

UC Davis

UC Davis Previously Published Works

Title

Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis.

Permalink

<https://escholarship.org/uc/item/2g4367dw>

Journal

Genome research, 8(3)

ISSN

1088-9051

Author

Eisen, JA

Publication Date

1998-03-01

DOI

10.1101/gr.8.3.163

Peer reviewed

Table 2. Examples of Conditions in Which Similarity Methods Produce Inaccurate Predictions of Function

Evolutionary Pattern and Tree of Genes and Functions ¹	Gene With Unknown Function ²	Highest Hit Method		Phylogenomic Method		Comments
		Predicted Function ³	Accurate?	Predicted Function ⁴	Accurate?	
<p>A. Functional change during evolution.</p>	<p>1 ●</p> <p>2 ●</p> <p>3 ●</p> <p>4 ■</p> <p>5 ■</p> <p>6 ■</p>	<p>●</p> <p>●</p> <p>●</p> <p>●</p> <p>●/■</p> <p>●/■</p>	<p>+</p> <p>+</p> <p>+</p> <p>-</p> <p>±</p> <p>±</p>	<p>●</p> <p>●</p> <p>●/■</p> <p>●/■</p> <p>■</p> <p>■</p>	<p>+</p> <p>+</p> <p>±</p> <p>±</p> <p>+</p> <p>+</p>	<ul style="list-style-type: none"> Phylogenomic method cannot predict functions for all genes, but the predictions that are made are accurate. Highest hit method is misleading because function changed among taxa but hierarchies of similarity do not correlate with the function (see Boll Raff 1996).
<p>B. Functional change & rate variation.</p>	<p>1 ●</p> <p>2 ●</p> <p>3 ●</p> <p>4 ■</p> <p>5 ■</p> <p>6 ■</p>	<p>●</p> <p>●</p> <p>■</p> <p>●</p> <p>●</p> <p>■</p>	<p>+</p> <p>+</p> <p>-</p> <p>-</p> <p>-</p> <p>+</p>	<p>●</p> <p>●</p> <p>●/■</p> <p>●/■</p> <p>■</p> <p>■</p>	<p>+</p> <p>+</p> <p>±</p> <p>±</p> <p>+</p> <p>+</p>	<ul style="list-style-type: none"> Similarity based methods perform particularly poorly when evolution varies between taxa. Molecular phylogenetic methods can allow for rate variation and reconstruct gene history reasonably accurately.
<p>C. Gene duplication and rate variation.</p>	<p>1A ●</p> <p>2A ●</p> <p>3A ●</p> <p>1B ■</p> <p>2B ■</p> <p>3B ■</p>	<p>●</p> <p>●</p> <p>■</p> <p>■</p> <p>■</p> <p>●</p>	<p>+</p> <p>+</p> <p>-</p> <p>+</p> <p>+</p> <p>+</p> <p>-</p>	<p>●</p> <p>●</p> <p>●</p> <p>■</p> <p>■</p> <p>■</p> <p>■</p>	<p>+</p> <p>+</p> <p>+</p> <p>+</p> <p>+</p> <p>+</p> <p>+</p>	<ul style="list-style-type: none"> Most-similarity based methods are not ideally set up to deal with cases of gene duplication since orthologous genes do not always have significantly more sequence similarity to each other than to paralogs (Eisen et al. 1995; Zaslavskiy et al. 1996; Tatusov et al. 1997). Similarity-based methods perform particularly poorly when rate variation and gene duplication are combined. This even applies to the COG method (Table 1) since it works by classifying levels of similarity and not by inferring gene history. Nevertheless, the COG method is a significant improvement over other similarity based methods in classifying orthologs. Phylogenetic reconstruction is the most reliable way to infer gene duplication events and thus determine orthology.

¹ The true tree is shown but it is assumed that it is not known. Different colors and symbols represent different functions. Numbers correspond to different species.

² The function of all other genes is assumed to be known.

³ The top hit can be determined from the tree by finding the gene that is the shortest evolutionary distance away (as determined along the branches of the tree).

⁴ It is assumed that the tree of the genes can be reproduced accurately by molecular phylogenetic methods (see Fig. 1).

PHYLOGENETIC PREDICTION OF GENE FUNCTION

EXAMPLE A

METHOD

EXAMPLE B

