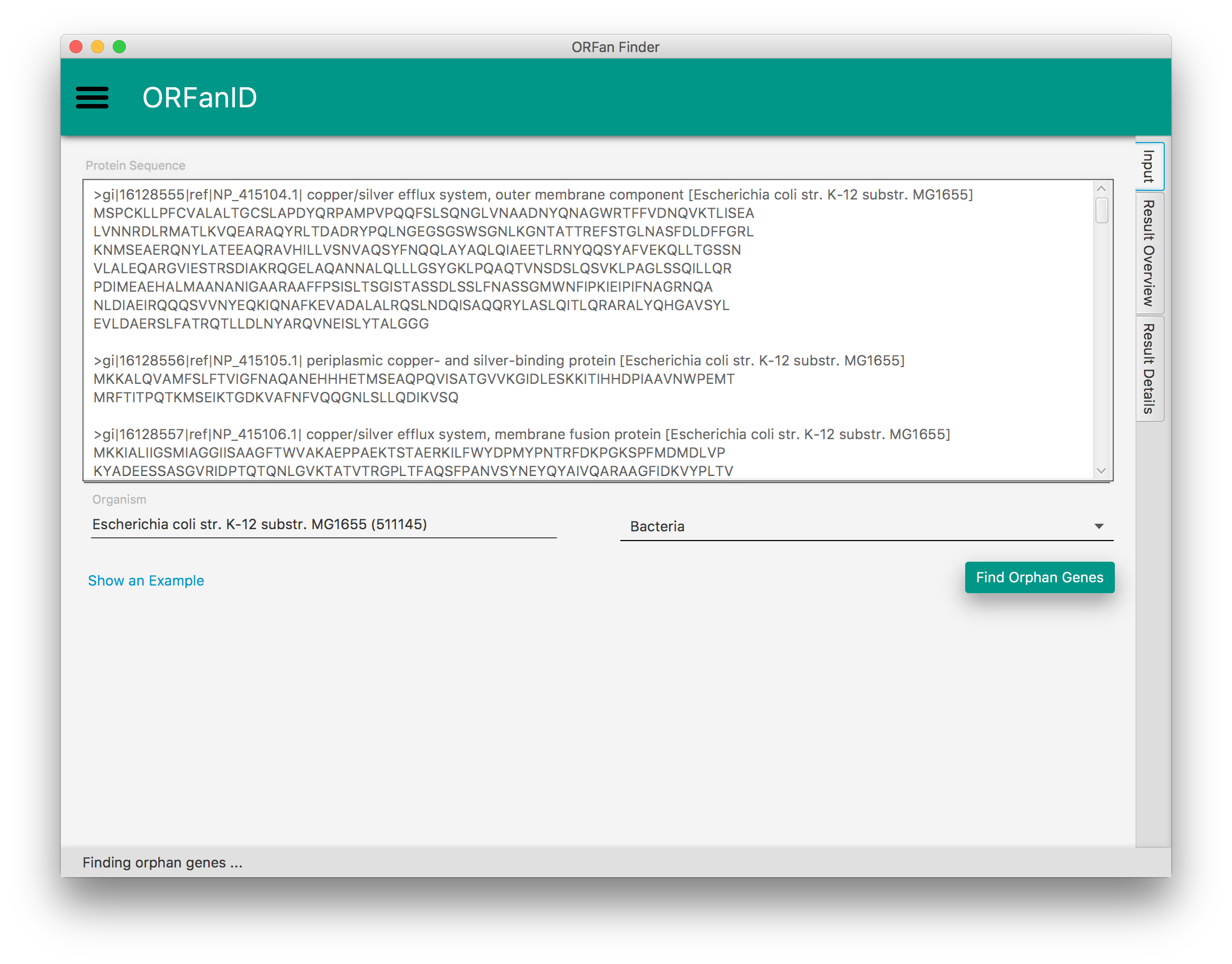
# Materials

34 gene sequences[[1]](#footnote-1) from *Escherichia coli* str. K-12 substr. MG1655(Tax ID: 511145) were used as an example. BLASTP command-line 2.5.0+ were used to run BLAST programme.

# Methods and Procedures

ORFanID accepts amino acid sequence of a gene or multiple genes in FASTA format with their NCBI gene IDs(gi) in the comment lines, organism name and the taxonomy level of the selected organism to restrict the search.

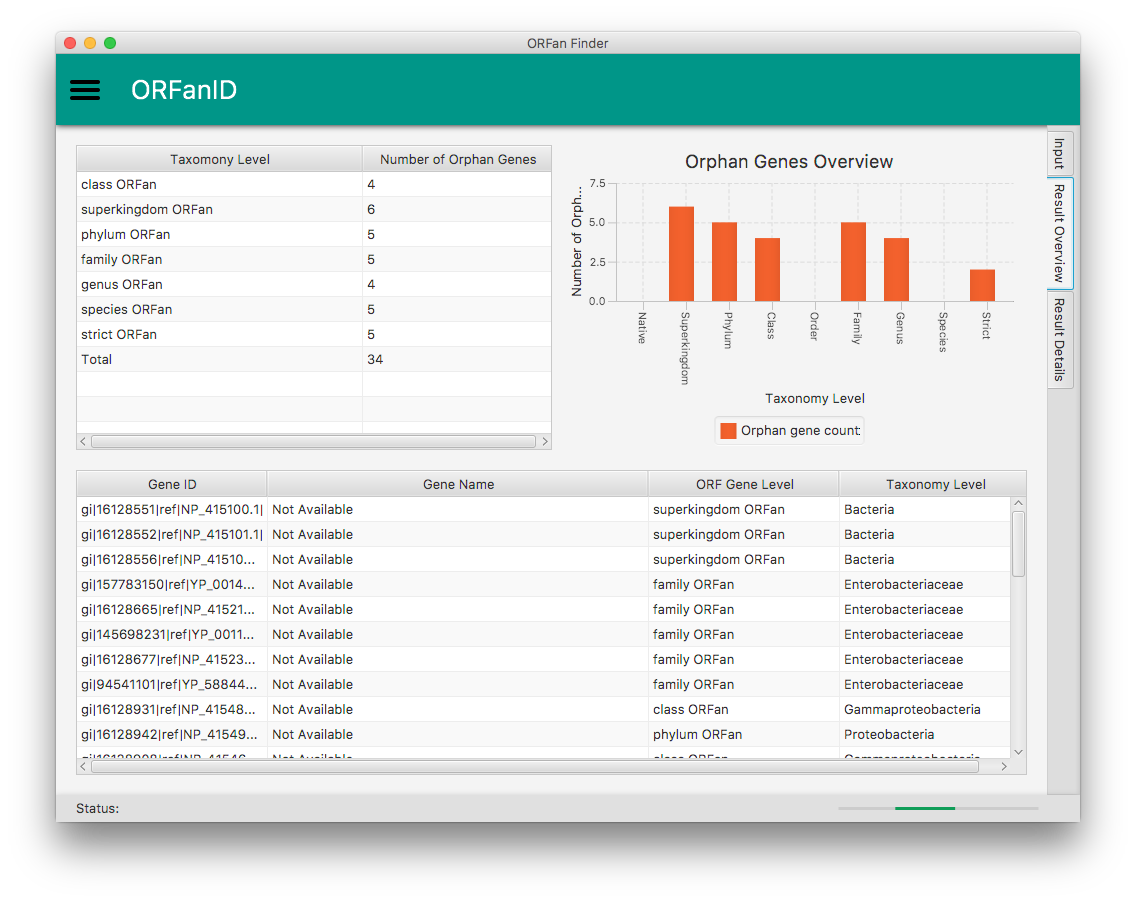


ORFanID engine executes several bash scripts as a pipeline as follows:

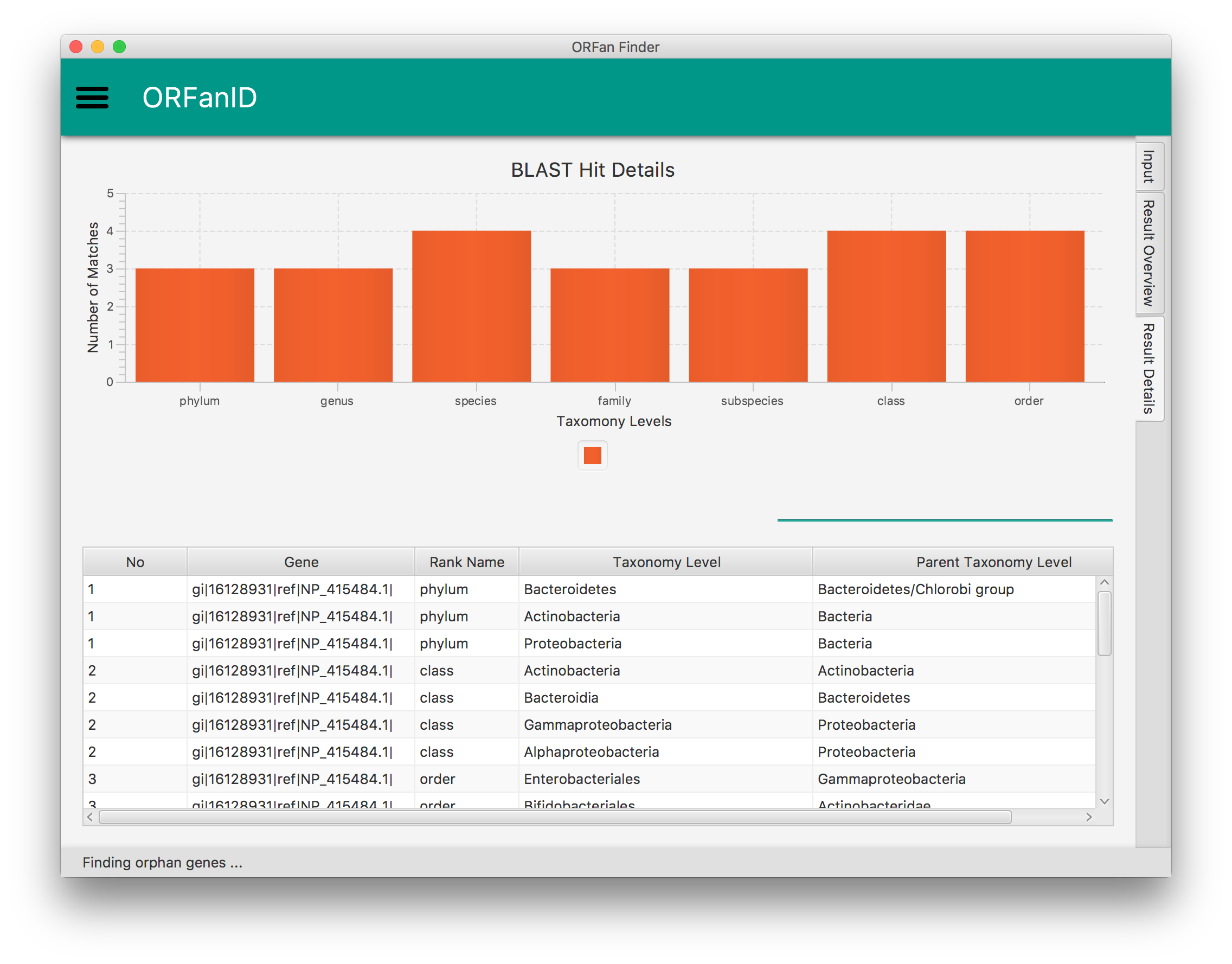
First, ORFanID extracts the gene ID(s) from the comment line(s) of the input FASTA file. Secondly, all the genes will be submitted to the online NCBI BLASTP programme to find homologous sequence against NCBI non-redundant protein database. Accuracy of the results can be adjusted based on the settings such as e-value and maximum number of targets sequences and the taxonomy level (to restrict the search), and results will be reported in a tabular format. Thirdly, genes will be categorized according to filtering algorithm by ORFanFinder command-line tool, based on the homology sequences found from blast results. Output of the ORFanFinder will be further formatted to produce tables and charts in order to visualize results interactively.

# RESULTS

ORFanID interactively visualizes results produced by ORFanFinder tool with Google Material designing concepts.



Top left table summarizes the results by showing the number of ORFan genes found on each taxonomy level of the selected species (Bacteria, in this example) and top right chart visualize the same data graphically to improve user-friendliness. Bottom table shows their categorization level of ORFan genes as well as their actual taxonomy level for each individual gene. Once user selects a gene from the bottom table, the evidences of homologous sequence found from BLASTP will be loaded in to “Result Details” tab.



Bottom table shows Taxonomy levels of different homologous BLAST hits for a selected gene. Number of BLAST matches for each taxonomy are interactively visualized in a chart. For example, if there are hits for Family level, but not in Phylum, Class, Order, then that gene specifically restricted to Family, therefore it’s a Family level ORFan gene. All the tables in the software are sortable by any column and results can be easily filtered with fuzzy search functionality.

ORFanID Web-based application provides certain additional functionalities such as table pagination, exporting results as PDF or CVS formats.

|  |
| --- |
| ../../Desktop/Screen%20Shot%202017-04-24%20at%203.29.13%20PM.p |
| ../../Desktop/Screen%20Shot%202017-04-24%20at%203.29.51%20PM.p |

1. 16127995,16129226,16129671,16129959,226524729,226524725,226524728,16130368,145698296,16130372,16129524,16130369,226524699,16131575,157783150,16128665,145698231,94541101,16128677,16128908,**16128931**,49176107,90111296,90111306,16128942,16128985,16129084,16129100,16129101,16128551,16128552,16128556,16128557,16128555 [↑](#footnote-ref-1)