

Brief guide to orthology prediction methods

Excerpt from "[Orthology prediction methods: A quality assessment using curated protein families](#)", Kalliopi Trachana, Thomas A. Larsson, Sean Powell, Wei-Hua Chen, Tobias Doerks, Jean Muller and Peer Bork, BioEssays Volume 33, Issue 10, pages 769-780.

Box 1 Comparison of orthology prediction methods

Orthology prediction methods can be classified based on the methodology they use to infer orthology into (i) graph-based and (ii) tree-based methods [12, 16, 17]. Different graph-based methods are designed to assign orthology for two (pairwise) or more (multiple) species. Graph-based methods assign proteins into OGs based on their similarity scores, while tree-based methods infer orthology through tree reconciliation.

Pairwise species methods (e.g. BHR, InParanoid, RoundUp):

Based on these methods, orthologs are best bi-directional hits (BBH) between a pair of species. BRH [46] is the first automated method and does not detect paralogs. InParanoid [47] implements an additional step for the detection of paralogs. RoundUp [48] uses evolutionary distances instead of BBH. In addition to the restriction of only two-species at a time, these methods are disadvantageous for long evolutionary distances.

Multi-species graph-based methods (e.g. COG, eggNOG, OrthoDB, OrthoMCL, OMA):

Due to the fast implementation and high scalability, there are many graph-based methods for multi-species comparisons. So far, all of them use either BLAST or Smith-Waterman (e.g. PARALIGN, SIMG) as sequence-similarity search algorithms. However, they are quite diverse regarding the clustering algorithms. COG, eggNOG, and OrthoDB share the same methodology: they identify three-way BBHs in three different species and then merge triangles that share a common side. OrthoMCL is a probabilistic method that uses a Markov clustering procedure to cluster BBH into OGs. OMA removes from the initial graph BBHs characterized by high evolutionary distance; a concept similar to RoundUp. After that, it performs clustering based on maximum weight cliques. Unique database characteristics are the hierarchical groups (OGs in different taxonomic levels) and "pure orthologs" (generate groups of one-to-one orthologs without paralogs), which has been introduced only by OMA (indicated as ** in the figure). Hierarchical groups can substitute the view of phylogenetic trees.

Multi-species tree-based methods (e.g. TreeFam, Ensembl Compara, PhylomeDB, LOFT):

Tree-based prediction methods can be separated into approaches that do (like EnsemblCompara, TreeFam, and PhylomeDB) and do not, e.g. LOFT [49], use tree-reconciliation. Tree-based methods also initially use homology searches; however, their criteria are more relaxed, as the orthology is resolved through tree topology. Although a reconciled phylogenetic tree is the most appropriate illustration of orthology/paralogy assignment, there are a few caveats to such an approach, namely their scalability and sensitivity to data quality.

For a more detailed and extensive discussion of the differences among orthology methodology, we recommend refs. [12, 16, 17].

Phylogenetic distribution describes the species range of each database. Homology search shows a few technical differences for recruiting orthologs. [§]: Supplies OGs whose members share only orthologous relationships. *: The user can compare any two genomes spanning a phylogenetic distance from bacteria to animals.

		Phylogenetic distribution	Paralogs	Homology search	Clustering strategy	Hierarchical groups
<p>Pairwise species comparison</p> <p>Multi-species comparison</p> <p>GRAPH-BASED METHODS</p>	BRH	ALL*	NO	BLAST	None	-
	InParanoid	ALL*	YES	BLAST	None	-
	RoundUp	ALL*	NO	Evol. Distance	None	-
	COG	ALL	YES	BLAST	Triangles	NO
	eggNOG	ALL	YES	BLAST	Triangles	YES
	OrthoDB	Eukaryotes	YES	PARALIGN	Triangles	YES
OMA	ALL	YES**	SIMD & Evol. Distance	Maximum weight cliques	YES	
OrthoMCL	ALL	YES	BLAST	Markov Clustering	NO	
<p>Family tree</p> <p>Species tree</p> <p>Reconciled tree</p> <p>TREE-BASED METHODS</p>	TreeFam	Metazoa	YES	BLAST & HMM	Hierarchical clustering	-
	Ensembl Compara	Metazoa	YES	BLAST	Hierarchical clustering	-
	PhylomeDB	ALL	YES	BLAST [§]	None	-

For the [reference list](#), please go to the original [article](#).

This text and figure correspond to Box 1 of the original article and are therefore inseparable.