

Kraken

- [What is Kraken?](#)
- [Running Kraken](#)
- [Kraken parameters](#)
- [Kraken task report](#)
- [References](#)

What is Kraken?

Kraken is a taxonomic sequence classifier that assigns taxonomic labels to short DNA sequences, typically from microbiome or metagenomic studies (1). Kraken classifies reads to a best-match location in a taxonomic tree (the lowest common ancestor), so not all sequences will be classified to a particular level such as species. Kraken matches k -mers (nucleotide sequences of k bases in length) within a read to a database of k -mer sequences from known genomes with established taxonomic relationships to perform its classifications. Partek Flow currently wraps Kraken version 2.0.8-beta (2), but Kraken 0.10.5-beta is also available (to roll back to the older version, see [Task Management](#)).

Running Kraken

Kraken takes FASTQ files as input. Reads can be single- or paired-end.

- Click a data node with FASTQ files
- Click the **Metagenomics** section of the toolbox
- Click **Kraken**
- Choose a Kraken database
- Configure parameters
- Click **Finish** to run (Figure 1)



The screenshot shows a configuration panel for Kraken. At the top, there is a 'Database name' dropdown menu with 'Bacteria' selected. Below this is a 'Generate unclassified reads' section with an information icon and an unchecked checkbox. The 'Advanced options' section is separated by a horizontal line and contains an 'Option set' dropdown menu with '-- Default --' selected, followed by a 'Configure' button. At the bottom of the panel are two buttons: 'Back' and 'Finish', with a mouse cursor hovering over the 'Finish' button.

Figure 6. Configuring Kraken

Kraken generates a *Taxonomic data* node. This data can be used as input for the [Alpha & Beta diversity](#) and [Choose taxonomic level](#) tasks. If you want to obtain the Kraken output files, select the new data node and choose **Download data** from the toolbox on the right.

Kraken parameters

Database name

Partek distributes the Kraken databases Bacteria, Human, Plasmids, Viruses, and MiniKraken. MiniKraken includes sequences from bacterial, archaeal, and viral genomes in RefSeq; however, it contains only 2.7% of the k -mers from the original database. Running Kraken using the MiniKraken database is significantly faster and less resource-intensive than using a full database, but will not give as complete a result.

Generate unclassified reads

If enabled, an *Unaligned reads* data node including reads that could not be classified by Kraken is produced. Default is *Disabled*.

Advanced options - Quick operation

If enabled, uses the first hit or hits (--quick). Default is *Disabled*.

Kraken task report

The Kraken task report presents a table and graph summarizing the results (Figure 2)

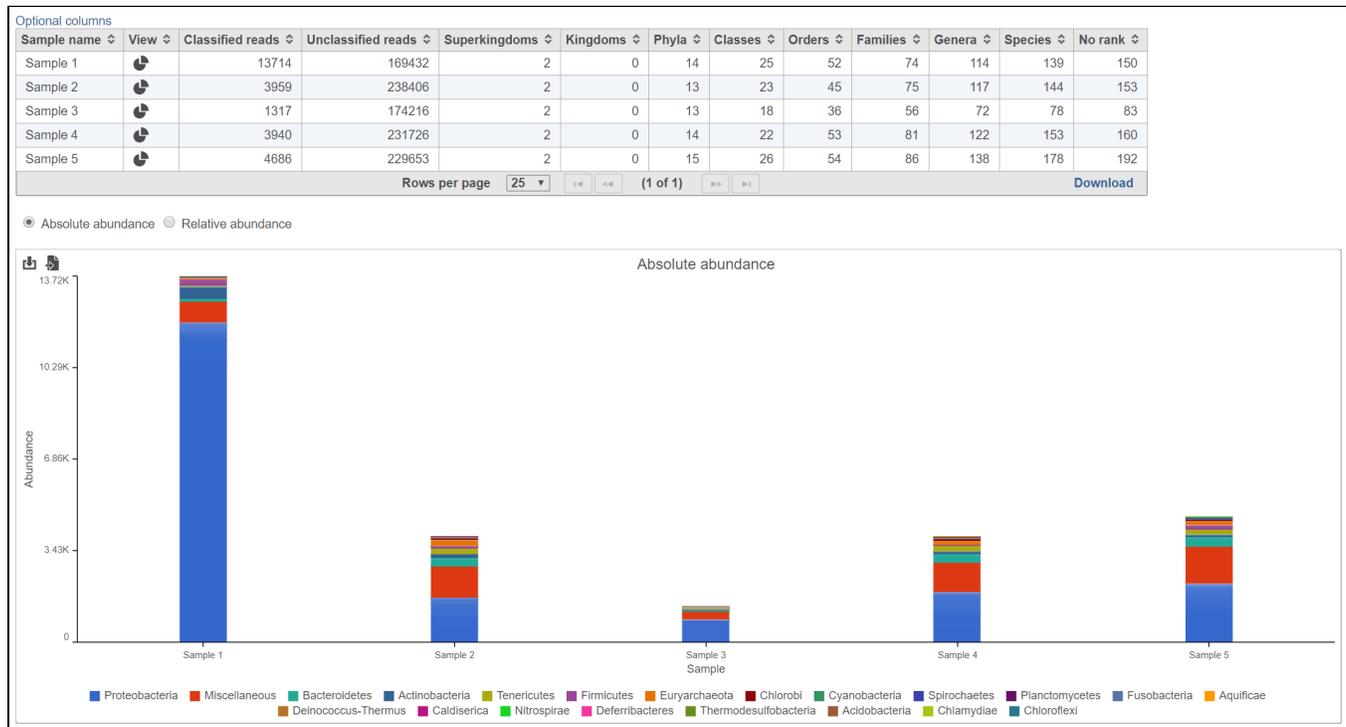


Figure 7. Kraken task report

Table

The table lists each sample and gives the following values:

Classified reads

The number of reads that were classified by Kraken for each sample.

Unclassified reads

The number of reads that were not classified by Kraken for each sample.

Taxonomic levels columns (Superkingdom, Kingdom, Phyla, Classes, Orders, Families, Genera, Species, No rank)

Lists the number of different taxa that were detected within each taxonomic level among the classified reads for each sample. For example, in Figure 2, there were 139 different species detected in Sample 1.

Bar chart

The stacked bar chart shows the abundance or relative abundance of the different phyla. The legend lists the color of each phyla. Mouse over a bar to view the breakdown of families within the phyla (Figure 3). Use the radio button to switch between **Absolute abundance** and **Relative abundance**.

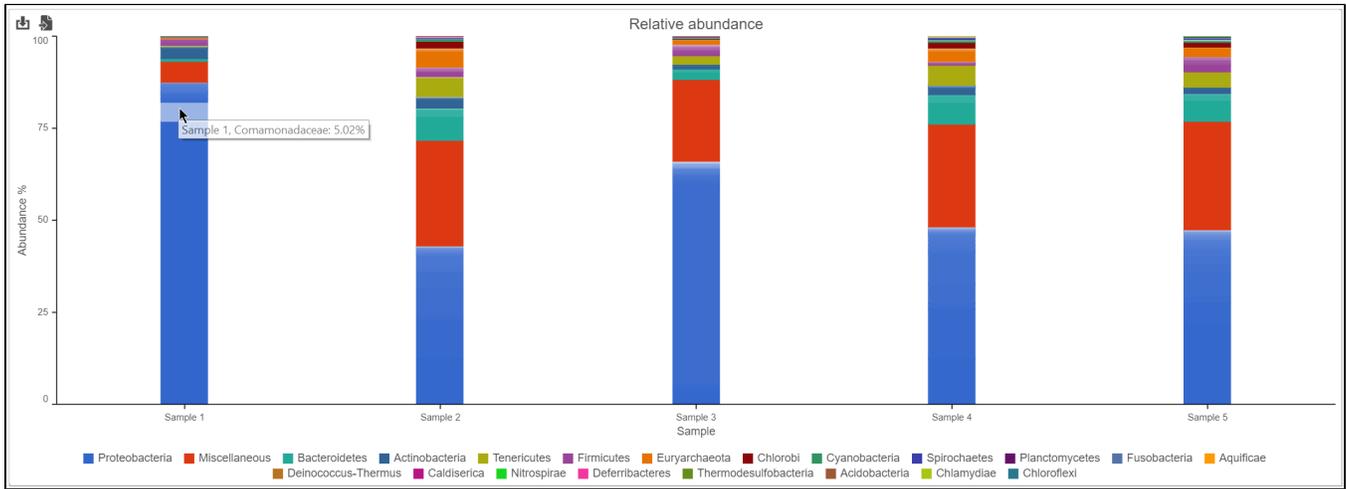


Figure 8. Kraken task report bar chart

Hierarchical pie chart

Click the  in the *View* column of the table opens an interactive Hierarchical pie chart that can be used to explore the taxonomies present in each sample (Figure 4). Mousing over a section of the pie chart gives the number of reads classified to this level, the number of reads in its children, the percentage of total reads represented by this group, and the percentage of the reads in the root that are represented by this group.

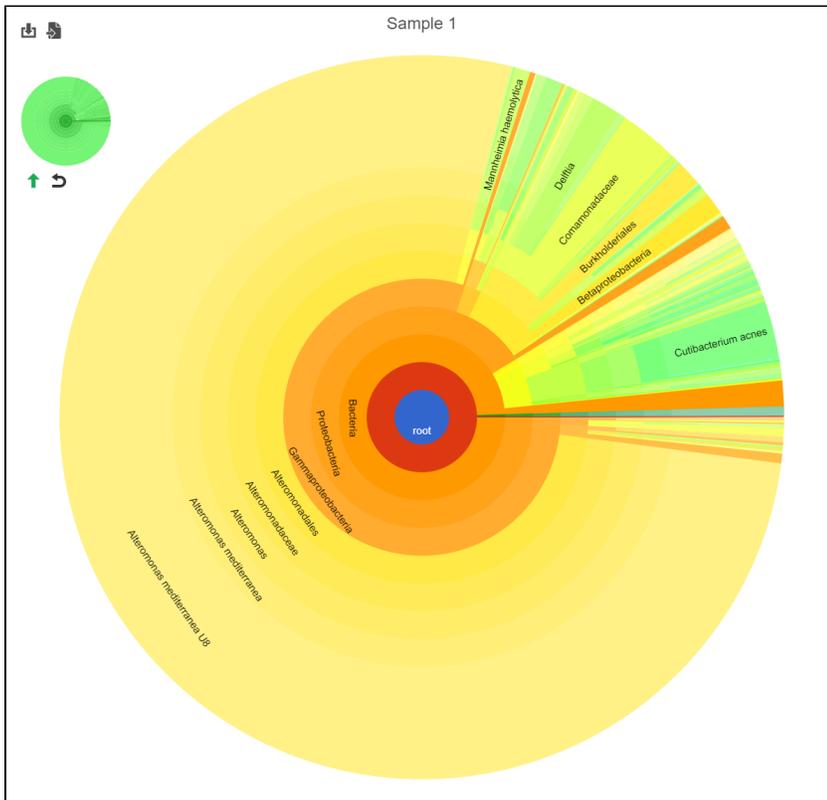


Figure 9. Hierarchical clustering pie chart

Clicking a section of the pie chart zooms in to that section by setting the selected level as the root (Figure 5).

References

1. Wood DE, Salzberg SL: Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 2014, 15:R46.
2. Wood DE, Lu J & Langmead B: Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019, 20:257

Additional Assistance

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.



Your Rating: ☆☆☆☆☆ Results: ★★★★★ 22 rates