

# Is there a relationship between evolutionary rate and age of genes?

Niv Sabath, Eran Elhaik, and Dan Graur

Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA



## Abstract

It has recently been claimed that older genes tend to evolve more slowly than newer ones (Alba and Castresana, 2005). By simulation of genes of equal age, we show that the inverse correlation between age and rate is an artifact caused by our inability to detect homology when evolutionary distances are large. Since evolutionary distance increases with time of divergence and rate of evolution, homologs of fast evolving genes are frequently undetected in distantly related taxa and are, hence, misclassified as "new." This misclassification causes the mean genetic distance of "new" genes to be overestimated, and the mean genetic distance of "old" genes to be underestimated.



## The study of Alba and Castresana 2005

### Summary of Alba and Castresana, 2005:

- Human-mouse orthologous gene pairs were defined through reciprocal Blast.
- "Age" of each human-mouse gene pair was assigned by the phylogenetic distribution of their homologs among six genomes (Figure 1).
- Rate of evolution was calculated for each pair as the number of nonsynonymous substitutions per nonsynonymous site (Ka).
- A negative correlation was found between the rate of substitution and "age" (Figure 2).
- This implies that "young" genes evolve faster than "old" genes.

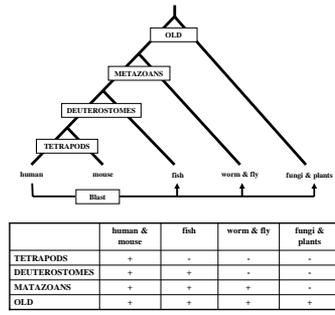


Figure 1. "Age" classification of human-mouse gene pair by the phylogenetic distribution of their homologs among six genomes. (Alba and Castresana, 2005. *Mol. Biol. Evol.* 22:598-606).

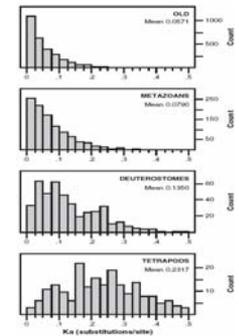


Figure 2. Distribution of nonsynonymous substitution rates (Ka values) in genes of four different "age" groups. The vertical bar represents the number of genes. "Age" of genes is claimed to be negatively correlated with the rate of evolution. (Alba and Castresana, 2005. *Mol. Biol. Evol.* 22:598-606).

## Our study

### Summary of our simulation study:

- We generated terminal sequences with different evolutionary rates, whose phylogenetic relationships are known using DAWG (DNA Assembly with Gaps) simulation program (Cartwright, 2005. *Bioinformatics* 21S3:31-38) (Figure 3).
- We used Blast to detect homology between gene A to genes C, D and E.
- A homolog was assumed present in C, D or E (E-value <= 10<sup>-4</sup>), otherwise, we inferred absence.
- In a manner analogous to Alba and Castresana's (2005. *Mol. Biol. Evol.* 22:598-606) OLD, METAZOANS, DEUTEROSTOMES, and TETRAPODS, each simulated A-B pair was assigned to one of four age groups: SENIORS, ADULTS, TEENAGERS, and TODDLERS (Figure 3).
- Rate of evolution was calculated for each A-B pair as the number of nucleotide differences per aligned sites.
- A negative correlation was found between the rate of substitution and "age" (Figure 4).

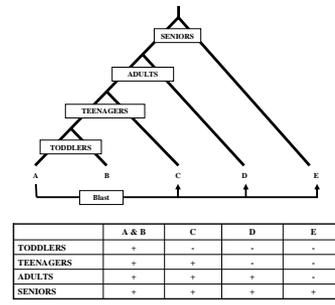


Figure 3. "Age" classification of A-B gene pair by the ability to detect homologs of A in C, D and E lineages.

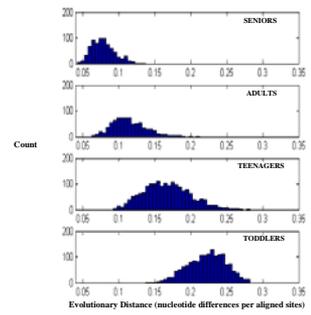


Figure 4. Observed distribution of evolutionary distances in four inferred "age" groups. Since all genes in the simulation are of equal age, the division into four "age" group is an artifact of the inference protocol. The evolutionary distance was calculated between lineages A and B. "Age" of genes is shown to be negatively correlated to the rate of evolution.

## Results and Conclusions

The distribution of the genetic distances and by implication the evolutionary rates, of the simulated genes are found to be negatively correlated with the "age" of the assigned group. Since all our simulated genes have the same evolutionary age, and since our results are similar to those obtained by Alba and Castresana (2005), we conclude that the inverse relationship between evolutionary rate and gene age is an artifact caused by our inability to detect similarity when genetic distances are large. Since genetic distance increases with time of divergence and rate of evolution, it is difficult to identify homologs of fast evolving genes in distantly related taxa. Thus, fast evolving genes may be misclassified as "new." The only conclusion that can be drawn from Alba and Castresana's (2005) study is that slowly evolving genes evolve slowly.

### Take Home Message:

From sequence similarity, one CAN infer homology.  
From absence of sequence similarity, one CANNOT infer absence of homology.

### References

Alba, M. M. and Castresana J. 2005. Inverse relationship between evolutionary rate and age of mammalian genes. *Mol. Biol. Evol.* 22:598-606.

Elhaik, E., Sabath N. and Graur D. 2006. The "Inverse relationship between evolutionary rate and age of mammalian genes" is an artifact of increased genetic distance with rate of evolution and time of divergence. *Mol. Biol. Evol.* 23:1-3.

Cartwright, R. A. 2005. DNA assembly with gaps (Dawg): simulating sequence evolution. *Bioinformatics* 21S3:31-38.

**Acknowledgment**  
This work was supported by NSF grant DBI-0543342. Poster production courtesy of TLC<sup>2</sup>