles) and for 38 yeast proteins associated with replicative senescence (open circles). The solid line is the maximum-likelihood fit of equation (1.1) to the full dataset, with $\gamma = 1.055$ and $k_c = 29.21$. The dashed line is the fit of the pleiotropy model (equation (2.2)) to the 38 proteins associated with replicative senescence (see table 1 for parameter values).

proteins; $\chi^2_1 = 11.45$, $p = 0.0007$). The other traits, however, do not show a significant deficit of proteins with $k < 3$ (cell size: 5 out of 16; UV sensitivity: 9 out of 30; DNA silencing: 2 out of 10; salt tolerance: 10 out of 26; $p > 0.1$ in all cases).

The 38 genes associated with senescence were identified in numerous laboratories. If the analysis is limited to 22 genes that have been identified in a single laboratory (that of L. Guarente), we still find that $\bar{k}'$ is significantly higher than expected (15 000 randomizations, $p = 0.001$).

Fitting the full protein–protein interaction network to equation (1.1), I found likelihood estimates of $\gamma = 1.055$ and $k_c = 29.21$. If connectivity increases the likelihood that

a protein is associated with a trait, then we should see a significant effect of $\theta(k)$ in the model. For proteins associated with senescence, adding $\theta(k)$ into the simpler $p(k)$ model explained a significant amount of the variation (likelihood ratio test: $\chi^2_2 = 18.66$, $p = 0.000\ 089$; table 1; figure 2). Similarly, $\theta(k)$ was highly significant for the cell cycle mutants ($\chi^2_2 = 17.1$, $p = 0.000\ 19$). None of the other traits examined showed a significant effect of $\theta(k)$ (table 1).

This pattern is further illustrated by determining whether the slope, s , of $\theta(k)$ versus k is greater than zero. Figure 4 shows the joint MLE for slope s and the inflection point i . 50%, 75% and 95% CI are also plotted to illustrate the landscape surface for the joint estimates of s and i . Note that in figure 4, $s > 0$ at the 95% confidence level for replicative senescence and cell cycle but for no other trait, implying that the probability that a protein is associated with ageing or cell cycle increases with its connectivity.

A key assumption underlying the model is that connectivity is related to pleiotropy. In strong support of this assumption, I found a highly significant positive correlation between connectivity and pleiotropy across 2611 proteins ($F_{1,2609} = 119.6$, $p = 3.02 \times 10^{-27}$), albeit with a substantial amount of scatter ($r^2 = 0.043$; figure 5). In addition, I found that, similar to the results for connectivity, genes associated with senescence had a higher than expected degree of pleiotropy ($t_{2647} = 5.645$, $p < 0.0001$; figure 6).

Consistent with a previous study (Yu *et al.* 2004), proteins found within the nucleus had a significantly higher average connectivity (1.579 ± 0.035 , $\ln(k) \pm \text{s.e.m.}$) than non-nuclear proteins (1.251 ± 0.029 ; $t_{2772} = 63.74$, $p < 0.0001$). Using the MLE model described above, among proteins found inside the nucleus, those associated with ageing showed a significantly higher connectivity (2.086 ± 0.217), compared with all nuclear proteins ($s = 4.477$, $\chi^2_2 = 13.12$, $p[s > 0] = 0.0014$). Similarly, average connectivity among non-nuclear proteins (1.867 ± 0.298) was higher than all non-nuclear proteins, although the effect was marginally non-significant with a two-tailed probability ($s = 0.063$, $\chi^2_2 = 4.98$, $p[s > 0] = 0.087$).

Figure 4. Estimates of slope (s) and inflection point (i) for $\theta(k)$ for six traits in yeast: (a) ageing; (b) cell cycle; (c) UV sensitivity; (d) DNA silencing; (e) salt tolerance; and (f) cell size. For a given trait, if s is significantly greater than zero, then the probability that a randomly chosen protein is associated with that trait increases with connectivity. The figures show the maximum-likelihood estimates for s and i (marked with an open circle) and the 50% (blue), 75% (gold) and 95% (red) confidence ellipses. Note that for proteins associated with replicative senescence and cell cycle, the 95% confidence ellipse supports an estimate of $s > 0$. In none of the other four traits shown is s estimated to be greater than zero at the 95% confidence level.

4. DISCUSSION

I have shown here that proteins associated with senescence interact with significantly more other proteins than expected by chance. I propose several possible mechanisms that might account for this finding.

First, the pattern I observed is consistent with the antagonistic pleiotropy theory of senescence (Williams 1957). According to Williams's theory, senescence evolves because genes with deleterious effects on fitness late in life spread owing to their beneficial (i.e. antagonistic pleiotropic) effects at earlier ages. I suggest here that the more pleiotropic effects a gene has, then simply by chance the more likely it is that some of these effects will be found at different ages,

and will be of opposite sign in their effects on fitness. I assumed here that connectivity was a surrogate for pleiotropy, and indeed, I found that proteins with relatively high connectivity tended to have a higher degree of pleiotropy than expected, and that the average degree of pleiotropy was highest among proteins associated with senescence.

This argument assumes that the highly pleiotropic mutants that have been shown to extend longevity reduce other fitness traits. Although few studies have looked specifically for genes with antagonistic pleiotropic effects in yeast, there is evidence that such genes exist. For example, *sch9* mutants have a threefold increase in lifespan, but grow at a slower rate and produce smaller colonies (Longo & Finch 2003).

Figure 5. Relationship between the degree of pleiotropy (see text for definition) and connectivity for yeast proteins plotted on a log-log scale. (a) The size of the bubbles is proportional to the number of proteins with that value. (b) This illustrates the average and standard error of $\ln(k)$ for proteins with a given degree of pleiotropy.

Although the results described here are consistent with the antagonistic pleiotropy theory of ageing, other mechanisms could also account for the high connectivity among ageing genes. For example, the probability that a null mutation in yeast is lethal increases with the connectivity of the protein that is knocked out (Jeong *et al.* 2001). More generally, the phenotypic effect of a mutant may increase with its connectivity. If mutations with weak phenotypic effects are hard to detect, then mutagenesis assays could be biased towards proteins that are most highly connected. This bias may depend, however, on the accuracy with which traits can be measured. Some traits, such as cell size, can be measured with a high degree of accuracy. In this case, we should be able to identify mutants even if they have relatively small effects. By contrast, longevity is a difficult trait to measure, and is highly variable even within isogenic populations (Khazaeli *et al.* 1998). Thus, it may be especially hard to detect genes with small effects on longevity.

Weakly connected proteins may have only a small effect on longevity, but the scale-free distribution of protein interactions (figure 3) means that these proteins are the most common. In addition, the broad distribution of

Figure 6. Geometric mean (+ 1 s.e.m.) of number of functions (i.e. pleiotropy) per gene, by trait. Means for each trait were compared with the full dataset mean using a two-tailed *t*-test with unequal variances. Mean degree of pleiotropy for proteins associated with ageing and the cell cycle both had higher averages than the overall mean, as denoted by the asterisks (cell cycle: $t_{67.1} = 2.63$, $p = 0.0106$; ageing: $t_{37.8} = 4.42$, $p < 0.0001$).

connectivities may further magnify the potential difference in mutational effects among proteins. Although recent molecular studies suggest that senescence is influenced by a few genes with large effect, we may inadvertently miss many weakly connected genes with small phenotypic effects. These genes may turn out to be important for senescence when combined with one another in larger numbers. In addition, they may influence senescence by acting as modifiers of genes with major effects. Studies have demonstrated that such epistatic interactions can be an important factor in the genetic architecture of lifespan (Leips & Mackay 2002; Spencer *et al.* 2003). Proteins with high connectivity may affect ageing if they have a disproportionate number of epistatic interactions with other proteins, which appears to be true for yeast (Tong *et al.* 2004). Weakly connected proteins, whose effects may be difficult to detect, may be identified not by testing for their direct effects, but rather by looking for their footprint as modifiers of previously identified genes, using the existing network structure as our guide (e.g. Letovsky & Kasif 2003; Tong *et al.* 2004).

One concern with protein network connectivity studies is that there is a substantial false-positive rate in measures of protein-protein interactions, perhaps as high as 50% (von Mering *et al.* 2002; Sprinzak *et al.* 2003). These high error rates will add substantial noise to any study, but are unlikely to bias the results presented here. Given the high error variance, the findings are all the more striking. In addition, proteins associated with ageing may appear to have higher connectivity if they tend to be associated with some third, unexamined trait that leads to higher connectivity. I tested the possibility that proteins associated with senescence may have had higher connectivity as a result of being disproportionately found in the nucleus (Yu *et al.* 2004), where connectivity is higher. In fact,

these proteins were found equally inside and outside of the nucleus, and within each location I continued to see higher connectivity among proteins associated with senescence. Similarly, the effect of the evolutionary age on protein connectivity (Eisenberg & Levanon 2003) did not influence the patterns that I found here (D. E. L. Promislow, unpublished data).

Ultimately, we will need direct experimental tests to determine the specific mechanism—antagonistic pleiotropy, variation in the magnitude of mutational effects, or epistatic interactions—that accounts for the pattern I have observed here. Although the mechanism underlying my results remains to be determined, a useful prediction that arises from these results is simply that more highly connected proteins are more likely to be associated with ageing. For those looking to identify novel ageing genes, these results suggest that one should focus on proteins with relatively high connectivity and/or a high degree of pleiotropy.

There are additional ways in which a network perspective may inform the evolution of senescence. Recent studies have used network structure and function to predict pattern formation during early development (e.g. von Dassow *et al.* 2000; Salazar-Ciudad & Jernvall 2002; Albert & Othmer 2003), and to explain why networks are robust to damage (Albert *et al.* 2000; Wagner 2000; Bergman & Siegal 2003). From the perspective of the biology of senescence, we may benefit from a shift in focus from how networks function to how they fail. Studies on network robustness have considered why some networks are less likely to fail than others (Albert *et al.* 2000). However, we do not yet have an explicitly evolutionary theory of how network components evolve in biological systems, and how that evolutionary history may bear on the likelihood that specific components fail, even as the overall network continues to function reasonably well. Once we have identified the critical components within a network that play a central role in the ageing process, we can then concentrate our energy on identifying the genes or proteins that are active in those components, and determine how best to extend the lifespan of those particular components of the network.

The model described here was motivated by theories for

the evolution of senescence, but it also explained patterns seen in cell cycle mutants. The similarity in pattern between these two traits is not surprising, given that in yeast, the timing of the cell cycle is inextricably tied to replicative senescence (Gershon & Gershon 2000). Perhaps proteins associated with senescence have high connectivity simply because many of the known ageing mutants are also cell cycle mutants. Analysis of the ageing mutants by using FUNSPEC (<http://funspec.med.utoronto.ca>) shows that the 'cell cycle' functional categorization is highly over-represented among proteins associated with replicative senescence (22 out of 38 proteins, $p < 3.52 \times 10^{-13}$). The functional relationship between cellular senescence and senescence at the organismal or demographic level remains a matter of debate (Hornsby 2002). As protein network data are now becoming available for multicellular organisms such as flies (see Giot *et al.* 2003) and nematode worms (see Li *et al.* 2004), we should be able to employ techniques similar to those used here to determine whether the relationship between connectivity and ageing mutations exists independent of any relationship between genes that effect senescence and those that effect cell division.

Although further work may be needed to determine just why we observe the striking pattern seen here for genes associated with replicative senescence, results from this study suggest a way to integrate evolutionary and molecular models of senescence, and may help to guide the search for additional genes that influence longevity.

J. Burger, V. Corby, M. Doebeli, K. Fedorka, H. Fraser, D. Gomulkiewicz, D. Hoyt, L. Keshet, C.-H. Kuo, A. Keyser, N. Newland, A. Norris, S. Otto, A. Peters, E. Promislow, H. Qin, B. Starzomski, M. Whitlock and N. Wolff provided comments during the development of this work, and two anonymous reviewers made suggestions on the manuscript. H. Fraser provided open access to his yeast network data, E. Eisenberg provided data on protein ages, and E. Promislow provided programming assistance. A significant portion of this work was performed in the Department of Zoology at the University of British Columbia, which provided invaluable logistical support. This work was supported by funding from NSF (DEB 0214022), NIH (AG021298) and the Ellison Medical Foundation.

APPENDIX A

Gene names and protein connectivity (k) for UV radiation sensitivity (Birrell *et al.* 2001), cell size (Jorgensen *et al.* 2002), senescence (including both decreased and increased lifespan) (Kaeberlein *et al.* 2002), cell cycle (Stevenson *et al.* 2001), salt tolerance (de Jesus Ferreira *et al.* 2001) and DNA silencing (Smith *et al.* 1999).

UV sensitivity	k	cell size	k	short lifespan	k	long lifespan	k
YBR098W	2	YDR156W	1	YAL021C	28	YBL103C	3
YCR066W	4	YDR457W	2	YBR275C	2	YDL042C	20
YDL059C	30	YER020W	12	YDL022W	1	YDR110W	10
YDL101C	28	YGL003C	6	YDR217C	4	YDR293C	3
YDR004W	1	YGL016W	3	YDR227W	11	YER020W	12
YDR076W	4	YGR111W	1	YDR422C	8	YGL178W	4
YDR217C	4	YHL011C	3	YDR477W	25	YHL003C	2
YDR386W	34	YHR034C	2	YER095W	41	YIL053W	5
YDR419W	1	YHR066W	48	YER111C	3	YJL005W	14
YEL037C	28	YHR158C	9	YGL208W	7	YJL194W	4
YER095W	41	YJL187C	10	YGR132C	1	YKL109W	3
YER162C	4	YKR020W	3	YHR008C	2	YLR310C	11

(Continued.)

Appendix A. (*Continued.*)

YER173W	10	YLR403W	8	YJR104C	8	YMR273C	16
YGL087C	4	YMR216C	4	YLR113W	7	YNL098C	7
YGL163C	4	YNL227C	3	YLR442C	16	YNL330C	45
YGL175C	2	YNL248C	2	YML032C	8	YOR101W	3
YGR258C	1			YML109W	26		
YHR025W	2			YMR106C	63		
YHR134W	1			YMR190C	3		
YJL092W	7			YMR284W	10		
YJL124C	19			YOR358W	3		
YJR052W	4			YPL254W	15		
YLR288C	11						
YML095C	16						
YMR201C	6						
YNL107W	2						
YOR290C	9						
YOR346W	1						
YOR368W	4						
YPL022W	37						
<hr/>							
cell cycle	<i>k</i>	cell cycle	<i>k</i>	salt tolerance	<i>k</i>	DNA silencing	<i>k</i>
YBR031W	102	YLR220W	1	YDR040C	1	YBR275C	2
YBR079C	29	YML016C	3	YDR166C	1	YBR278W	13
YBR085W	10	YML068W	1	YDR335W	7	YCR084C	12
YBR089C-A	23	YMR043W	1	YEL061C	1	YDR227W	11
YBR143C	34	YMR072W	1	YER053-A	2	YDR439W	1
YBR221C	27	YMR142C	42	YGL019W	24	YJR140C	3
YCL011C	30	YMR186W	12	YGL020C	1	YMR263W	3
YCL028W	15	YMR273C	16	YGL060W	3	YNL102W	6
YDL084W	3	YMR308C	62	YGL061C	5	YNL330C	45
YDL134C	15	YNL061W	66	YGR102C	2	YOL006C	11
YDL137W	22	YNL076W	1	YIL034C	6	YOR217W	3
YDL155W	4	YNL126W	5	YJL102W	1		
YDL192W	7	YNL141W	1	YJL165C	1		
YDR002W	4	YNL233W	4	YKL182W	18		
YDR113C	4	YOR063W	50	YLR295C	7		
YDR381W	16	YOR212W	20	YLR399C	6		
YEL009C	3	YOR232W	13	YLR417W	2		
YER016W	8	YOR326W	26	YML006C	3		
YFL005W	4	YPL020C	1	YML057W	10		
YFL010C	2	YPL078C	5	YNL307C	6		
YFL037W	44	YPL127C	2	YNR029C	2		
YFL039C	35	YPL145C	1	YOR039W	35		
YFR028C	25	YPL203W	9	YOR181W	29		
YGL003C	6	YPR119W	12	YOR275C	3		
YGL116W	9	YPR029C	4				
YGL190C	45	YPR032W	3				
YGR082W	1						
YGR162W	59						
YHR086W	37						
YHR129C	6						
YJL164C	8						
YJR053W	5						
YKL009W	6						
YKL042W	1						
YKL148C	1						
YKL193C	17						
YKL195W	1						
YKR002W	28						
YLR026C	6						
YLR153C	4						
YLR167W	1						
YLR176C	19						

REFERENCES

- Albert, R. & Barabási, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97.
- Albert, R. & Othmer, H. G. 2003 The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* **223**, 1–18.
- Albert, R., Jeong, H. & Barabási, A. L. 2000 Error and attack tolerance of complex networks. *Nature* **406**, 378–382.
- Bartke, A., Chandrashekar, V., Dominici, F., Turyn, D., Kinney, B., Steger, R. & Kopchick, J. J. 2003 Insulin-like growth factor 1 (IGF-1) and aging: controversies and new insights. *Biogerontology* **4**, 1–8.
- Bergman, A. & Siegal, M. L. 2003 Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**, 549–552.
- Birrell, G. W., Giaever, G., Chu, A. M., Davis, R. W. & Brown, J. M. 2001 A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl Acad. Sci. USA* **98**, 12 608–12 613.
- Charlesworth, B. 2001 Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation accumulation theory of aging. *J. Theor. Biol.* **210**, 47–65.
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A. & Johnston, M. 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76.
- de Jesus Ferreira, M. C., Bao, X., Laize, V. & Hohmann, S. 2001 Transposon mutagenesis reveals novel loci affecting tolerance to salt stress and growth at low temperature. *Curr. Genet.* **40**, 27–39.
- Eisenberg, E. & Levanon, E. Y. 2003 Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **91**, 138 701, 1–4.
- Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. 2002 Evolutionary rate in the protein interaction network. *Science* **296**, 750–752.
- Gershon, H. & Gershon, D. 2000 The budding yeast, *Saccharomyces cerevisiae*, as a model for aging research: a critical review. *Mech. Ageing Dev.* **120**, 1–22.
- Giot, L. (and 48 others) 2003 A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736.
- Hekimi, S. & Guarente, L. 2003 Genetics and the specificity of the aging process. *Science* **299**, 1351–1354.
- Hornsby, P. J. 2002 Cellular senescence and tissue aging *in vivo*. *J. Gerontol. A. Biol. Sci. Med. Sci.* **57**, B251–B256.
- Hughes, K. A. 1995 The evolutionary genetics of male life-history characters in *Drosophila melanogaster*. *Evolution* **49**, 521–537.
- Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O'Shea, E. K. 2003 Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691.
- Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. 2001 Lethality and centrality in protein networks. *Nature* **411**, 41–42.
- Jorgensen, P., Nishikawa, J. L., Breitkreutz, B. J. & Tyers, M. 2002 Systematic identification of pathways that couple cell growth and division in yeast. *Science* **297**, 395–400.
- Kaeberlein, M., Jegalian, B. & McVey, M. 2002 AGEID: a database of aging genes and interventions. *Mech. Ageing Dev.* **123**, 1115–1119.
- Khazaeli, A. A., Pletcher, S. D. & Curtsinger, J. W. 1998 The fractionation experiment: reducing heterogeneity to investigate age-specific mortality in *Drosophila*. *Mech. Ageing Dev.* **105**, 301–317.
- Kowald, A. & Kirkwood, T. B. 1996 A network theory of ageing: the interactions of defective mitochondria, aberrant proteins, free radicals and scavengers in the ageing process. *Mutat. Res.* **316**, 209–236.
- Leips, J. & Mackay, T. F. 2002 The complex genetic architecture of *Drosophila* lifespan. *Exp. Aging Res.* **28**, 361–390.
- Letovsky, S. & Kasif, S. 2003 Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics* **19**(Suppl. 1), I197–I204.
- Li, S. (and 47 others) 2004 A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543.
- Lithgow, G. J., White, T. M., Melov, S. & Johnson, T. E. 1995 Thermotolerance and extended lifespan conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92**, 7540–7544.
- Longo, V. D. & Finch, C. E. 2003 Evolutionary medicine: from dwarf model systems to healthy centenarians? *Science* **299**, 1342–1346.
- Manly, B. F. J. 1997 *Randomization, bootstrap and Monte Carlo methods in biology*. New York: Chapman & Hall.
- Medawar, P. B. 1952 *An unsolved problem in biology*. London: H. K. Lewis.
- Murphy, C. T., McCarroll, S. A., Bargmann, C. I., Fraser, A., Kamath, R. S., Ahringer, J., Li, H. & Kenyon, C. 2003 Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283.
- Price, T. & Schluter, D. 1991 On the low heritability of life-history traits. *Evolution* **45**, 853–861.
- Rose, M. R. 1991 *Evolutionary biology of aging*. Oxford University Press.
- Rose, M. R. & Charlesworth, B. 1980 A test of evolutionary theories of senescence. *Nature* **287**, 141–142.
- Salazar-Ciudad, I. & Jernvall, J. 2002 A gene network model accounting for development and evolution of mammalian teeth. *Proc. Natl Acad. Sci. USA* **99**, 8116–8120.
- SAS Institute 1995 *JMP*. Cary, NC: SAS Institute Inc.
- Smith, J. S., Caputo, E. & Boeke, J. D. 1999 A genetic screen for ribosomal DNA silencing defects identifies multiple DNA replication and chromatin-modulating factors. *Mol. Cell Biol.* **19**, 3184–3197.
- Spencer, C. C., Howell, C. E., Wright, A. R. & Promislow, D. E. L. 2003 Testing an 'aging gene' in long-lived *Drosophila* strains: increased longevity depends on sex and genetic background. *Aging Cell* **2**, 123–130.
- Sprinzak, E., Sattath, S. & Margalit, H. 2003 How reliable are experimental protein–protein interaction data? *J. Mol. Biol.* **327**, 919–923.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. 2002 Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193.
- Stevenson, L. F., Kennedy, B. K. & Harlow, E. 2001 A large-scale overexpression screen in *Saccharomyces cerevisiae* identifies previously uncharacterized cell cycle genes. *Proc. Natl Acad. Sci. USA* **98**, 3946–3951.
- Tatar, M., Bartke, A. & Antebi, A. 2003 The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346–1351.
- Tong, A. H. (and 49 others) 2004 Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813.
- von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. 2000 The segment polarity network is a robust developmental module. *Nature* **406**, 188–192.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. 2002 Comparative assessment of large-scale datasets of protein–protein interactions. *Nature* **417**, 399–403.

- Wagner, A. 2000 Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361.
- Williams, G. C. 1957 Pleiotropy, natural selection, and the evolution of senescence. *Evolution* **11**, 398–411.
- Yu, H., Zhu, X., Greenbaum, D., Karro, J. & Gerstein, M. 2004 TOPNET: a tool for comparing biological sub-networks,

correlating protein properties with topological statistics. *Nucleic Acids Res.* **32**, 328–337.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.