Benchmarking Data Visualization

Garrayo J. [1][2], Fernández-Rodríguez V. [1][2], Haas J. [3], Sundesha V. [1][2], Valencia A. [1][2][4], Gelpí J.L. [1][2][5], Capella-Gutiérrez S. [1][2]

[1] Barcelona Supercomputing Center (**BSC**), Barcelona Spain.
[2] Spanish National Bioinformatics Institute (**INB-ISCIII**), Barcelona Spain.
[3] University of Basel & Swiss Institute of Bioinformatics (**SIB**), Basel Switzerland.
[4] Catalan Institution for Research and Advanced Studies (**ICREA**), Barcelona Spain.
[5] Universitat de Barcelona (**UB**), Barcelona Spain.

Abstract. Scientific benchmarking provides a mechanism for users to systematically identify the most effective methods for the problem at hand, sets a minimum requirement for new tools and resources, and guides the development of more accurate inference methods. However, understanding the results of specific scientific benchmark approaches is not a trivial task for end-users who might not have the background knowledge to correctly interpret visualized results.

Plot and classify the benchmarking results

Here we proposed three alternative methods for converting traditional 2D plots into a more intuitive classification system for helping non-experts to correctly understand these results. We have used real data from the **Quest for Orthologs** initiative to perform the data transformations using …
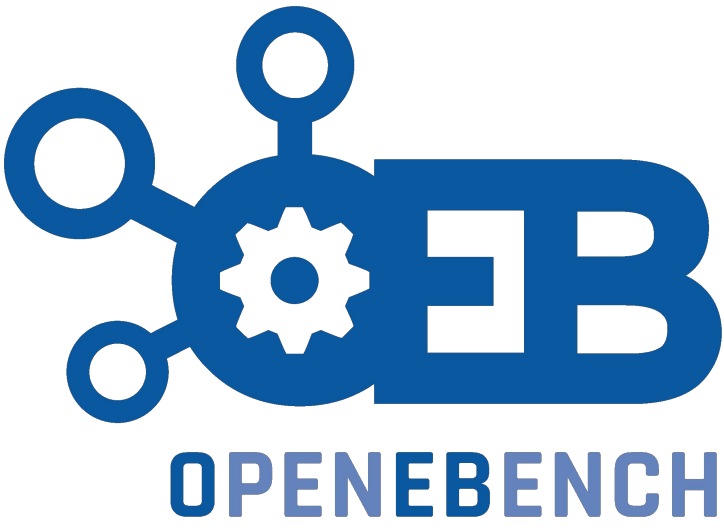
◆ **Square quartiles** (Figure 1 - A)
◆ **Diagonal quartiles** (Figure 1 - B)
◆ **Clustering** (Figure 1 - C)
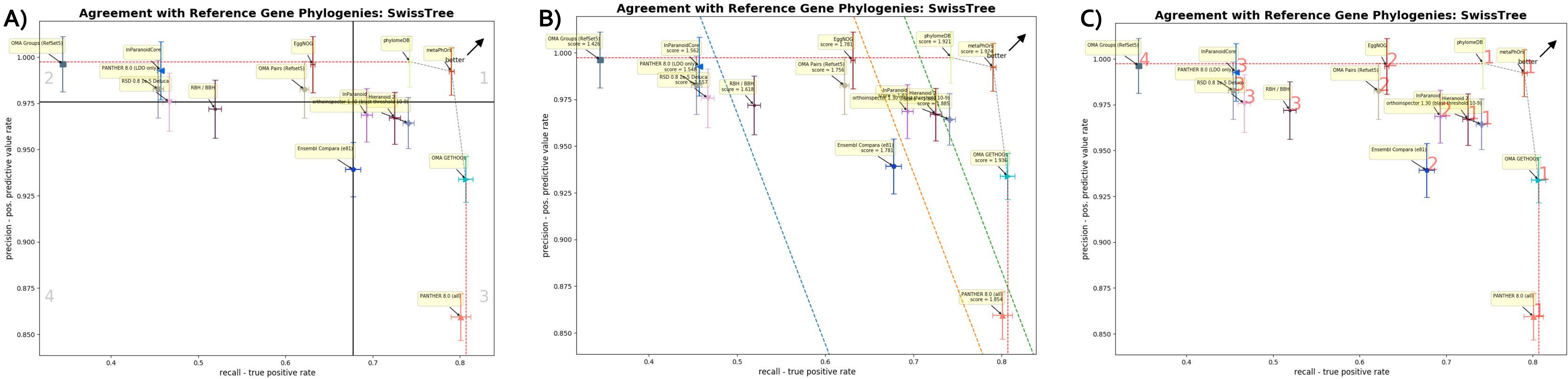
Figure 1. Agreement with Reference Gene Phylogenies (SwissTree) benchmark results, from the Quest for Orthologs community (http://orthology.benchmarkservice.org), classified using different methods. **A)** square quartiles - split up the plotting area in four parts according to the values of the metrics. Tools in the top right rectangle display the best performance. **B) Diagonal quartiles** - obtained by assigning a score to each participant according to its distance to the *optimal performance* corner. Tools in the right triangle display the best performance. **C) Clustering** - K-Means unsupervised learning algorithm to cluster the participant tools and ranked those groups according to the score of clusters' centroids. Tools in the cluster No. 1 (see red numbers) display the best performance. **D)** Participants classification using the three methods described on A), B) and C).

D)

TOOL	Quar_sqr	Quar_diag	Cluster
Ensembl Compara (e81)	3	3	2
EggNOG	2	3	2
PANTHER 8.0 (LDO only)	2	4	3
Hieranoid 2	3	2	1
InParanoidCore	2	4	3
OMA GETHOGs	3	1	1
orthoinspector 1.30	3	1	1
RBH / BBH	4	3	3
phylomeDB	1	1	1

Tabular format results

Tables can be used to assist on the transformation of scientific benchmarking results from experts perspective to non-expert perspective. Those transformation can be done for individual benchmarking cases (Figure 1 - D), or can be obtained for all categories considered in the benchmarking campaign (Figure 2). Importantly, tools can be classified into different categories depending on the transformation applied (Figure 3). Therefore, it is important to closely work with Scientific communities on the best transformation methods for their results.

DIAGONAL QUARTILES CLASSIFICATION RESULTS											
TOOL / BENCH. METHOD -->	EC Conservation	GO Conservation	G-STD Eukaryota	G-STD Fungi	G-STD LUCA	G-STD Vertebrata	STD Eukaryota	STD Fungi	SwissTree	TreeFam-A	# RANKING #
Ensembl Compara (e81)	2	2	4	3	4	3	2	4	3	3	# 11
EggNOG	4	3	3	3	4	1	4	2	3	2	# 9
PANTHER 8.0 (LDO only)	4	4	2	2	3	4	3	3	4	4	# 13
OMA Pairs (Refset5)	3	3	4	4	3	4	2	4	3	3	# 14
InParanoidCore	4	3	4	1	3	3	3	3	4	4	# 12
OMA GETHOGs	1	1	3	3	1	4	3	3	1	3	# 7
Hieranoid 2	2	4	2	3	2	3	1	1	2	2	# 6
orthoinspector 1.30	1	2	1	2	2	1	2	1	1	1	# 1
RBH / BBH	3	3	1	2	1	1	3	3	3	3	# 8
phylomeDB	3	2	3	1	4	2	1	1	1	1	# 4
OMA Groups (RefSet5)	4	4	4	4	4	4	4	4	4	4	# 15
PANTHER 8.0 (all)	1	1	1	4	2	3	1	2	2	1	# 3
RSD 0.8 1e-5 Deluca	3	4	2	1	1	2	4	4	4	4	# 10
InParanoid	1	1	1	1	1	3	1	1	2	2	# 2
metaPhOrs	2	1	4	3	4	1	1	1	1	1	# 5

Figure 2. Diagonal quartiles classification results in table format. Swisstree benchmark (see Figure 1) is highlighted in orange and tools which show the best performance for each of the benchmarks are highlighted in green. The column on the right assigns a ranking to each of the tools according to their performance across all benchmarking categories.

BENCHMARKING METHOD -->	EC Conservation			GO Conservation			G-STD Eukaryota			G-STD Fungi			G-STD LUCA			G-STD Vertebrata			STD Eukaryota			STD Fungi			SwissTree			TreeFam-A			# RANKING #		
TOOL / QUARTILES -->	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	SQR	DIAG	CLUST	#SQR	#DIAG	#CLUST
Ensembl Compara (e81)	3	2	2	3	2	3	4	4	3	2	4	4	2	3	3	2	3	2	3	2	3	4	4	2	3	3	2	4	3	3	# 15	# 11	# 9
EggNOG	2	4	3	1	3	3	3	3	2	4	3	2	4	4	1	1	1	2	4	4	3	1	2	2	2	3	2	1	2	2	# 4	# 9	# 11
PANTHER 8.0 (LDO only)	2	4	3	2	4	4	3	2	2	1	2	2	4	3	3	2	4	3	3	2	4	3	4	3	2	4	3	2	4	4	# 13	# 13	# 12
OMA Pairs (Refset5)	2	3	3	2	3	3	2	4	3	2	4	4	3	3	4	3	4	3	2	2	3	2	4	2	2	3	2	4	3	3	# 12	# 14	# 13
InParanoidCore	2	4	3	2	3	2	4	4	3	2	1	2	2	3	3	2	3	3	2	3	3	3	3	2	2	4	3	4	4	3	# 1	# 12	# 14
OMA GETHOGs	3	1	1	3	1	1	3	2	3	3	3	3	1	1	1	4	4	2	2	3	3	3	3	2	3	1	1	3	2	1	# 11	# 7	# 6
Hieranoid 2	1	2	2	2	4	4	1	2	4	2	3	4	1	2	4	2	3	1	1	1	2	1	1	2	3	2	1	3	2	2	# 5	# 6	# 7
orthoinspector 1.30	3	1	2	3	2	2	3	1	2	3	2	1	3	2	2	1	1	1	2	2	1	1	1	1	3	1	1	3	1	1	# 9	# 1	# 2
RBH / BBH	4	3	2	2	3	3	1	1	2	2	2	1	1	1	1	1	1	2	3	3	3	3	3	3	4	3	2	3	3	3	# 7	# 8	# 8
phylomeDB	3	3	2	3	2	2	2	3	3	1	1	1	2	4	3	1	2	2	3	1	1	2	1	1	1	1	1	1	1	2	# 2	# 4	# 5
OMA Groups (RefSet5)	2	4	4	2	4	4	2	4	4	2	4	4	4	4	4	2	4	4	2	4	4	4	4	4	2	4	4	3	4	4	# 3	# 15	# 15
PANTHER 8.0 (all)	3	1	1	3	1	1	3	1	1	2	4	3	2	2	2	3	3	3	3	1	2	3	1	1	3	2	1	3	1	1	# 14	# 3	# 3
RSD 0.8 1e-5 Deluca	2	3	3	2	4	3	2	2	2	1	1	1	1	1	2	3	4	4	3	4	4	2	4	4	2	4	3	4	4	4	# 10	# 10	# 10
InParanoid	3	1	1	2	1	1	1	1	1	1	1	1	2	1	1	2	3	2	3	1	1	2	2	2	2	2	1	2	2	2	# 6	# 2	# 4
metaPhOrs	3	2	2	3	1	2	4	4	3	3	3	2	4	3	3	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	# 5	# 5	# 4

Figure 3. Overall results for all benchmarking categories and transformation methods (squares, diagonals, clusters) with tools ranking (right columns). Swisstree benchmark (see Figure 1) is highlighted in orange. Colors in the '**Ranking**' section refer to the comparison among the results obtained using the different methods. Light red means worse results than in the previous classification while dark red means even worse results than in previous classification/s. Conversely, green means better results than the previous transformation while light green means even better results than the previous classification/s.

Contact:

Javier Garrayo javier.garrayo@bsc.es
Salvador Capella salvador.capella@bsc.es

Barcelona Supercomputing center.
C/Jordi Girona, 29, 08034 Barcelona. Spain