

# All-in-One Microbiome Analysis with OmicsBox



# OmicsBox

David Seide, Gabriele Nocchi, Jesús Gutiérrez, Robert Nica, Mariana Monteiro, Alejandro Rocamonde, Carlos Menor, Francisco Salavert, Stefan Götz  
BioBam Bioinformatics S.L., Valencia, Spain

## Abstract

Metagenomics has many applications in different fields such as medicine, engineering, agriculture among others, as well as, benefits from DNA sequencing technology improvements and the continuous development and evolution of dedicated bioinformatics tools and databases. However, there is no universally suitable analysis strategy that guarantees best results for all applications. Experimental design, different input data and hypothesis require detailed knowledge about used tools, databases and possible combinations of the analysis steps. This is a major challenge for non bioinformatics experts.

Here the novel Metagenomics Module of OmicsBox is presented. It employs best practices in an intuitive way with many analysis options. It works in an out-of-the-box way without complex installations or specific hardware requirements. Up-to-date tools and databases are directly available via OmicsBox cloud. Raw read sequencing data, either 16S amplicon based or shotgun data can be used. Analysis steps can be combined in many possible ways and custom pipelines can be created with a workflow editor. This allows to obtain biological insights in a timely manner. All analysis steps can be carried out on a standard laptop or workstation with 4GB of RAM running Windows, Linux or macOS.

## Main Features

**Taxonomic Classification** Identify present species (Bacteria, Archaea, Virus) with Kraken and visualize results with multi-level pie-charts (Krona) as well as inter-sample comparison bar charts.

**Metagenomic Assembly** Choose between MetaSPAdes and MEGAHIT to assemble large datasets easy and fast in the cloud.

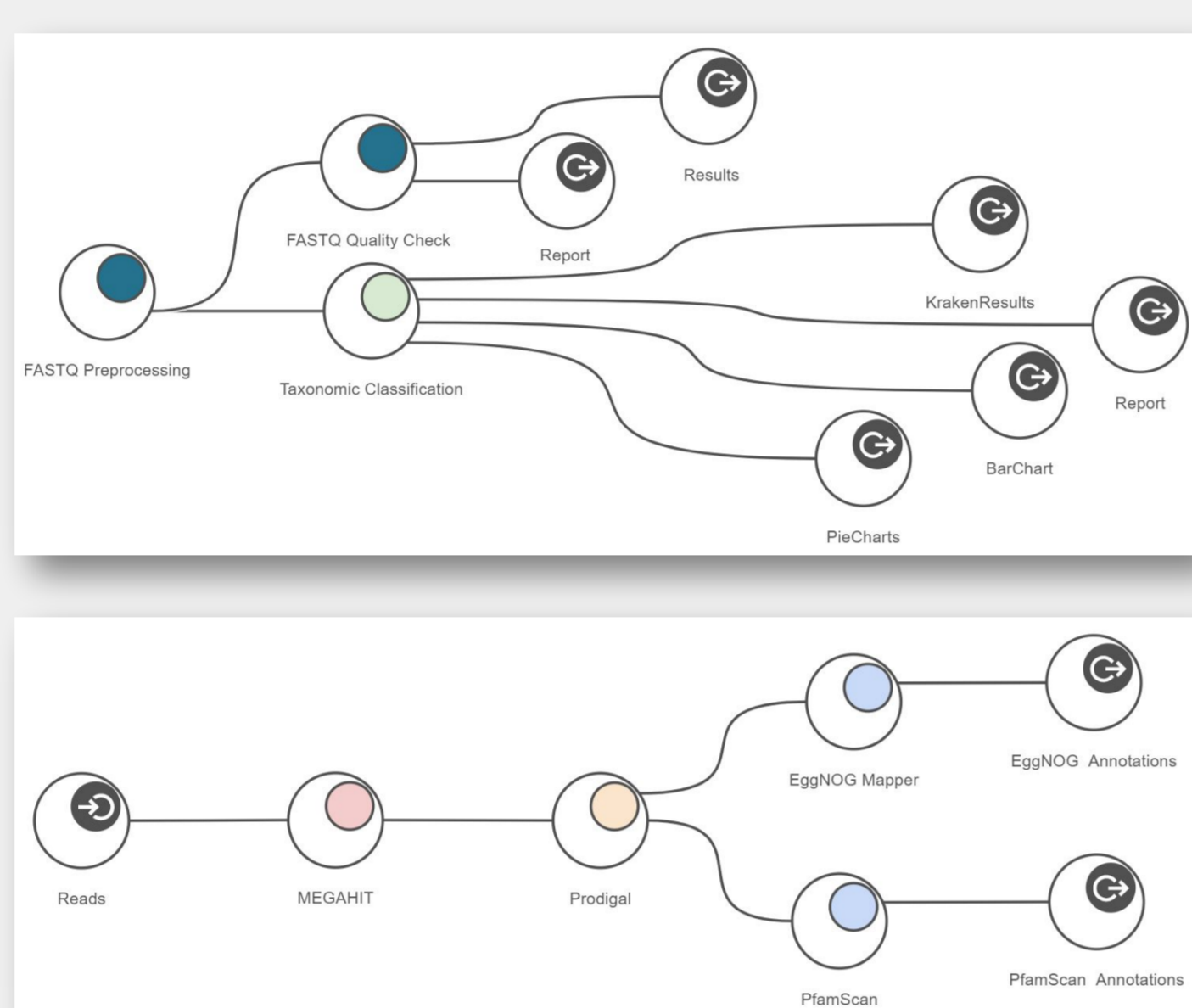
**Gene Prediction** Use FragGeneScan directly on read data and Prodigal for assembled contigs to identify and extract possible genes and proteins.

**Functional Annotation** Use EggNOG-Mapper and PfamScan to obtain high-throughput functional annotations. Results can be represented and compared visually with GO graphs and charts.

## Workflows

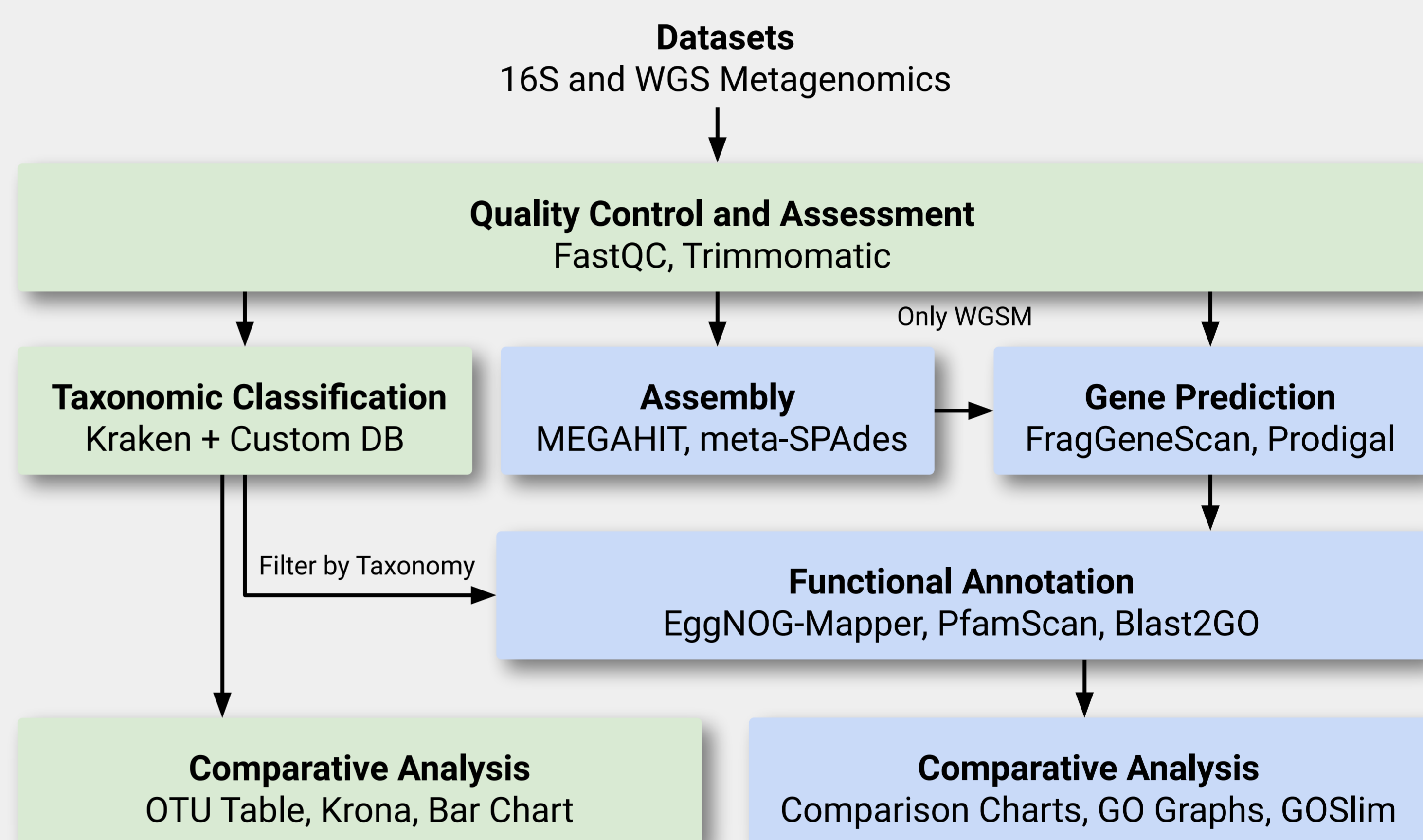
### Taxonomic Classification

The workflow in the figure on the right shows taxonomic classification of metagenomics data with OmicsBox. First, the reads are preprocessed, and a report is generated. Kraken is used to identify and count all the different OTUs for further interpretation. The spreadsheet-like result can be filtered and organized. The PDF report gives a clean overview about the most abundant OTUs at different levels for each sample.



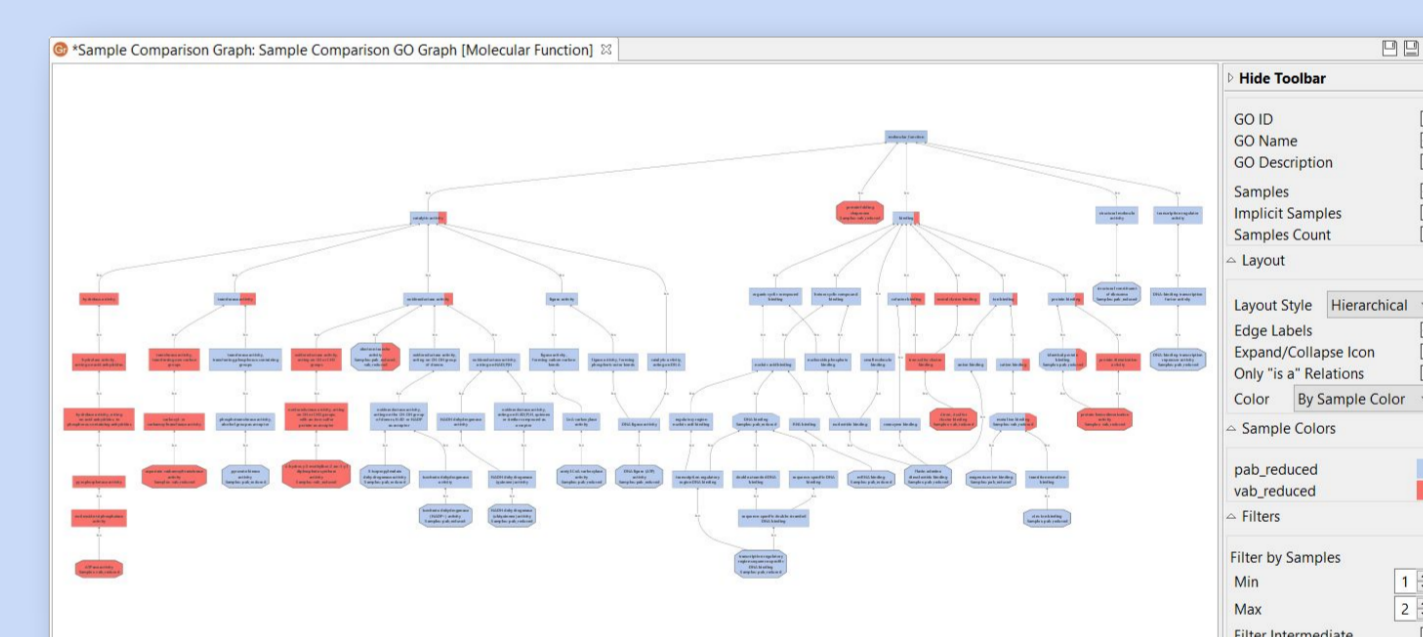
### Functional Annotation

Functional characterization of metagenomics data is a complex task. The OmicsBox Metagenomics module allows you to design streamlined workflows to easily combine the typically resource-demanding assembly step with gene prediction, as well as high-throughput functional annotation for large metagenomics datasets. The workflow in the bottom figure shows an example workflow of the combination of MEGAHIT with Prodigal. Fast and comprehensive functional annotation is achieved with the integration of EggNOG-Mapper and PfamScan.



## Colored GO Graph

Compare GO annotation abundances visually between samples and take into account the GO hierarchy. Nodes are colored partially depending in which samples they are more abundant.



## Functional Abundance Comparison

Compare annotation abundances (GO, Pfam, EggNOG, etc.) between different samples with many individual options. Choose the chart style, sample colors and names, and sort by similarity or difference among others.



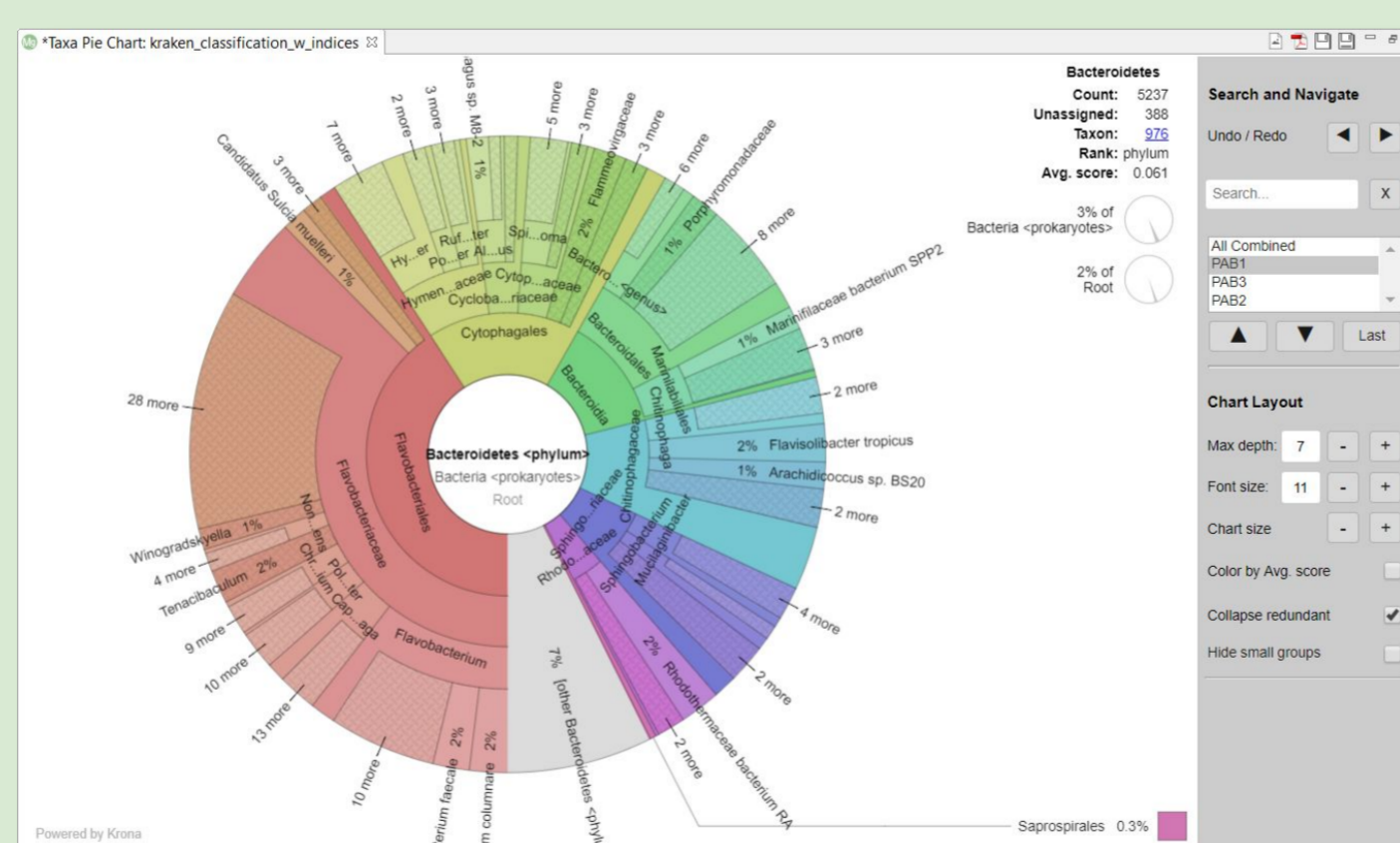
## OTU Abundances Table

Sort and filter the taxonomic classification results, and extract all reads that belong to selected OTUs. Continue the assembly with reads that belong to selected groups of OTUs and obtain a more specific assembly. Refine your reports with customized distribution charts in various styles.

Rank	OTU	Scientific Name	PA81	PA82	PA83
1	142978	Bacteria	102978	73332	85347
2	11404	Bacteria	11404	8853	10163
3	1234	Proteobacteria	7715	6298	7131
4	28216	Betaproteobacteria	3326	2984	3565
5	80840	Burkholderiales	2826	2536	3037
6	28211	Alphaproteobacteria	2360	1612	1741
7	1236	Gamma proteobacteria	1427	1173	1225
8	201174	Actinobacteria	1421	988	1191
9	30304	Comamonadaceae	1308	1130	1130
10	17160	Actinobacteria	1248	1130	1130
11	356	Rhizobiales	1130	1130	1130
12	119060	Burkholderiaceae	556	556	556
13	976	Bacteroidetes	545	545	545
14	1239	Firmicutes	502	502	502
15	72274	Pseudomonadales	447	447	447
16	204437	Sphingomonadales	410	276	283

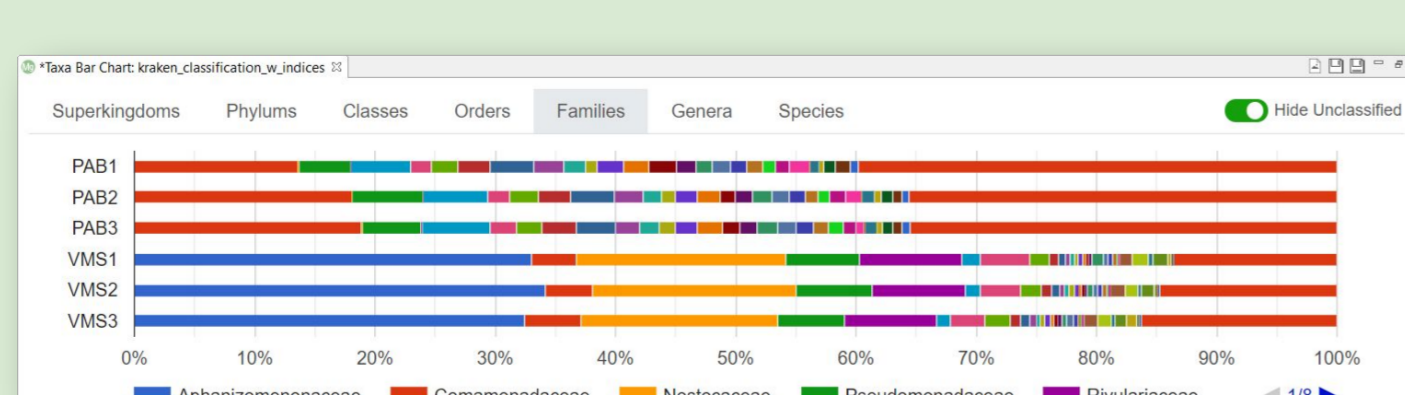
## OTU Abundances Pie Chart

Interactively visualize all OTUs and their abundances with the Krona chart library. Browse through each sample's rank hierarchy and compare abundances or color each OTU according to the corresponding confidence scores determined by Kraken.



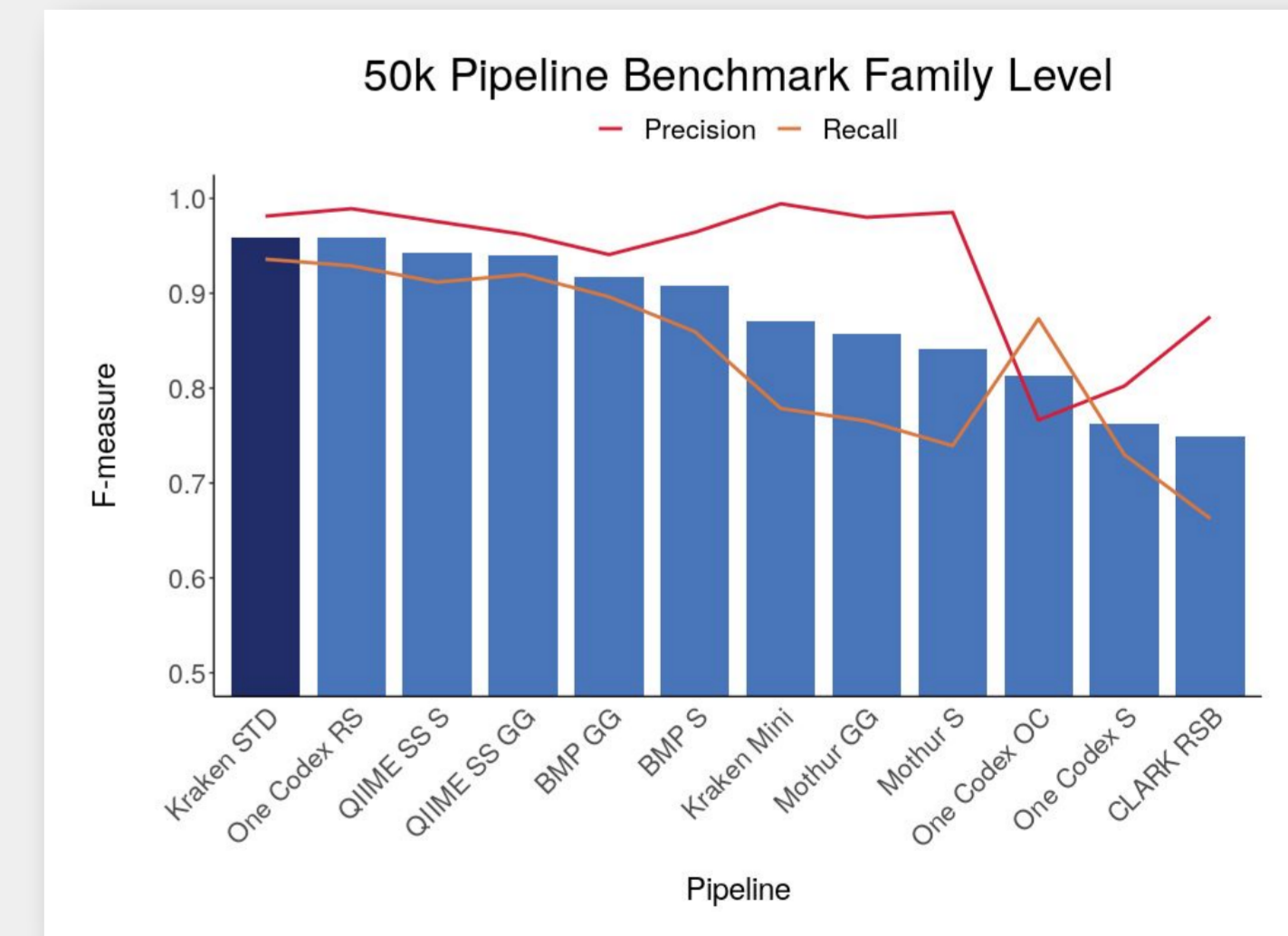
## Inter-Sample Comparison Bar Chart

Summarize and compare OTUs abundance compositions at certain ranks and draw conclusions on what factors influence your data.



The OmicsBox Metagenomics Module allows to perform best-practices microbiome analysis steps while successfully eliminating the hurdles of complex installations and high-performance computing requirements. Exhaustive benchmarking for 16S (Ref: Siegwald L, et al.) showed best-in-class performance. Taxonomic classifications with Kraken (Ref: Wood DE. and Salzberg SL.) for bacteria, archaea, viral data showed remarkable F1-scores (see chart). A WGS metagenomics study of two different soda lake microbial communities (see QR code) showed that the generation of functional annotations and the inter-sample comparisons were straight-forward and offer insights for the interpretation of the underlying biological conditions.

## Results



Datasets and comparison taken from Siegwald L, et al. PLoS One. 2017;12(1):e0169563. doi:10.1371/journal.pone.0169563 and compared against the current Kraken standard database performance.



Link to soda lake study use case: <https://bit.ly/2VWoa3F>

Please visit [www.biobam.com/omicsbox](http://www.biobam.com/omicsbox)

BioBam is a leading bioinformatics solution provider which accelerates research in disciplines such as agricultural genomics, microbiology and environmental NGS studies, among others. BioBam is committed to the development of user-friendly software solutions for biological research.

