

OpenEBench Scientific Benchmarking

Vicky Sundesha ^{1,2}, Javier Garrayo ^{1,2}, Laia Codó ^{1,2}, Dmitry Repchevsky ^{1,2}, Eva Martin del Pico ^{1,2}, Víctor Fernández-Rodríguez ^{1,2}, José María Fernández-González ^{1,2}, Juergen Haas ⁵, Eduard Porta-Pardo ^{1,2}, Analia Lourenco ⁶, Adrian Altenhoff ⁵, Christophe Dessimoz ⁷, Isabel Cuesta ⁸, Sara Monzon ⁸, Alfonso Valencia ^{1,2,3}, Josep Ll Gelpi ^{1,2,4}, Salvador Capella-Gutiérrez ^{1,2}

¹ Barcelona Supercomputing Center (BSC), Barcelona Spain.

² Spanish National Bioinformatics Institute (INB), Barcelona Spain.

³ Catalan Institution for Research and Advanced Studies (ICREA), Barcelona Spain.

⁴ Universidad de Barcelona (UB), Barcelona Spain.

⁵ Swiss Institute of Bioinformatics (SIB), Lausanne/Basel Switzerland.

⁶ University of Vigo (UVIGO), Vigo Spain.

⁷ University of Lausanne (UNIL), Lausanne Switzerland.

⁸ Instituto de Salud Carlos III (ISCIII), Madrid Spain.



OpenEBench (<https://openebench.bsc.es>), is the ELIXIR benchmarking and technical monitoring platform for bioinformatics tools, web servers and workflows.

OpenEBench is part of the ELIXIR Tools platform (<https://www.elixir-europe.org/platforms/tools>). Its development is led by the INB/ELIXIR-ES (<https://inb-elixir.es>)

Central node at the Barcelona Supercomputing Center (BSC) in collaboration with different partners across Europe.

Scientific Benchmarking

Scientific benchmarking determines the performance of bioinformatics resources in the context of predefined reference datasets and metrics reflecting specific scientific challenges. These metrics allow to objectively evaluate the relative scientific performance of the different participating resources. It is even possible to understand what are the software potential biases, strengths and weaknesses and/or under which conditions do they perform better or worst. Scientific benchmarking can also be used as a quality control mechanism for newly developed resources and/or releases of established ones.

OpenEBench Benchmarking model

There are four main aspects to the scientific benchmarking model which have been designed in affinity with several established scientific communities:

- ❑ **Communities:** Unbiased and objective evaluations of bioinformatics resources are challenging to set-up and can only be effective when built and implemented around community driven efforts. There are several communities from different scientific domains engaged in the OpenEBench platform e.g. TCGA, QfO, CAMEO, GMI Initiative, CAID, etc. The collaboration with the DREAM Challenges will contribute towards the sustainability of OpenEBench.
- ❑ **Benchmarking events:** A benchmarking event corresponds to a benchmarking experiment which is held at a certain period of time. It starts when the organizing community opens one or more challenges by providing participants with specific input datasets, allows them to submit their predictions, and make public the results of such event.
- ❑ **Participants:** Participants are users e.g. developers and/or advanced end-users, who want to understand the performance of specific software as compared with others in the field.
- ❑ **Datasets:** Those are one of the key elements in any benchmarking event. There are five types of datasets defined in OpenEBench:
 - ❑ **Input datasets** which are provided by the community to the participants in order to be processed by their tools, web-servers and/or workflows.
 - ❑ **Reference datasets** are the Gold/Silver Standard datasets that are also provided by the community but are not visible. Metrics will be computed considering these datasets.
 - ❑ **Participant datasets** are datasets uploaded by the participants to be benchmarked against the reference datasets.
 - ❑ **Assessment datasets** contain information about the results of metrics applied to a certain participant dataset.
 - ❑ **Aggregation datasets** are obtained by summarizing and/or computing statistics on assessment datasets. This allow to compare results across participants.

OpenEBench benchmarking architecture

The architecture at OpenEBench is divided into three levels (*Figure 1*). **Level 1** is based on hosting, accessing and visualizing previously generated benchmarking data; **Level 2** consists in computing community-specific metrics comparing participants predictions against previously defined reference datasets by the Communities. This process follows the workflow define in *Figure 2*. **Level 3** holds the whole benchmarking effort within OpenEBench by allowing participants to upload their workflows packaged in software containers. Then, technical and scientific performance are assessed by executing those workflows in the same environment. In this architecture, each level depends on the previous one, that is, each level can be fed in data coming from upper levels.

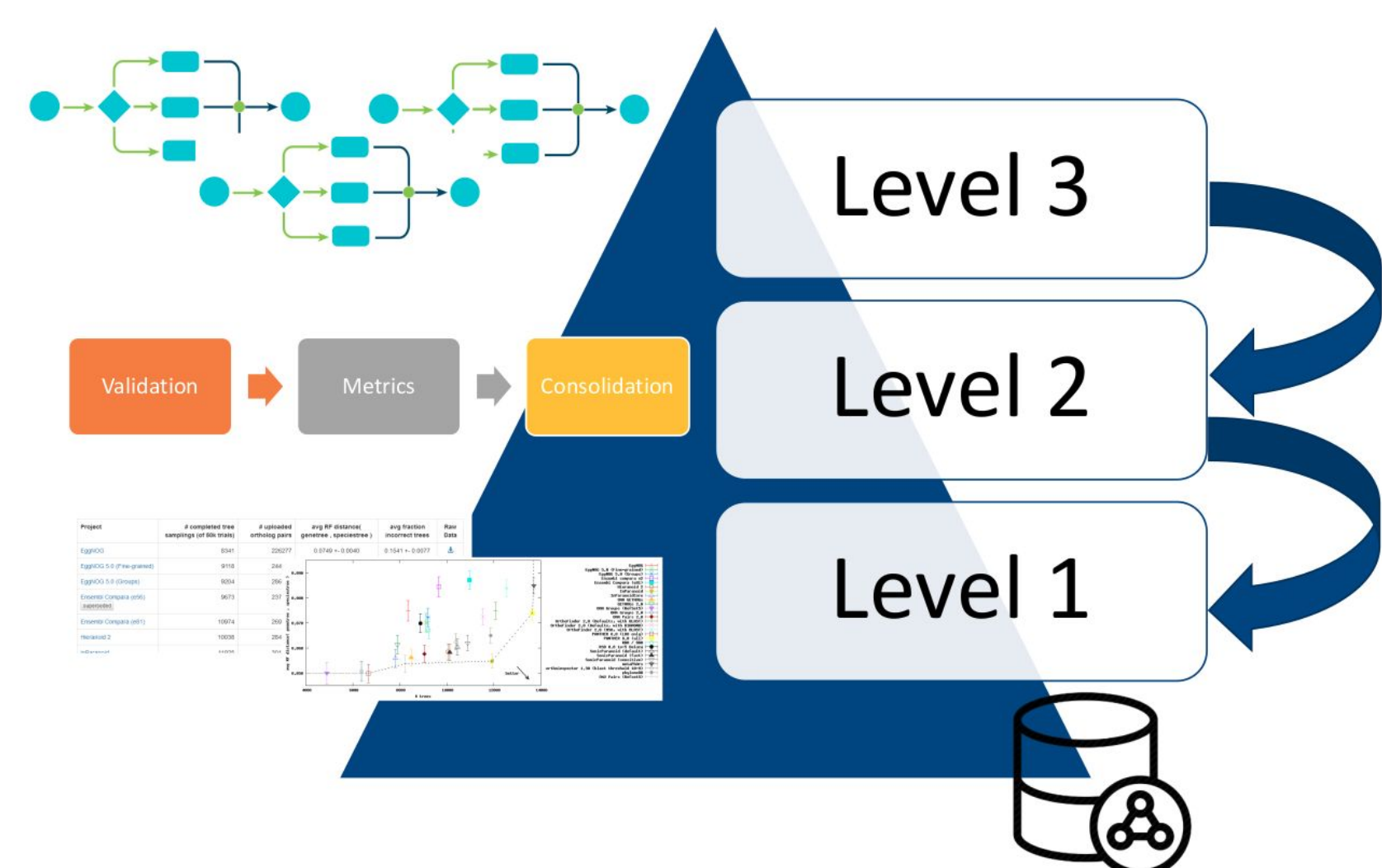


Figure 1. Benchmarking architecture. The three levels of the openEBench architecture

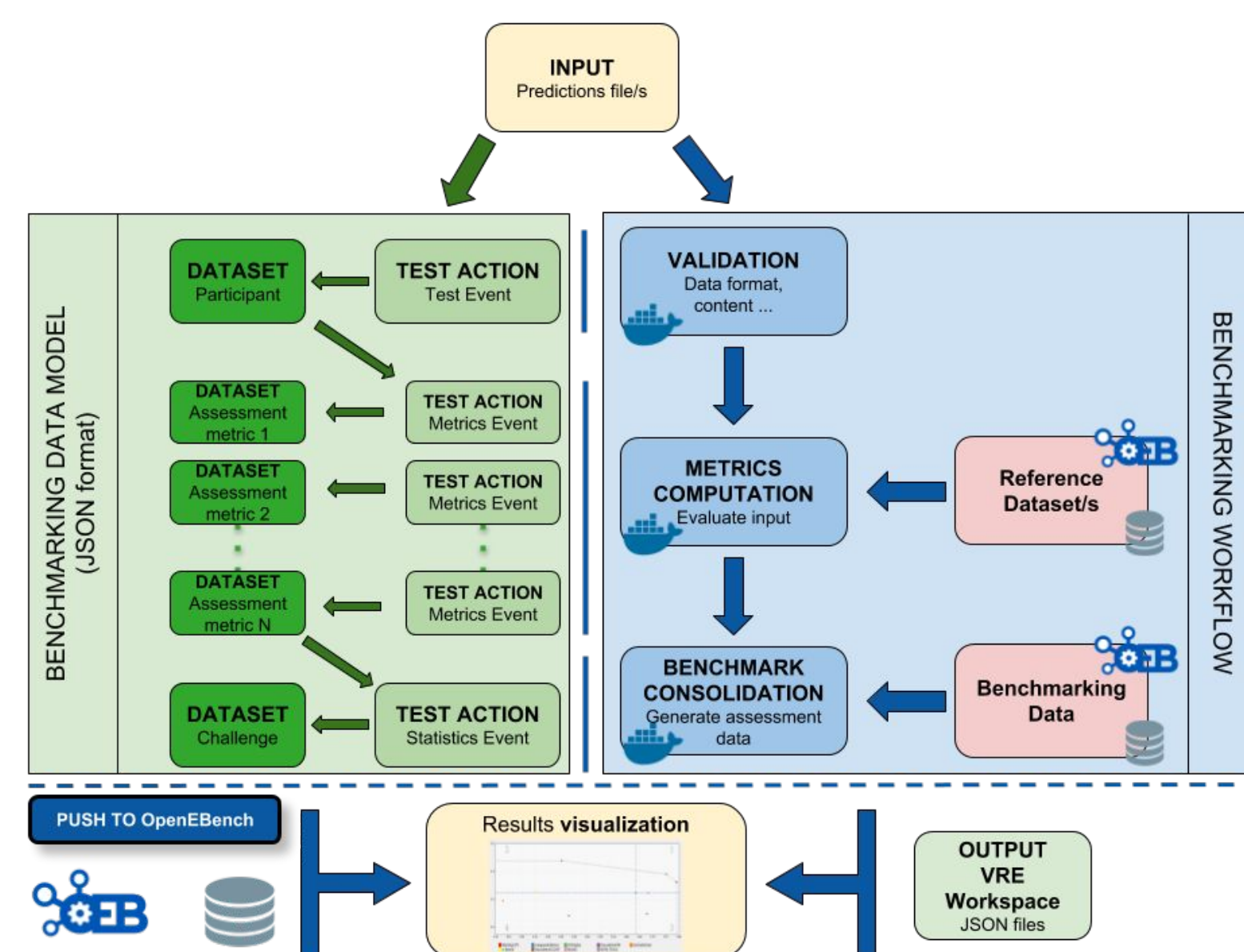


Figure 2. Benchmarking Workflow. OpenEBench workflows receive one or more input files with participants predictions (participant datasets) and are run in three steps (blue box): format validation, metrics computation - predictions are benchmarked against community gold/silver standard, and benchmark consolidation - aggregation of several challenge results in a single dataset. While this pipeline is executed several files are generated following the schemas of the Benchmarking Data Model (green box); those files are then pushed to the database and results can be visualized in the interface.

Quest for Orthologs benchmarking Level:1 use case

The Quest for Orthologs consortium (QfO) which is currently involved in OpenEBench is a joint effort to benchmark, improve and standardize orthology predictions through collaboration, the use of shared reference datasets, and evaluation of emerging new methods.

In the QfO benchmarking service orthology, predictions are evaluated using several challenges such as the *Species Tree Discordance* test or the *Gene Ontology Conservation* test, using specific metrics like the *Robinson-Foulds distance* and/or the *Schlicker similarity*.

OpenEBench offers a complete web interface which allows users to visualize those results (*Figures 3 & 4*), evaluate the available resources and choose the most appropriate software for the problem at hand.

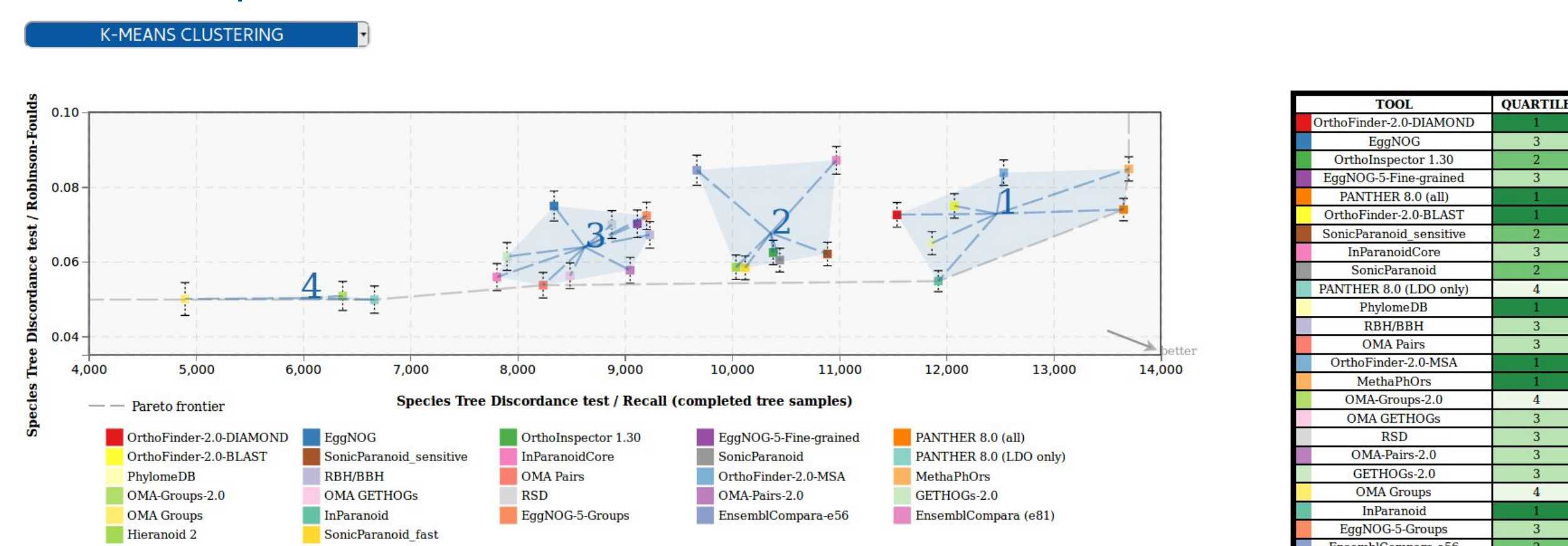


Figure 3. QfO Species Tree Discordance challenge results visualization. Community metrics can be visualized in a 2D plot and transformed to tabular format by applying different classification methods - quartiles, clustering, etc... This facilitates the evaluation and interpretation of the results to both expert and non- expert benchmarking users.

CHALLENGE + TOOL +	ECtest#	GOtest#	G_STD_Eukaryota#	G_STD_Fungi#	G_STD_LUCA#	G_STD_Vertebrate#	STD_Eukaryota#	STD_Fungi#	SwissTree#	TreeFam-A#
Hieranoid 2#	3	3	2	2	3	3	2	2	2	2
InParanoid#	2	3	2	2	2	3	1	1	2	2
InParanoidCore#	3	4	3	2	4	4	3	3	3	3
MethaPhOrs#	2	3	3	2	3	1	1	1	1	1
OMA CETHOGs#	2	3	3	3	2	3	3	2	1	2
OMA Groups#	4	4	4	4	4	4	4	4	4	4
OMA Pairs#	3	4	3	3	3	3	3	3	2	3
OMA-Groups-2.0#	4	4	4	4	4	4	4	4	4	4
OMA-Pairs-2.0#	3	4	3	3	3	3	3	3	2	2
OrthoFinder-2.0-BLAST#	2	3	1	1	2	1	1	1	1	1
OrthoFinder-2.0-DIAMOND#	2	3	1	1	2	1	1	1	1	1
OrthoFinder-2.0-MSA#	2	3	1	1	2	1	1	1	1	1
OrthoInspector 1.30#	2	3	2	2	2	2	2	2	2	1
PANTHER 8.0 (LDO only)#	3	4	2	2	3	3	4	2	3	3

Figure 4. Benchmarking Event Results Table. Results from a whole Benchmarking Event can be visualized in a single 'Summary Table', which shows the overall participants' performance across the different Challenges in the Benchmark.

OpenEBench Virtual research environment (OEB VRE)

OpenEBench Infrastructure also consists of a VRE which is used in the different levels. More information on the VRE can be found at the poster with the title: **OpenEBench Virtual Research Environment**.

Contact:

Salvador Capella: salvador.capella@bsc.es

Vicky Sundesha: vicky.sundesha@bsc.es

Barcelona Supercomputing Center (BSC).

Jordi Girona, 31, 08034. Barcelona, Spain.



ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559.

