UC Davis
UC Davis Previously Published Works

Title

Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis

Permalink

Journal

ISSN

Author

Eisen, Jonathan A

Publication Date

DOI

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?	
A. Functional change during evolution.	1 ●	●	+	●	+	• 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 h... but hierarchies of similarity do not correlate with the function (see Bol... Raff 1996).
	2 ●	●	+	●	+	
	3 ●	●	+	●/■	±	
	4 ■	●	-	●/■	±	
	5 ■	●/■	±	■	+	
	6 ■	●/■	±	■	+	
B. Functional change & rate variation.	1 ●	●	+	●	+	• Similarity based methods perform particularly poorly when evolution... vary between taxa. • Molecular phylogenetic methods can allow for rate variation and reco... gene history reasonably accurately.
	2 ●	●	+	●	+	
	3 ●	■	-	●/■	±	
	4 ■	●	-	●/■	±	
	5 ■	●	-	■	+	
	6 ■	■	+	■	+	
C. Gene duplication and rate variation.	1A ●	●	+	●	+	• Most-similarity based methods are not ideally set up to deal with case... duplication since orthologous genes do not always have significantly m... sequence similarity to each other than to paralogs (Eisen et al. 1995; Za... al. 1996; Tatusov et al. 1997). • Similarity-based methods perform particularly poorly when rate variat... gene duplication are combined. This even applies to the COG method... Table1) since it works by classifying levels of similarity and not by inf... history. Nevertheless, the COG method is a significant improvement o... similarity based methods in classifying orthologs. • Phylogenetic reconstruction is the most reliably way to inferr gene du... events and thus determine orthology.
	2A ●	●	+	●	+	
	3A ●	■	-	●	+	
	1B ■	■	+	■	+	
	2B ■	■	+	■	+	
	3B ■	●	-	■	+	

[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 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).

PHYLOGENENETIC PREDICTION OF GENE FUNCTION

EXAMPLE A	METHOD	EXAMPLE B
2A	CHOOSE GENE(S) OF INTEREST	5
3A 2B 1A 2A 1B 3B	IDENTIFY HOMOLOGS	2 1 3 4 6 5
≡≡≡≡	ALIGN SEQUENCES	≡≡≡≡
1A 2A 3A 1B 2B 3B ← Duplication?	CALCULATE GENE TREE	1 2 3 4 5 6
1A 2A 3A 1B 2B 3B ← Duplication?	OVERLAY KNOWN FUNCTIONS ONTO TREE	1 2 3 4 5 6
1A 2A 3A 1B 2B 3B ← Duplication?	INFER LIKELY FUNCTION OF GENE(S) OF INTEREST	1 2 3 4 5 6 Ambiguous
Species 1 Species 2 Species 3 1A 1B 2A 2B 3A 3B ← Duplication	ACTUAL EVOLUTION (ASSUMED TO BE UNKNOWN)	1 2 3 4 5 6