

Predicting Aging-Genes in *Drosophila Melanogaster* by Integrating Network Topological Features and Functional Categories

Yan-Hui Li, Institute of Cardiovascular Sciences and Key Laboratory of Molecular Cardiovascular Sciences, Ministry of Education, Peking University Health Science Center, Beijing, China

Jian-Hui Li, Computer Network Information Center Chinese Academy of Sciences, Beijing, China

Xin Song, Computer Network Information Center Chinese Academy of Sciences, Beijing, China

Kai Feng, Computer Network Information Center Chinese Academy of Sciences, Beijing, China

Yuan-Chun Zhou, Computer Network Information Center Chinese Academy of Sciences, Beijing, China

ABSTRACT

*An important task of aging research is to find genes that regulate lifespan. Wet-lab identification of aging genes is tedious and labor-intensive activity. Developing an algorithm to predict aging genes will be greatly helpful. In this paper, we systematically analyzed topological features of proteins encoded by *Drosophila melanogaster* aging genes versus those encoded by non-aging genes in protein-protein interaction (PPI) network and found that aging genes are characterized by several network topological features such as higher in degrees. And aging genes tend to be enriched in certain functions were also found. Based on these features, an algorithm was developed to detect aging genes genome wide. With a posterior probability score describing possible involvement in aging no less than 1, 1014 novel aging genes were predicted by decision trees. Evidence supporting our prediction can be found.*

Keywords: *Aging Genes, Algorithm, Drosophila Melanogaster, Non-Aging, Prediction, Protein-Protein Interaction (PPI)*

DOI: 10.4018/jkdb.2012040102

1. INTRODUCTION

Drosophila melanogaster is one of the most important model organisms for aging research. In comparison with other model systems, the relative closeness to human being has made it extremely valuable and rapid progress has been made in this field for the past decade. Up to now, 75 *Drosophila melanogaster* aging genes have been collected in the database (de Magalhaes et al., 2009). However, as the average lifespan of *Drosophila melanogaster* is about four months, identification of aging genes in wet-lab is very time-consuming. Therefore, systematic investigation of characteristics of current collected *Drosophila melanogaster* aging genes to develop an algorithm for predicting novel ones should be valuable for its guiding wet-lab research.

No protein functions in isolation but through interaction with other proteins. In fact, investigating characters such as network topological features of proteins in protein-protein interaction (PPI) network has helped a lot in identifying proteins encoded by disease genes (Furney, Higgins, Ouzounis, & Lopez-Bigas, 2006; Lopez-Bigas & Ouzounis, 2004; Xu & Li, 2006). For aging research, it has been found that proteins encoded by aging genes tend to have higher degrees in PPI network in *S. cerevisiae* which was then suggested as a feature to prioritize candidate proteins encoded by aging genes (Ferrarini, Bertelli, Feala, McCulloch, & Paternostro, 2005; Promislow, 2004). The shortest-path network analysis was also suggested as a useful approach to identify genetic determinants of longevity in *S. cerevisiae* (Managbanag et al., 2008). And, we have proposed a network based approach to identify aging genes in *C. elegans* (Li, Dong, & Guo, 2010). How about these characters of aging genes in *Drosophila melanogaster*? And can other topological features be used for aging genes research such as *K-core*? (Platzer, Perco, Lukas, & Mayer, 2007). These questions are needed for investigation.

In this paper, using the PPI network downloaded from BioGRID (Stark et al., 2006), we systematically investigated topological features of proteins encoded by aging versus non-aging genes in *Drosophila melanogaster* for the first time. We found that proteins encoded by aging genes are characterized by several topological features such as degree. For functional analysis, we found that aging genes tend to be enriched in certain biological processes. The Kolmogorov-Smirnov (*KS*) test was employed to analyze the distribution of each property of aging and non-aging genes. The features with significant difference were used for the subsequent novel aging genes prediction by decision tree. A posterior probability score for possible involvement in aging has been calculated for each gene. Finally, 1014 aging genes have been predicted with posterior probability score no less than 1. Evidence supporting our prediction can be found.

2. MATERIALS AND METHODS

2.1. Data Source

Downloaded from [http://genomics.senescence.info/\(de Magalhaes, et al., 2009\)](http://genomics.senescence.info/(de%20Magalhaes,%20et%20al.,%202009)), the database contained a list of 75 *Drosophila melanogaster* aging genes but only 51 of them are covered by the protein-protein interaction network and thus used for analysis. The protein-protein interaction data were downloaded from BioGRID (Stark, et al., 2006), consisting of a total of 7,610 proteins and 34,946 interactions. The self-interactions and redundant interactions were deleted. The functional annotation data were downloaded from Gene Ontology (Ashburner et al., 2000). InParanoid is a comprehensive database of orthologs (O'Brien, Remm, & Sonnhammer, 2005). The orthologs between *Drosophila melanogaster* and *C. elegans* were obtained from this database.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/predicting-aging-genes-drosophila-melanogaster/77928

Related Content

Bayesian Multilevel Analysis of Postnatal Care Utilization in Ethiopia

Dereje Bekele Dessie and Abay Sahle Nigussie (2021). *International Journal of Applied Research in Bioinformatics* (pp. 12-21).

www.irma-international.org/article/bayesian-multilevel-analysis-of-postnatal-care-utilization-in-ethiopia/267821

Wave-SOM: A Novel Wavelet-Based Clustering Algorithm for Analysis of Gene Expression Patterns

Andrew Blanchard, Christopher Wolter, David S. McNabb and Eitan Gross (2012). *Computational Knowledge Discovery for Bioinformatics Research* (pp. 104-126).

www.irma-international.org/chapter/wave-som-novel-wavelet-based/66707

Work Time Allocation of Family Doctors in Rural Tajikistan

Nargis S. Juraeva, Martin W. Bratschi, Kaspar Wyss and Ismoil S. Komilov (2020). *International Journal of Applied Research in Bioinformatics* (pp. 28-37).

www.irma-international.org/article/work-time-allocation-of-family-doctors-in-rural-tajikistan/261868

Data Mining Approach for the Early Risk Assessment of Gestational Diabetes Mellitus

Saeed Rouhani and Maryam MirSharif (2018). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-11).

www.irma-international.org/article/data-mining-approach-for-the-early-risk-assessment-of-gestational-diabetes-mellitus/202360

Redundancy-Based Feature Selection

(2011). *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations* (pp. 203-222).

www.irma-international.org/chapter/redundancy-based-feature-selection/53904