GenAge: a genomic and proteomic network map of human ageing

João Pedro de Magalhães*, Olivier Toussaint

Research Unit on Cellular Biology, University of Namur (FUNDP), B-5000 Namur, Belgium

Accepted 6 July 2004

Available online 14 July 2004

Edited by Robert B. Russell

Abstract The aim of this work was to provide an overview of the genetics of human ageing to gain novel insights about the mechanisms involved. By incorporating findings from model organisms to humans, such as mutations that either delay or accelerate ageing in mice, we constructed the gene networks previously related to ageing: namely, the network related to DNA metabolism and the network involving the GH/IGF-1 axis. Gathering data about the interacting partners of these proteins allowed us to suggest the involvement in ageing of a number of proteins through a "guilt-by-association" methodology. To organize our data, we developed the first curated database of genes related to human ageing: GenAge. With over 200 entries, GenAge may serve as a reference database of genes related to human ageing. Moreover, we rendered the first proteomic network map of human ageing, which suggests a relationship between the genetics of development and the genetics of ageing. Our work serves as a framework upon which a systems-biology understanding of ageing can be developed. GenAge is freely available for academic purposes at: http://genomics.senescence.info/genes/.

© 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

Keywords: Gene network; Genetics; Proteomics; Senescence

1. Introduction

Despite its general interest not only to the academic community but to medicine and the public in general, human ageing is a poorly understood biological problem [1]. Findings in model organisms have shown that genes can regulate ageing [2,3] and so ageing clearly has a strong genetic component [4]. Yet since studying a process with the duration of ageing in humans is nearly impossible, it is unknown how these genes fit together and impact on human ageing. The only few genes shown to influence human ageing are the genes responsible for the so-called "progeroid syndromes" [5]. Therefore, partly because of technical difficulties in studying ageing in humans, partly because of the complexity of the ageing process, the genetics of human ageing remain largely a mystery.

The aim of this work is to create a comprehensive overview of the genetics of human ageing by incorporating findings from model organisms to humans. Namely, mutations that either delay [6] or accelerate [7] ageing in mice offer clues on the genetic mechanisms of mammalian ageing. Since discriminating causes from effects of ageing is troublesome, genetic interventions that alter the rate of ageing provide some of the few clear hints on the genetics of ageing. Our rationale is that by integrating these results, we may gain novel insights about the mechanisms involved. In addition, by gathering data about the interacting partners of these proteins, it may be possible to find additional clues about the players involved in these pathways. Eventually, we wanted to render the first genomic and proteomic network maps of human ageing to serve as a framework upon which a systems-biology understanding of ageing can be developed. As a result of our efforts, we developed the first curated database of genes related to human ageing: GenAge, which is freely available for academic purposes at: http://genomics.senescence.info/genes/.

2. Materials and methods

2.1. Gene selection

Great caution was taken to ensure that the selection of GenAge entries was as unbiased as possible. Initially, the few genes that appear to modulate ageing in humans, such as the progeroid genes [5], were selected, as were those that appear to influence mammalian ageing [8]. This list of genes influencing mammalian ageing is available online (http://genomics.senescence.info/genes/clues.html). Importantly, GenAge is not a database of health-promoting genes but rather a database of genes related to the fundamental human ageing process, as advocated before [1,4]. Consequently, care was taken to discriminate between genes that affect ageing and those that extend longevity by enhancing overall health.

In a second phase, a number of entries were selected in a "guilt-by-association" methodology: these consisted of proteins found highly associated with proteins or pathways previously directly linked with ageing. A functional clustering of pathways involved was also derived to identify the pathways of interest (http://genomics.senescence.info/genes/function.html), in line with previous reports [2].

Lastly, data from non-mammalian models was also employed. Since GenAge only includes human genes, many entries in GenAge represent human homologues of genes shown to affect ageing in model organisms. These were derived from a variety of sources such as AGEID [9], as done before [2]. Yet given the difficulty in extrapolating results from model organisms to human biology [10], GenAge is not a collection of genes identified in model organisms, but an integrative database that places each entry in the perspective of human biology. An extensive literature review, present in GenAge, was performed to assess whether genes found to affect ageing in model organisms may also relate to human ageing. If applicable, the following criteria were followed: (1) the influence of the gene in the model organism's ageing process; (2) the degree of conservation of the gene and protein in humans (e.g., the presence of 1:1 orthologues); (3) information on the phenotype resulting from polymorphisms or mutations in the human homologue; and (4) effects on ageing of the genetic manipulation – e.g., overexpression or disruption – of the gene's product(s) in mice. Given the subjective nature of the methodology, an explanation as to why the gene was selected is present in each GenAge entry (also see http://genomics.senescence.info/genes/allgenes.php), which also allows users

* Corresponding author. Present address: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Room 238, Boston, MA 02115, USA. Fax: +1-617-432-6513.
E-mail address: jp@senescence.info (J.P. de Magalhães).

0014-5793/$22.00 © 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2004.07.006

to discriminate between genes that may influence human ageing in different ways.

2.2. Software and hardware

The GenAge database is implemented using the MySQL 4.0.15 relational database management system. The query form Web pages are created dynamically in PHP or CGI scripts written in the Perl programming language. PathwayAssist 2.5 (Stratagene, La Jolla, CA) and InterViewer 3.7 (Inha University, WI Lab, South Korea) layout algorithms were used to analyse and visualise the protein–protein interaction networks. To automate Medline searches and perform text data-mining of abstracts, we used the Perl programming language; our programs are freely available online for non-commercial purposes (http://genomics.senescence.info/software/). Protein–protein interactions were obtained from the Human Protein Reference Database [11] and from both automated and manual Medline searches.

3. Results and discussion

Genes that appear to modulate mammalian ageing, including genes that may accelerate ageing in humans [5], served as basis for our work. The full list of genes as well as the respective references is available online (http://genomics.senescence.info/genes/clues.html), as previously published [8]. In line with previous efforts [2], we grouped genes shown to affect ageing in model organisms into functional clusters. We then attempted to reconstruct the gene networks influencing ageing by focusing on specific pathways. The rationale behind constructing maps of protein–protein interactions is that by taking a top-down view we may gain a better perspective of how networks work [12].

3.1. Gene networks influencing ageing

Given that most genes extending life span in mice appear to affect the GH/IGF-1 axis [2,13], we wanted to reconstruct the gene network involved. Genes belonging to this pathway and shown to influence ageing in mice served as our first selection: IGF1R, GH1, IGF1, GHR, GHRHR, PIT1, and PROP1 [6,8]. Protein–protein interactions for these proteins were obtained from the Human Protein Reference Database [11] as well as automated searches of Medline using the software we developed (see Section 2). This allowed us to suggest the involvement of a number of proteins as "guilt-by-association" by choosing the interaction partners of the proteins under study with the higher connectivity to the GH/IGF-1 pathway. "Guilt-by-association" genes included, not surprisingly, a number of proteins previously associated with insulin signalling, but also the STAT family, JUN, and two proteins involved in lipid metabolism: NR3C1 and LRP2. Using these proteins we could then derive a proteomic view of the GH/IGF-1 axis, including proteins shown to influence ageing and a number of additional proteins identified through our "guilt-by-association" methodology (Fig. 1).

The signal transduction from the transcription factor PROP1, passing through PIT1, to the key hormones GH1, IGF-1, and INS is represented. Interestingly, the signal

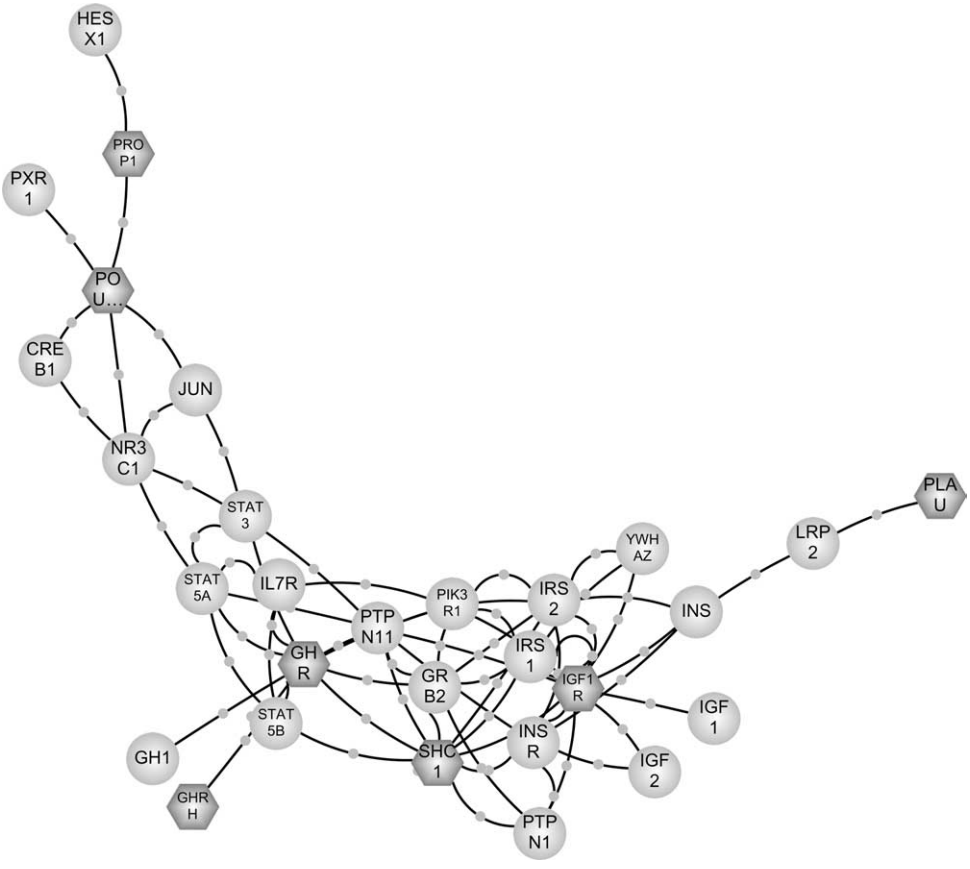

Fig. 1. Gene network of proteins thought to influence mammalian ageing whose effect may derive from alterations of the GH/IGF-1 axis. Proteins directly linked to ageing in mice (hexagons) as well as proteins associated with these (circles) are represented. Lines represent published protein–protein interactions as they appear in GenAge. Layout computed using the PathwayAssist software 2.5 (Stratagene, La Jolla, CA).

transduction affecting the GH/IGF-1 axis could involve JUN and the STAT family of signalling proteins. Moreover, SHC1 also appears to play a pivotal role in the pathway, which is intriguing since $p66^{shc}$, one of SHC1's splice variants, has been related to ageing in mice [14].

While the large majority of genes extending life span in mice are related to the GH/IGF-1 axis, the large majority of genes related to an accelerated ageing phenotype in mammals appear to be related to DNA metabolism [8]. Consequently, we also wanted to derive the gene network for these genes. The major difference when compared to the study of the GH/IGF-1 pathway is that the DNA metabolism network includes both genes related to ageing in mice and genes related to ageing in humans. Thus, care was taken to give more importance to genes related to human ageing. For instance, protein–protein interactions up to two degrees of freedom were investigated for proteins related to human ageing, while only one degree of freedom was used for proteins related to ageing in mice. Contrary to the GH/IGF-1 gene network, the connection between the different DNA metabolism proteins suggested to affect ageing is not so obvious (Fig. 2). Not surprisingly, since some of these proteins have been related to human ageing directly, we selected many more proteins in our "guilt-by-association" strategy. These represent possible players in the pathways involved and include PCNA, PRKCA, TP53, and ERCC6, which showed the higher connectivity.

It is also possible to analyse such pathways to find new controlling nodes. For example, it is possible to quantify the influence of each gene on ageing, either manually or by the selection process, and then attempt to find new regulatory nodes. Based on the gene networks we constructed, which obviously represent the current knowledge on these proteins and thus may be inaccurate, a few proteins appear to play pivotal roles. For instance, SHC1 appears to be pivotal in connecting the different aspects of the GH/IGF-1 pathway, while TP53 and WRN also emerge as crucial in the DNA metabolism network. Of course, each of these models is subject to a variety of interpretations. For example, if one sees ageing as an accumulation of damage, then DNA repair pathways may allow researchers to choose new targets of study or even predict ways to optimize DNA repair in animals to delay ageing and cancer.

3.2. GenAge: a curated database of genes related to human ageing

To organize our data and compile over 1000 bibliographical references collected to support our results, we developed the first curated database of genes related to human ageing: GenAge. GenAge is a relational database featuring methods to search and analyse genes or clusters of genes, study networks, and derive pathways. For instance, it is possible to search genes related to a certain function or present in a given cellular

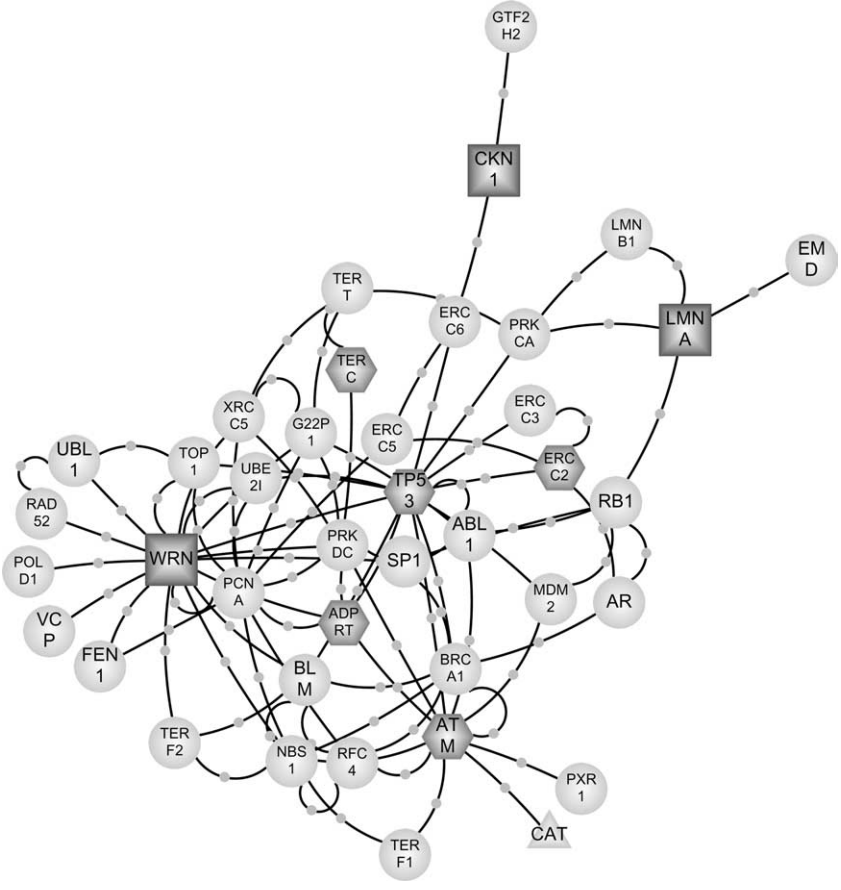

Fig. 2. Gene network of proteins thought to influence mammalian ageing whose function has been linked to DNA metabolism. Proteins directly linked to ageing in humans (squares) or mice (hexagons) as well as proteins associated with these (circles) are represented; one protein (CAT) associated with ageing in a non-mammalian organism is also present (triangle). Lines represent published protein–protein interactions as they appear in GenAge. Layout computed using the PathwayAssist software 2.5 (Stratagene, La Jolla, CA).

organelle, allowing users to analyse the genetic network of their choice and find novel relations between the genes involved. It is also possible to seek genes related to common pathways through, for example, protein–protein interactions and a number of visualization tools are available.

In GenAge, we incorporated genes found to affect ageing in humans and mice and the genes we identified through our "guilt-by-association" method. We also included a number of genes for which experimental evidence is contradictory in their association with ageing. For instance, some genes found to affect ageing in model organisms but unconfirmed in mammals were selected and it is possible for users to discriminate between different levels of selection (see Section 2). Through this list of genes, we were able to derive the first global view of the protein interaction network related to human ageing (Fig. 3). Of course, this global view gives us only a general impression of the performance of the network and it will require graph-theoretic methods for its analysis. Moreover, many putative genes are included with the added complication that many genes are frequently expressed in distinct organs or tissues. Even so, our results give insights into how the mechanisms of ageing fit together not as a collection of genes but like a mechanistic process where changes in one aspect of the cellular machinery affect cellular proliferation and body homeostasis. Although expected, it is gratifying to observe the topology of the network with three large functional clusters: growth and development on one end, DNA metabolism on the other end, and signalling pathways in the middle. It remains a challenge, however, to assess how the combination of all these factors explains the age-related loss of function and increase in vulnerability which we call ageing.

Probably, the most striking finding from our protein interaction map is the apparent influence on human ageing of genes involved in development. The suggestion that genes regulating development may also regulate the pace of ageing is not new [15]. Although premature, our results support such a connection between the genetic mechanisms regulating development and ageing. Links between proteins affecting ageing in model

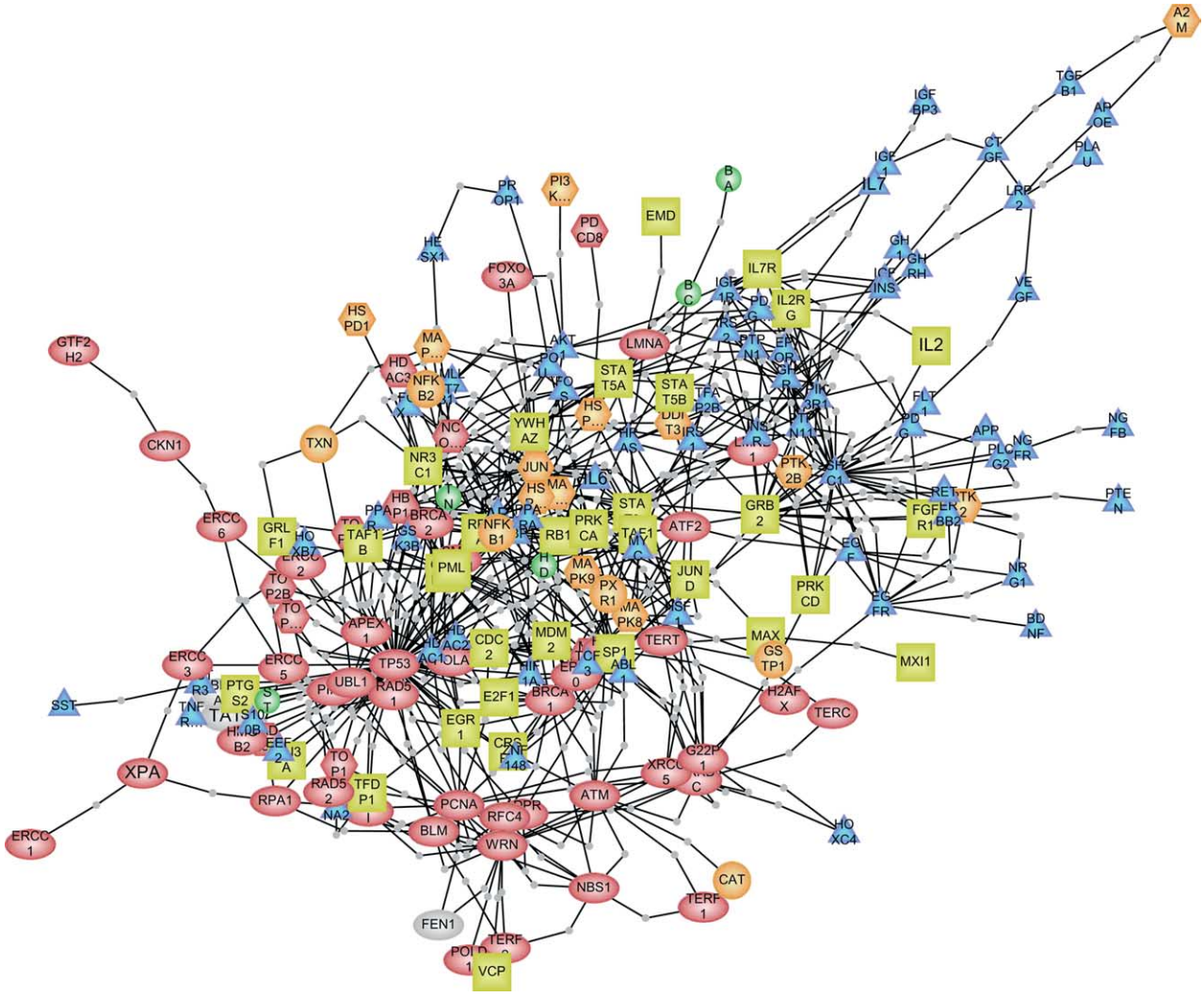

Fig. 3. Global view of the protein interaction network in GenAge build 6 (22/05/2004). All entries in GenAge for which there are reported protein–protein interactions were selected. Lines represent published protein–protein interactions as they appear in GenAge. Colour indicates the primary function of each protein as deducted from the literature: red, DNA repair and replication (oval), and DNA condensation (hexagon); blue, growth and development; yellow, transcriptional regulation and signalling; orange, redox and oxidative regulation (hexagon) and stress response (oval); green, apoptosis; grey, unknown or other. Layout computed using the PathwayAssist software 2.5 (Stratagene, La Jolla, CA).

organisms and proteins involved in development have been reported such as the association between $p66^{shc}$ and the forkhead family [16] or the apparent connection between the GH/IGF-1 pathway and development [13]. From GenAge, we found consistent examples of this connection between development and ageing, either directly or using our "guilt-by-association" method. For instance, two of the HOX genes have been included in GenAge due to their association with pathways that appear to be involved in ageing: HOXB7 and HOXC4. In fact, the possible role of HOX genes in ageing has been proposed before [17]. One hypothesis is that the cascade of events that regulates ontogeny fades away after sexual maturity resulting in ageing. This would explain as to why the mammalian ageing process appears as if timed by some unknown mechanism [4].

3.3. Concluding remarks

Global and local networks such as those presented herein provide a framework for researchers to investigate specific signalling and regulatory pathways involved in human ageing. They allow us to suggest other genes as possible players in ageing through a "guilt-by-association" methodology and reveal pivotal nodes in the pathways involved. For example, in yeast pathways highly connected proteins with a central role in the network's architecture are more likely to be essential than proteins with only a small number of links to other proteins [18]. While global maps (Fig. 3) will require a much more detailed analysis, we may obtain testable biological insights if we focus on smaller and more specialized subsystems [12]. Using a system-biology approach to the study of human ageing, local networks such as those derived for the GH/IGF-1 axis (Fig. 1) may be used to design experiments in animal models. Given the large number of age-related changes and players involved in ageing, the insights provided by GenAge are a powerful first step towards understanding the genomic and proteomic aspects of human ageing.

In the post-genome era, research into human ageing is made difficult by the lack of a coherent view of the genetic mechanisms of ageing [8]. GenAge offers a powerful new set of tools to understand the genetics of human ageing and provides the first glimpse of how the genome affects ageing. In addition, GenAge may serve as a reference database of genes related to human ageing. At present there are over 200 manually curated entries in GenAge (http://genomics.senescence.info/genes/allgenes.php), each featuring a combination of automatically extracted data – e.g., protein, open reading frame, and promoter sequences – and manually curated information – e.g., a description of the gene's relevance to human ageing and a selection of literature references. Hyperlinks in each entry also point the user to a variety of additional sources of information for a total of roughly 2000 hyperlinks.

At present, GenAge is mostly derived from genetic perturbations in animal models or inferred from human pathologies. One future prospect is to include entries showing differential expression during ageing. Yet interpreting differently expressed genes during ageing is not straightforward [8,19], particularly since our aim in GenAge is to identify genes that could play a causative role in human ageing. Clearly, it is much more difficult to determine cause versus effect in gene expression experiments than in mouse knockouts [8]. Even so, it is our goal to incorporate genes differently expressed with age and attempt to understand the regulatory basis of those changes. Integrating protein–protein interactions with gene expression, as well as other types of data, will allow us a better understanding of the network topology of human ageing.

GenAge is located at the Human Ageing Genomic Resources, which also include data-mining tools, software, and AnAge, a database of, where available, species longevity, phylogeny, metabolic rate, and ageing rate. Given that identifying the genes that determine the different rates of ageing amongst organisms is a major endeavour [2,4], AnAge is a unique resource in the comparative biology of ageing with over 2000 organisms listed. GenAge is freely available for academic purposes (http://genomics.senescence.info/genes/) at the Human Ageing Genomic Resources (http://genomics.senescence.info).

Acknowledgements: J.P. de Magalhães is supported by NIH-NHGRI CEGS grant to George Church. He thanks the FCT, Portugal, and the European Social Fund, III Quadro Comunitário de Apoio, for research support. O. Toussaint is a Research Associate from the FNRS, Belgium. Many thanks to Fabian Bastin, Jason Stajich, Joana Costa, Domingos Magalhães, Matt Kaeberlain, George Church, and everyone of the Linux and Perl/Bioperl communities for their invaluable assistance. Further thanks to the many biologists and researchers who have contributed with their knowledge to establish AnAge and GenAge.

References

[1] Hayflick, L. (2000) Nature 408, 267–269.
[2] Butler, R.N. et al. (2003) J. Gerontol. A 58, 581–584.
[3] Hekimi, S. and Guarente, L. (2003) Science 299, 1351–1354.
[4] de Magalhaes, J.P. (2003) Biogerontology 4, 119–120.
[5] Martin, G.M. and Oshima, J. (2000) Nature 408, 263–266.
[6] Liang, H., Masoro, E.J., Nelson, J.F., Strong, R., McMahan, C.A. and Richardson, A. (2003) Exp. Gerontol. 38, 1353–1364.
[7] Warner, H.R. and Sierra, F. (2003) Mech. Ageing Dev. 124, 581–587.
[8] de Magalhaes, J.P. and Toussaint, O. (2004) Ageing Res. Rev. 3, 125–141.
[9] Kaeberlein, M., Jegalian, B. and McVey, M. (2002) Mech. Ageing Dev. 123, 1115–1119.
[10] Pound, P., Ebrahim, S., Sandercock, P., Bracken, M.B. and Roberts, I. (2004) BMJ 328, 514–517.
[11] Peri, S. et al. (2003) Genome Res. 13, 2363–2371.
[12] Bray, D. (2003) Science 301, 1864–1865.
[13] Bartke, A., Coschigano, K., Kopchick, J., Chandrashekar, V., Mattison, J., Kinney, B. and Hauck, S. (2001) J. Gerontol. A 56, B340–B349.
[14] Migliaccio, E., Giorgio, M., Mele, S., Pelicci, G., Reboldi, P., Pandolfi, P.P., Lanfrancone, L. and Pelicci, P.G. (1999) Nature 402, 309–313.
[15] Martin, G.M. (1978) Birth Defects Orig. Artic. Ser. 14, 5–39.
[16] Nemoto, S. and Finkel, T. (2002) Science 295, 2450–2452.
[17] Venkataraman, K. and Futerman, A.H. (2002) FEBS Lett. 528, 3–4.
[18] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Nature 411, 41–42.
[19] Miller, R.A., Galecki, A. and Shmookler-Reis, R.J. (2001) J. Gerontol. A 56, B52–B57.